Physiological roles of sperm-borne small RNAs

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Cell and Molecular Biology

by

Andrew Schuster

Dr. Wei Yan, Dissertation Advisor

May, 2016
We recommend that the dissertation prepared under our supervision by

ANDREW SCHUSTER

Entitled

Physiological Roles Of Sperm-Borne Small RNAs

be accepted in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Dr. Wei Yan, Advisor

Dr. Patricia Berninsone, Committee Member

Dr. Subhash Verma, Committee Member

Dr. David AuCoin, Committee Member

Dr. Grant Hennig, Graduate School Representative

David W. Zeh, Ph. D., Dean, Graduate School

May, 2016
ABSTRACT

The discovery of RNA interference (RNAi) nearly two decades ago revealed a complex and ubiquitous layer of gene expression regulation in both plants and animals. Small noncoding RNAs (sncRNAs), the effectors of RNAi, are present and active in both somatic and germ cells, and appear to play an indispensable role in reproduction. Interestingly, despite cessation of transcription prior to their formation, diverse populations of sncRNAs are present in sperm. While it has been established that sperm-borne sncRNAs are delivered to the oocyte during fertilization, their function and physiological significance in the early embryo remains unclear. To better understand the importance of the paternal sncRNA contribution in the early embryo, this dissertation investigates sperm-borne sncRNAs in multiple contexts. Both miRNA and miRNA / endo-siRNA deficient sperm were generated for fertilization studies, to determine whether the paternal contribution of either class of sncRNA was required for successful early embryo development (Chapter II). To improve our knowledge of the conservation and characteristics of mammalian sperm RNA content, catalogs of both large and small RNA expression were compiled for multiple species, using a methodology that emphasized consistency between datasets (Chapter III). The recent discovery that sperm-borne sncRNAs are involved in epigenetic inheritance prompted an investigation into whether they are active in vinclozolin-induced, epigenetically inheritable disease phenotypes (Chapter IV). In these studies, sperm-borne sncRNAs were consistently found to play a non-trivial physiological role in many aspects of mammalian development.
ACKNOWLEDGEMENTS

I express my sincere gratitude to Dr. Wei Yan for the opportunity to work in his lab. His support and encouragement has been indispensable in my development as a scientist. My time in his lab has exposed me to countless new topics and technologies, many of which I may never have encountered otherwise, and has left me with a richer view of the world. I am also grateful to the other members of the Yan lab for all of their help over the years.

I also would like to extend my thanks to the other members of my committee, Dr. Hennig, Dr. Verma, Dr. AuCoin, and Dr. Berninsone, for providing me with insightful comments on my work and enabling my growth as a scientist.

And, of course, I also would like to thank my friends and family for their support, as well as their continued acceptance of “grad school” as the panacea of excuses.

Lastly, I would like to thank my own fledgling family – the emotional and intellectual support I have received from Liz, along with the less tangible support from Buddy, is immeasurable. Without them, I would not be where I am today.
TABLE OF CONTENTS

ABSTRACT................................................................................................................................. i

ACKNOWLEDGEMENTS........................................................................................................... ii

TABLE OF CONTENTS.............................................................................................................. iii

LIST OF FIGURES.................................................................................................................. vi

LIST OF TABLES..................................................................................................................... viii

CHAPTER I: Introduction to small noncoding RNAs and their relation to mammalian sperm................................................................................................................................. 1

Small noncoding RNAs… in sperm?..................................................................................... 1

Micro RNA............................................................................................................................... 3

Endogenous small interfering RNA.................................................................................... 5

PIWI-interacting RNA........................................................................................................... 6

tRNA-derived small RNA..................................................................................................... 9

Sperm-borne sncRNA and the embryo.............................................................................. 11

Sperm-borne sncRNA and epigenetic transgenerational inheritance.............................. 13

REFERENCES....................................................................................................................... 17

CHAPTER II: Sperm-borne miRNAs and endo-siRNAs are important for fertilization and preimplantation embryonic development................................................................. 24

ABSTRACT............................................................................................................................... 25

INTRODUCTION....................................................................................................................... 26

RESULTS.................................................................................................................................... 28

DISCUSSION............................................................................................................................. 39

ACKNOWLEDGEMENTS........................................................................................................ 44

MATERIALS AND METHODS............................................................................................... 45

FIGURES.................................................................................................................................... 51

TABLES..................................................................................................................................... 60

REFERENCES.......................................................................................................................... 62
MATERIALS.............................................................................................................. 164
METHODS.............................................................................................................. 165
NOTES..................................................................................................................... 170
FIGURES.............................................................................................................. 171
REFERENCES........................................................................................................ 177
LIST OF FIGURES

1-1. RNAi small RNA biogenesis........................................................................................................ 4
1-2. Possible tsRNA subclasses.......................................................................................................... 10
1-3. Epigenetic transgenerational inheritance.................................................................................. 15

2-1. Inactivation of Dicer and Drosha in spermatogenic cells leads to oligo asthenoteratozoospermia........................................................................................................................................ 51
2-2. Dicer and Drosha cKO sperm display altered miRNA and endo-siRNA expression profiles, as revealed by sncRNA-Seq analyses........................................................................................................ 52
2-3. Histogram showing the rate of development of embryos derived by ICSI using WT oocytes and Dicer cKO sperm with or without sperm RNA supplementation.................................. 53
2-4. Histogram showing the rate of development of embryos derived by ICSI using WT oocytes and Drosha cKO sperm with or without sperm RNA supplementation............ 54
2-5. Histogram showing the rate of live-born pups derived from ICSI using WT oocytes and WT, Dicer cKO, or Drosha cKO sperm with or without supplementation of sperm-borne RNAs.................................................................................................................. 55
2-6. Histogram showing the rate of development of 2PN embryos derived from ICSI using Drosha cKO oocytes and WT, Dicer cKO and Drosha cKO sperm.......................................... 56
2-7. Heat map showing miRNA profiles in WT oocytes, WT sperm and 2PN embryos derived from ICSI using WT, Dicer cKO and Drosha cKO sperm, as determined by low-input sncRNA-Seq.......................................................... 56
2-8. Heat map showing expression profiles of 96 genes in single oocytes and preimplantation embryos derived from ICSI using WT, Dicer and Drosha cKO sperm at 2PN, 2-cell, 4-cell, morula and blastocyst stages................................................................. 57
2-9. Fold changes of dysregulated genes with or without significantly altered targeting miRNAs in Dicer and Drosha cKO sperm in 2PN and 2-cell embryos derived from ICSI using Dicer and Drosha cKO sperm................................................................. 58

3-1. Sperm RNA quality control........................................................................................................ 95
3-2. Sperm large RNA .................................................................................................................. 96
3-3. Sperm small RNA contents in four mammalian species including the mouse, rat, rabbit, and human.................................................................................................................. 97
3-4. Gene ontology and 5’ halve targeting example........................................................................ 99
3-S1. SpermBase home page ................................................................. 100
3-S2. Sperm-borne piRNA .................................................................................. 101
3-S3. Length distribution of sncRNA-Seq reads that aligned to mitochondrial genes.. 102
3-S4. Length distribution of abundant mouse total sperm mitosRNAs.................. 103
3-S5. Length distribution of abundant human total sperm mitosRNAs.................. 104
3-S6. Length distribution of abundant rat total sperm mitosRNAs..................... 105
3-S7. Length distribution of abundant rabbit total sperm mitosRNAs............... 106
3-S8. Length distribution of sncRNA-Seq reads aligned to tRNA genes.............. 107

4-1. Differentially expressed sncRNAs in vinclozolin lineage F3 sperm............... 141
4-2. Relative numbers of predicted 5’ halve gene targets in 5’ UTR, CDS, and 3’ UTR sequences ........................................................................................................................................... 143
4-3. GO term analyses of dysregulated sncRNAs in the vinclozolin lineage F3 sperm ......................................................................................................................................................... 144

S1-1. Flowchart outlining our novel endo-siRNA identification method............ 171
S1-2. Sequery v1.0 window example ..................................................................... 172
S1-3. Graphical examples of dsRNA precursors for endo-siRNA..................... 173
S1-4. Example of putative nat-cis-dsRNA precursor for endo-siRNA-T32 ......... 174
S1-5. Example of putative nat-trans-dsRNA precursor for endo-siRNA-T19 ....... 175
S1-6. Example of putative hairpin precursor for endo-siRNA-T18 ................. 176
## LIST OF TABLES

2-1. Preimplantation development of embryos derived from ICSI using WT oocytes and Dicer cKO sperm with or without sperm RNA supplementation ........................................... 60

2-2. Preimplantation development of embryos derived from ICSI using WT oocytes and Drosha cKO sperm with or without sperm RNA supplementation ........................................... 60

2-3. Pups born through transfer of 2-cell embryos derived from ICSI using WT oocytes and WT, Dicer cKO, Drosha cKO sperm with or without supplementation of sperm-borne RNAs .............................................................................................................................................. 61

2-4. Preimplantation development of embryos derived from ICSI using Drosha cKO oocytes and WT, Dicer cKO and Drosha cKO sperm ............................................................................. 61

3-1. Summary of the lysis procedure used for each of the four species studied .......... 108

3-2. Predicted miRNA and 5’ halve gene targets are significantly present in early development stages ............................................................................................................................ 108

3-S1. xi and θ values used for RNAhybrid .................................................................... 109

3-S2. Conserved sperm-borne coding genes ................................................................ 109

3-S3. Enriched GO terms (conserved total sperm coding genes) .................................. 110

3-S4. Small RNA distribution ...................................................................................... 110

3-S5. Conserved sperm miRNA .................................................................................. 110

3-S6. Conserved miRNA gene target predictions ....................................................... 111

3-S7. Enriched GO terms (conserved miRNA predicted targets) ................................ 111

3-S8. 5’ halve gene target predictions ....................................................................... 112

3-S9. Enriched GO terms (5’ halve predicted targets) ................................................ 112

4-1. Differential expressed sncRNAs in vinclozolin lineage F3 sperm ...................... 146

4-2. Proximity of differential expressed sncRNAs to DMRs ........................................ 147

4-3. Correlation between genes proximal to F3 sperm DMRs and predicted gene targets ................................................................................................................................. 148

4-4. Correlation between genes proximal to F3 PGC DMRs and predicted gene targets ................................................................................................................................. 149
CHAPTER I

Introduction to small noncoding RNAs and their relation to mammalian sperm

The development of a single cell embryo into a fully formed organism requires a highly fine-tuned program of gene expression, and thus, a robust regulatory mechanism. Since the discovery of RNA interference almost two decades ago, small noncoding RNAs (sncRNAs) have been implicated in many systems of epigenetic regulation. The developing embryo is no exception – sncRNAs play an integral role throughout embryogenesis (Suh and Blelloch, 2011). At conception, the sperm delivers a repertoire of sncRNA, along with the paternal genome, to the oocyte (Ostermeier et al., 2004). The importance of these paternal sncRNAs in early development, as well as the development of the sperm themselves, has only recently been elucidated. In this chapter, we will review the basic characteristics and functions of the different sncRNA classes present in mammalian sperm and early embryos.

Small noncoding RNAs… in sperm?

During fertilization, the mature spermatozoa (i.e., sperm) enters the oocyte and releases, among other components, its genome into the cell. To successfully traverse the female reproductive tract and penetrate the oocyte, sperm must be highly specialized; sperm possess a strong flagellum and hydrodynamic, damage-resistant head that make them particularly well-equipped to safely deliver their genetic information to the oocyte (Johnson et al., 2011; Miller, 2014). These morphologically distinct cells are derived from primordial germ cells (PGCs), which undergo a series of dramatic transformations called spermatogenesis. PGCs differentiate into spermatogonia, which proliferate via mitosis,
generating primary spermatocytes (Kornbluth and Fissore, 2015; Yadav and Kotaja, 2014). The primary spermatocyte undergoes meiosis, ultimately resulting in four haploid round spermatids (Yan, 2009). In the next phase of spermatogenesis, referred to as spermiogenesis, round spermatids acquire their canonical sperm head and tail structure through a process called elongation. During elongation, the bulk of the round spermatids’ cytoplasm is shed and the nucleus (located in the sperm head) becomes highly condensed as histones are replaced by protamines (Yan, 2009; Miller, 2015). It is also during this time that transcription ceases in the developing male germ cell (Rathke et al, 2014; Kierszenbaum and Tres, 1975). The restructuring that takes place during spermiogenesis allows the sperm to accomplish its primary objective – the delivery of the paternal genome to the oocyte (Johnson et al, 2011).

Considering that sperm are streamlined to be efficient delivery vectors of the paternal genome, contain almost no cytoplasm, and are transcriptionally quiescent, it may seem odd that they would contain any sncRNAs, which are mainly known as regulators of gene expression. After their initial discovery, sperm-borne RNAs were widely thought to be either contaminants from other cell types or merely the remnants of spermatogenesis (Krawetz, 2005; Miller, 2014). It is now fairly established that sperm do indeed contain their own populations of RNA, both large (>200 nucleotides) and small (<200 nucleotides), and that many of these transcripts are selectively retained during elongation (Krawetz et al, 2011; Miller and Ostermeier, 2006; Sendler et al, 2013; Jodar et al, 2013). Additionally, sperm deliver their RNA contents to the oocyte during fertilization, further suggesting that their presence in sperm is not trivial (Ostermeier et al, 2004). In light of these observations, the potential functional role(s) sperm-borne
RNA play in the early embryo has been studied with increasing intensity in recent years.

To contextualize the current research on the biological significance of sperm-borne sncRNAs in the early embryo, the canonical activities of the classes of sncRNA observed in sperm must first be discussed.

**Micro RNA**

Micro RNAs (miRNAs) are perhaps the most widely studied of the sncRNA classes. These sncRNA are, on average, 22 nucleotides (nt) in length and are single-stranded (Bartel, 2004). Mature miRNA are derived from much longer transcripts, known as primary miRNA (pri-miRNA) transcripts. These pri-miRNAs are typically transcribed by RNA polymerase II (pol II), and can contain a single mature miRNA or several (polycistronic) different mature miRNA species (Holley and Topkara, 2011). The pri-miRNA is processed in the nucleus by a protein complex consisting of DiGeorge syndrome critical region 8 (DGCR8) and Drosha. DGCR8 recognizes the ‘hairpin’ (also referred to as a ‘stem loop’) secondary RNA structures present in the pri-miRNA and binds to the transcript, allowing Drosha (a member of the RNase III enzyme family) to cleave the pri-miRNA at the base of the hairpin(s), thereby generating the precursor miRNA (pre-miRNA) (Holley and Topkara, 2011; Sperber et al., 2014). The pre-miRNA departs the nucleus via Exportin5, a nuclear export protein, and is cleaved by another RNase III family enzyme, Dicer, in the cytoplasm (Kim et al., 2016). Dicer produces a short (~22 nt) miRNA duplex by removing the loop at the end of the hairpin structure. This cleavage is stabilized by TAR RNA binding protein (TRBP), which complexes with Dicer (Chendrimada et al., 2005; Romero-Cordoba et al., 2014). One of the duplex
strands, called the ‘guide’ strand, associates with the RNA-induced silencing complex (RISC), which contains Dicer, Argonaute 2, and TRBP, while the other is discarded. RISC determines which of the two RNAs will act as the guide strand through the structure of the duplex – the strand with the less thermodynamically stable 5’ end is often selected (Czech and Hannon, 2011; Holley and Topkara, 2011). After the guide strand

Figure 1-1 – RNAi small RNA biogenesis (a) miRNA biogenesis. Primary miRNAs (pri-miRNAs) are processed by Drosha, creating precursor miRNAs (pre-miRNA) in the nucleus. The pre-miRNAs are exported into the cytoplasm and cleaved by Dicer, then loaded onto an Argonaute protein (part of the RISC; not shown). (b) endo-siRNA biogenesis. Precursor transcripts can be double-stranded cis or trans (cis-dsRNA and trans-dsRNA, respectively), or single-stranded containing a hairpin structure. Precursors are cleaved in a Dicer-dependent, Drosha-independent mechanism. The mature endo-siRNA is loaded onto Argonaute 2 (part of the RISC; not shown). (c) piRNA biogenesis. The precursor transcripts to piRNA are single-stranded. Primary (pre-pachytene) piRNAs are first produced, followed by secondary piRNAs, in a MILI and MIW12 dependent pathway called the ping-pong mechanism. Pachytene piRNA are generated by a mechanism similar to the one shown for primary piRNA production. Adapted from (Siomi et al., 2011).
(i.e., the mature miRNA) is loaded onto RISC, it is able to perform its canonical function – post-transcriptional regulation of gene expression, also referred to as RNA interference (RNAi) (Figure 1-1a) (Holley and Topkara, 2011).

In order to regulate gene expression, RISC binds to sequences complementary to its associated mature miRNA; these complementary sequences typically lie within the 3’ untranslated region (3’ UTR) of the mRNA (Holley and Topkara, 2011). The match between the miRNA and mRNA does not need to be perfect; on the contrary, miRNA-mRNA binding is heavily dependent on the ‘seed region,’ a short (~8 nt) conserved sequence at the 5’ end of the miRNA (Friedman et al, 2009). Reliance on the seed region allows a single miRNA species to target several hundred unique mRNAs – not surprisingly, over half of all human mRNAs are thought to be targeted by a miRNA (Bartel, 2009; Friedman et al, 2009). When a miRNA is partially complementary to its target mRNA, the transcript is deadenylated and further destabilized via hydrolysis of its 5’ cap, and as a result, the mRNA is susceptible to 5’ → 3’ exonucleolytic degradation (Jonas and Izaurrealde, 2015). Occasionally, a miRNA will perfectly complement its target mRNA – in this scenario, the mRNA is directly cleaved by the Argonaute protein in RISC (Ameres and Zamore, 2013). In both cases, translation of the targeted mRNA is prevented and thus, the expression of the gene is ‘silenced’ (Holley and Topkara, 2011).

**Endogenous small interfering RNA**

The ability to post-transcriptionally repress gene expression is not exclusive to miRNAs. Endogenous small interfering RNAs (endo-siRNAs) are ~21 nt in length and single-stranded in their mature form, and also like miRNAs, participate in RNAi (Song et
The longer precursor transcripts that contain the endo-siRNA can be single-stranded (containing a long hairpin structure) or double-stranded. The latter consists of two partially complementary transcripts originating from the same or separate loci (cis and trans, respectively) (Song et al, 2011; Okamura and Lai, 2008; Nilsen, 2008). Endo-siRNA biogenesis and function is Dicer-dependent and, unlike miRNAs, Drosha-independent (Luo et al, 2016). The cleavage of the precursor transcripts by Dicer generates a duplex from which one single-stranded (mature) endo-siRNA associates with the Argonaute protein present in RISC (Okamura and Lai, 2008; Luo et al, 2016). The RISC-bound endo-siRNA guides the complex to an mRNA with a perfectly complementary sequence (Tam et al, 2008; Xia et al, 2012). Targeted mRNAs are cleaved directly by RISC, thereby silencing gene expression (Figure 1-1b) (Xia et al, 2012).

Unlike miRNAs, which have been observed in a myriad of mammalian somatic and germ cells, endo-siRNAs are mostly confined to reproduction (Golden et al, 2008; Ivey and Srivastava, 2015). In mammals, endo-siRNA have been observed in testes (Song et al, 2011), oocytes (Tam et al, 2008), and embryonic stem cells (Babiarz et al, 2008), as well as skin (Xia et al, 2012) and breast cells (Chen et al, 2012). Other mammalian cells may also possess endo-siRNAs which have yet to be annotated – the identification of novel endo-siRNA is discussed in Supplementary Chapter I.

**PIWI-interacting RNA**

The third and final member of the canonical RNAi family of sncRNAs, PIWI-interacting RNA (piRNA), is dissimilar to miRNAs and endo-siRNAs in several respects.
Unlike their RNAi relatives, piRNAs are Dicer-independent for their biogenesis or function (Bortvin, 2013). As their name suggests, piRNAs associate with PIWI proteins, which are germline-specific members of the Argonaute family (Siomi et al, 2011). In mice, three PIWI proteins, MILI, MIWI, and MIWI2, comprise the main components of the piRNA biogenesis machinery (Iwasaki et al, 2015). All three are abundantly expressed in male germ cells, but only MILI and MIWI have been observed in murine oocytes (Aravin et al, 2008; Ding et al, 2013; Roovers et al, 2015). The failure of spermatogenesis in MILI, MIWI, and MIWI2 knockout mice established piRNAs as a requirement for successful male germ cell development, whereas the deletions of these genes had no effect on female fertility (Kuramochi-Miyagawa et al, 2004; Carmell et al, 2007; Deng and Lin, 2002). MILI is expressed first in male PGCs at embryonic day 12.5 (E12.5), followed by MIWI2 at E15.5 (Aravin et al, 2008). MILI expression continues throughout the development of the male germ cell, whereas MIWI2 expression ceases at post-natal day 3 (P3) (Aravin et al, 2008; Bortvin, 2013). MIWI is expressed much later during spermatogenesis, beginning in pachytene spermatocytes and increasing as the spermatocytes become round spermatids (Bortvin, 2013; Vourekas et al, 2012). In male germ cells, piRNAs can be divided into two sub-classes: pre-pachytene and pachytene. Pre-pachytene piRNA are expressed first in PGCs and are dependent on MIWI2 and MILI for their biogenesis. The pachytene piRNA, which are dependent on MILI and MIWI, are expressed much later at the pachytene stage of meiosis and are abundant in round spermatids (Aravin et al, 2007; Vourekas et al, 2012; Ortogero et al, 2014).

In addition to being expressed at different periods of development, pre-pachytene and pachytene piRNA contrast in their biogenesis, function, and nucleotide sequence.
Pre-pachytene piRNA are generated through a positive feedback loop called the ping-pong mechanism. First, precursor transcripts derived from repetitive sequences (e.g., transposable elements) are processed to form primary piRNA (~26 nt) which associate with MILI. The MILI-primary piRNA complex binds to a complementary sequence in another transcript, cleaving it 10 nt from the 5’ end of the primary piRNA, generating a secondary piRNA (~28 nt), while also silencing the transcript. The secondary piRNA, which associates with MIWI2, targets the original precursor transcript, thus repeating the cycle (Figure 1-1c) (Aravin et al., 2008; Beyret et al., 2012; Gunawardane et al., 2007). Primary piRNAs have a bias at their 5’ end (1st nt) for uracil; as a result, secondary piRNAs have both a 1st and 10th nt bias for uracil and adenine, respectively (Aravin et al., 2008; Gunawardane et al., 2007). The ability to silence transcripts allows the pre-pachytene piRNAs to protect the integrity of the genome from deleterious transposon and repetitive sequence expression that coincides with the expression of MILI and MIWI2.

Starting at E8, PGCs undergo a period of epigenetic reprogramming in which the bulk of genomic DNA methylation is erased (E8 – E13.5) and then reset via de novo methylation (E14 – E17), thus establishing the germline-specific methylome. During this wave of reprogramming, pre-pachytene piRNA-mediated silencing is critical in ensuring that temporarily demethylated and up-regulated transposons and repetitive sequences are unable to wreak havoc in the genome (Ly et al., 2015; Yan, 2014; Bortvin, 2013).

The biogenesis of pachytene piRNA is independent of the ping-pong mechanism (Beyret et al., 2012). Instead, pachytene piRNAs are derived from long single-stranded transcripts, each of which contains a myriad of different piRNA species (Li et al., 2013a; Li et al., 2013b). The processed piRNA (~30 nt) associates with MIWI (complexed with
other proteins) and silences gene expression by binding partially complementary sequences in the 3’ UTRs of mRNA and mediating their degradation, not unlike miRNA. Similar to primary pre-pachytene piRNAs, pachytene piRNAs have a 1st nt bias towards uracil (Vourekas et al, 2012; Ortega et al, 2014; Aravin et al, 2007). Pachytene piRNAs are responsible for eliminating the majority of the mRNAs present in elongating spermatids, thereby controlling which coding transcripts are retained in the mature sperm (Gou et al, 2014).

**tRNA-derived small RNA**

While sperm are fairly diverse in their sncRNA content, tRNA-derived small RNAs (tsRNAs) were recently identified as the dominant class of sncRNA in terms of expression (Peng et al, 2012). This delayed observation is not surprising – relative to the aforementioned sncRNA, the biogenesis and functions of tsRNAs remain fairly enigmatic in mammals. Mature tRNAs can be cleaved by several different mechanisms, generating multiple sub-classes of tsRNAs. In terms of sequence length, tRNA fragments (tRFs) are 19 – 26 nt, while tRNA halves are typically 30 – 35 nt (Raina and Ibbá, 2014; Gebetsberger and Polacek, 2013). Additionally, tRFs and tRNA halves are categorized by the portion of the mature tRNA from which they were derived – either the 5’ (tRF-5 and 5’ halves) or 3’ (tRF-3 and 3’ halves) portion of the tRNA, divided at the anticodon loop (Figure 1-2) (Gebetsberger and Polacek, 2013). In sperm, the vast majority of the tsRNAs are 5’ halves (Peng et al, 2012). For the sake of brevity, 5’ halves will be the focus of this section.
The biogenesis of sperm-borne 5’ halves has not been fully elucidated. Recent observations by Sharma et al suggest that epididymosomes, vesicles that fuse with sperm as it passes through the epididymis after spermiogenesis, deliver 5’ halves to sperm; however, the mechanism by which these sncRNA are produced remains unknown (Sharma et al, 2016). In other instances of 5’ halve expression in mammals, Angiogenin, a member of the RNase A family, cleaves mature tRNA within the anticodon loop; this cleavage is induced by stress conditions (Yamasaki et al, 2009; Saikia and Hatzoglou, 2015; Ivanov et al, 2011; Emara et al, 2010; Raina and Ibba, 2014). Angiogenin is present in human testis but has not been observed in sperm, suggesting that sperm-borne 5’ halves may be generated through a non-canonical mechanism (Koga et al, 2004). Cytosine-5 methylation (m^5C) in mature tRNA may also play a role in regulating 5’ halve
expression. Depletion of the methyltransferases responsible for tRNA m$^5$C marks, Nsun2 and Dnmt2, leads to increased Angiogenin-mediated 5′ halve production (Blanco et al., 2014; Schaefer et al., 2010; Tuorto et al., 2012; Hussain et al., 2013).

Previous studies suggest that 5′ halves act as suppressors of translation in mammalian cells, including sperm (Raina and Ibba, 2014; Fu et al., 2015; Chen et al., 2016). In one in vitro study, 5′ halves inhibited translation initiation by displacing eIF4G/eIF4A from 5′ ends of mRNA (Ivanov et al., 2011). Sperm-borne 5′ halves were recently postulated to regulate gene expression by targeting the promoters of genes (Chen et al., 2016). As will be discussed in later chapters, our research suggests that sperm-borne 5′ halves affect gene expression via complementary sequences in the 5′ untranslated regions (5′ UTR) of mRNAs (Schuster et al., 2016). Further investigation is required before the true biological significance of 5′ halve abundance in sperm can be elucidated.

**Sperm-borne sncRNA and the embryo**

The retention of specific RNAs in sperm, coupled with the resource commitment associated with retaining these transcripts, implies that sperm-borne RNAs hold some importance in reproduction (Hosken and Hodgson, 2014; Krawetz, 2005; Miller, 2014). RNA is only one of the putative epigenetic contributions of the male during fertilization. During spermiogenesis, the majority of the histones are replaced with protamines in order to condense the paternal genome for delivery; however, a small portion of histones remain, in a seemingly non-random pattern (Yan, 2009; Miller, 2015; Hammoud et al., 2009). The retention of histones at specific promoters may allow the paternal genome to
quickly express important early development genes. Not surprisingly, histones were significantly enriched at development-related loci, such as the \textit{HOXD} cluster, a set of transcription factors that play a vital regulatory role in morphogenesis (Sheth \textit{et al}, 2013; Hammoud \textit{et al}, 2009). In addition to post-transcriptional silencing, miRNAs can regulate gene expression by influencing transcription itself by targeting complementary sequences in promoters, inducing chromatin restructuring (Kim \textit{et al}, 2008; Younger and Corey, 2011; Ross and Kassir, 2014). It is possible that sperm-borne miRNAs influence gene expression by regulating which histones are retained in the sperm, thus playing a regulatory role in the early embryonic transcriptome (Jodar \textit{et al}, 2013; Krawetz \textit{et al}, 2011).

Sperm-borne miRNAs likely regulate early embryonic gene expression through their canonical post-transcriptional silencing activity as well. In a study by Hammoud \textit{et al}, eight sperm-specific miRNAs (i.e., absent in the oocyte, present in the sperm) were inhibited via miRNA sponges (i.e., transcripts that contain multiple copies of sequences complementary to the miRNA you are attempting to inhibit – miRNA silencing is ‘absorbed’ by the relatively abundant sponges) in the zygote (Ebert and Sharp, 2010; Hammoud \textit{et al}, 2014). As a result, the blastulation of injected embryos decreased, demonstrating the potentially critical roles sperm-borne sncRNA can play in early development (Hammoud \textit{et al}, 2014). Not surprisingly, the fertility potential of sperm has previously been correlated with altered miRNA expression in both humans (Salas-Huetos \textit{et al}, 2015; Abu-Halima \textit{et al}, 2013) and bulls (Fagerlind \textit{et al}, 2015; Govindaraju \textit{et al}, 2012). Whether these alterations are responsible for the observed infertility of the sperm or are just an indirect effect of another defect is unclear; despite this, these altered
miRNA profiles have the potential to be used as biomarkers of fertility (Krawetz, 2005; Miller and Ostermeier, 2006). Chapter II provides additional evidence for the importance of sperm-borne sncRNAs in early development. Sperm obtained from both Dicer and Drosha conditional knockout (cKO) mice were tested for their fertility potential, against a wild type (WT) control. In both cases, the deficit of miRNAs and endo-siRNAs (Dicer-cKO) or just miRNAs (Drosha-cKO) lead to a decline in developmental potential of embryos produced by intracytoplasmic sperm injection (ICSI) of the cKO sperm into WT oocytes; developmental potential could be rescued by supplementing ICSI with WT sperm RNA, indicating that the loss of paternal miRNAs and endo-siRNAs had a negative impact on embryogenesis (Yuan et al., 2016).

Sperm-borne sncRNA and epigenetic transgenerational inheritance

During fertilization, the father contributes its share of the zygote’s genetic information, and in many cases, an epigenetic contribution is made as well. The epigenomes of many different organisms are susceptible to environmental insults in the form of toxins, behavior, stress, and so forth (Yan, 2014; Skinner et al., 2012; Nilsson and Skinner, 2015; Wu et al., 2015). The primordial germ cells (PGCs) provide the basis for the male germ cell, culminating in mature sperm (Wu et al., 2015; Ly et al., 2015). PGCs are particularly susceptible to perturbations, partially because of the epigenetic reprogramming that occurs starting at embryonic day 8 (E8) (Ly et al., 2015). From E8 to E13.5, the PGC migrates to the fetal gonad. At this time, the bulk of the paternal methylome is erased, in order to create a fresh slate for the subsequent period of reprogramming (Ly et al., 2015; Seisenberger et al., 2013; Skinner et al., 2011) At E14, the
PGC undergoes *de novo* methylation, including imprinted genes, differentially methylated regions (DMRs), and retrotransposons – this has been estimated to last either to E17 or as long as post-natal day 2 (P2) (Ly *et al*., 2015; Yan, 2014; Morselli *et al*., 2015). During these periods of reprogramming, certain exposures and diseases are able to manipulate the epigenome of the developing embryo (Yan, 2014; Wu *et al*., 2015).

Whether these alterations in the epigenome are caused by direct changes to the methylome or to the transcriptome (or perhaps by some other mechanism) is not fully understood (Ly *et al*., 2015; Yan, 2014). Adult epigenomes can also be altered, and the associated phenotype can be passed down to future generations through sperm (Wu *et al*., 2015). In some cases, paternally transmitted epimutations and their effects persist until the F3 generation, at which point the ancestor had never been directly exposed to the environmental factor – this phenomenon is called epigenetic transgenerational inheritance, or ETI (Figure 1-3) (Skinner, 2011; Yan, 2014). ETI can be induced by conditions such as altered behavior (e.g., chronic stress), abnormal diet (e.g., high-fat, high-sugar diet), and exposure to toxins in the environment (e.g., BPA, vinclozolin) (Nilsson and Skinner, 2015; Chen *et al*., 2016; Grandjean *et al*., 2015; Rodgers *et al*., 2013; Rodgers *et al*., 2015; Manikkam *et al*., 2013). If the phenotype is paternally transmitted, sperm must contain the effector(s) of ETI. One possible mode of transmission would be sperm-borne sncRNAs, as they are delivered to the oocyte during fertilization (Yan, 2014; Ostermeier *et al*., 2004). Furthermore, paternal sncRNA have already been found to be associated with certain phenotypes and functions in this context, so it seems only logical that they may be effectors and transmitters of ETI (Yan, 2014; Lane *et al*., 2014).

Sperm miRNA expression can be affected by toxins and stress, among other variables
(Stowe et al., 2014; Rodgers et al., 2013; Schuster et al., 2016). The latter could be induced by microinjecting a combination of eight miRNA found differentially expressed in the sperm of stressed mice, suggesting a direct connection between sperm silencing sncRNAs and epigenetically inherited phenotypes (Rodgers et al., 2015). In addition to miRNAs, sperm-borne 5’ halves are also affected by ETI (Chen et al., 2016; Schuster et al., 2016;)

**Figure 1-3 – Epigenetic transgenerational inheritance.** When a pregnant mother is exposed to an environmental factor (e.g., vinclozolin), the child (F1) as well as the child’s germline (F2) are directly exposed in utero. Therefore, the great-grandchildren of the mother who was originally exposed (F3) is the first generation to never come into direct contact with the environmental factor. Epimutations that are transferred stably until at least the F3 generation are examples of epigenetic transgenerational inheritance. Adapted from (Nilsson and Skinner, 2015).
Phenotypes can also be inherited epigenetically as paramutations. Paramutations are epigenetic states of alleles that influence gene expression, are passed down in a non-Mendelian fashion, and alter the epigenetic status of the other allele for that gene (Pilu, 2011; Yuan et al, 2015). Examples include the Sox9, Kit, and Cdk9 paramutations, which are associated with excessive early growth, white tail tip, and cardiac hypertrophy phenotypes, respectively (Grandjean et al, 2009; Wagner et al, 2008; Yuan et al, 2015). These paramutations and their associated phenotypes could be induced via microinjection of specific, paternally-delivered RNAs (e.g., miR-124 and the Sox9 paramutation) into the zygote (Liebers et al, 2014; Rassoulzadegan and Cuzin, 2015). Furthermore, both the Sox9 and Kit paramutations could not be transmitted in the absence of Dnmt2, the tRNA methyltransferase involved in 5’ halve production (Kiani et al, 2013). It is possible that certain aberrant 5’ halves expression profiles are the basis for different paternally transmitted phenotypes.

All of these aforementioned studies suggest a meaningful physiological role for sperm-borne sncRNAs in the early embryo; however, the extent of this role is still unknown. In this dissertation, the importance of paternal miRNA and endo-siRNA in the development of the preimplantation embryo (Chapter II), characteristics of sncRNA populations in mammalian sperm (Chapter III), and a putative role for sperm-borne sncRNA in vinclozolin-induced ETI (Chapter IV) are explored. Through these studies, we hope to further our knowledge of the importance of sncRNA observed in sperm.
REFERENCES


Hosken and Hodgson. 2014. Why do sperm carry RNA? Relatedness, conflict, and


Sperber et al. 2014. miRNA sensitivity to Drosha levels correlates with pre-miRNA secondary structure. RNA, 20, 621 – 631.


Yuan et al. 2016. Sperm-borne miRNAs and endo-siRNAs are important for fertilization and preimplantation embryonic development. *Development*, 143, 635 – 647.
CHAPTER II

Sperm-borne miRNAs and endo-siRNAs are important for fertilization and preimplantation embryonic development

Shuiqiao Yuan¹, Andrew Schuster¹, Chong Tang¹, Tian Yu¹, Nicole Ortogero¹, Jianqiang Bao¹, Huili Zheng¹, Wei Yan¹.²*

¹Department of Physiology and Cell Biology, University of Nevada School of Medicine, 1664 North Virginia Street, MS0575, Reno, NV 89557, USA

This chapter can be found in the following publication:

Yuan et al. 2016. Sperm-borne miRNAs and endo-siRNAs are important for fertilization and preimplantation embryonic development. Development, 143, 635 – 647.

*Corresponding author:

Wei Yan M.D., Ph.D.
Professor
University of Nevada School of Medicine
Center for Molecular Medicine, Room 207B
1664 North Virginia Street, MS/0575
Reno, NV 89557
Tel: 775 784 7765
Fax: 775 784 4362
Email: wyan@medicine.nevada.edu
URL: http://www.medicine.nevada.edu/yan
ABSTRACT

Although it is believed that mammalian sperm carry small noncoding RNAs (sncRNAs) into oocytes during fertilization, it remains unknown whether these sperm-borne sncRNAs truly have any function during fertilization and preimplantation embryonic development. Germline-specific Dicer and Drosha conditional knockout (cKO) mice produce gametes (i.e., sperm and oocytes) partially deficient in miRNAs and/or endo-siRNAs, thus providing a unique opportunity for testing whether normal sperm (paternal) or oocyte (maternal) miRNA and endo-siRNA contents are required for fertilization and preimplantation development. Using the outcome of intracytoplasmic sperm injection (ICSI) as a readout, we found that sperm with altered miRNA and endo-siRNA profiles could fertilize wild-type (WT) eggs, but embryos derived from these partially sncRNA-deficient sperm displayed a significant reduction in developmental potential, which could be rescued by injecting WT sperm-derived total or small RNAs into ICSI embryos. Disrupted maternal transcript turnover and failure in early zygotic gene activation appeared to associate with the aberrant miRNA profiles in Dicer and Drosha cKO spermatozoa. Overall, our data support a crucial function of paternal miRNAs and/or endo-siRNAs in the control of the transcriptomic homeostasis in fertilized eggs, zygotes and two-cell embryos. Given that supplementation of sperm RNAs enhances both the developmental potential of preimplantation embryos and the live birth rate, it might represent a novel means to improve the success rate of assisted reproductive technologies in fertility clinics.
INTRODUCTION

Spermatozoa (i.e., sperm), as the male gamete, are responsible for not only delivering the paternal genome, but also for providing essential factors for fertilization and preimplantation embryonic development (Krawetz, 2005; Sone et al, 2005). Recent data also suggest that in addition to the paternal proteins and genome, sperm also contribute the paternal epigenome in the forms of specific paternal DNA methylation, retained histones and sperm-borne RNAs during fertilization and early embryonic development (Gannon et al, 2014; Jenkins and Carrell, 2012). Both coding and noncoding RNAs have been detected in sperm (Krawetz, 2005; Pessot et al, 1989) and small noncoding RNA (sncRNA) species including microRNAs (miRNAs) (Amanai et al, 2006; Yan et al, 2008), endogenous small interfering RNAs (endo-siRNAs) (Song et al, 2011), piwi-interacting RNAs (piRNAs) (Ghildiyal and Zamore, 2009; Grivna et al, 2006), and tRNA-derived small RNAs (Kawano et al, 2012; Peng et al, 2012) appear to be abundant in sperm. Similarly, oocytes also contain various species of sncRNAs (McGinnis et al, 2015; Sirotkin, 2012). Interestingly, only endo-siRNAs have been shown to be essential for oogenesis and the other types of sncRNAs appeared to be either nonfunctional or dispensable during oogenesis (Dixon et al, 2012; Hong et al, 2008; Leese et al, 2001). Zygotic expression of sncRNAs plays an essential role in early development (Ohnishi et al, 2010; Viswanathan et al, 2009; Yang et al, 2008). However, it is unclear whether sperm-borne (i.e., paternal) or oocyte-borne (i.e., maternal) sncRNAs are required for fertilization and/or pre-implantation development.

The biogenesis of miRNAs and endo-siRNAs requires two RNase III enzymes, Drosha and Dicer (also known as Dicer1) (Lee et al, 2006; Suh and Blelloch, 2011; Yang
and Lai, 2011). Drosha is responsible for the processing of miRNA primary transcripts into precursor miRNAs (pre-miRNAs) in the nucleus, whereas Dicer cleaves the pre-miRNAs into mature miRNAs in the cytoplasm. The miRNA production requires both RNase III enzymes, whereas endo-siRNA production needs only Dicer (Lee et al, 2006; Suh and Blelloch, 2011; Yang and Lai, 2011). By generating male germline-specific Dicer or Drosha conditional knockout (cKO) mice, we have demonstrated that these cKO male mice produce a small number of normal-looking sperm despite low sperm counts and reduced motility (Wu et al, 2012). Therefore, we decided to utilize these cKO male mice to investigate whether the Dicer or Drosha cKO spermatozoa are deficient in miRNAs and endo-siRNAs; if so, we will use these miRNA- and endo-siRNA-deficient sperm to perform intracytoplasmic sperm injection (ICSI) to study whether ablation of paternal sncRNAs could lead to defects in fertilization and/or preimplantation development, and whether supplementation of normal sperm-derived sncRNAs and total RNAs could rescue the defects. By generating female germ cell-specific Drosha cKO mice, we also obtained miRNA-deficient oocytes (Yuan et al, 2014), with which ICSI was performed using Dicer or Drosha cKO sperm to examine the relative importance of paternal and maternal sncRNAs. Here, we report that sperm-borne sncRNAs are indeed important for preimplantation embryonic development.
RESULTS

miRNAs and/or endo-siRNAs are partially deficient in Dicer and Drosha cKO spermatozoa

We previously generated male germ cell-specific Dicer (Stra8-iCre; Dicer<sup>lox/del</sup>, herein called Dicer cKO) and Drosha (Stra8-iCre; Drosha<sup>lox/del</sup>, herein called Drosha cKO) conditional KO mice by crossing the Stra8-iCre line with Dicer or Drosha loxP mice (Wu et al, 2012). Both Dicer and Drosha cKO males were infertile as a result of low sperm counts, low sperm motility and abnormal sperm morphology (Figure 2-1A – F), resembling oligoasthenoteratozoospermia (OAT) in humans. We collected cKO epididymal spermatozoa and analyzed Dicer and Drosha mRNA levels using real-time quantitative polymerase chain reaction (qPCR). Both Dicer and Drosha mRNA levels were reduced drastically in cKO sperm (Figure 2-1G), which is consistent with our previous data showing significantly reduced levels of Dicer and Drosha in Dicer and Drosha cKO pachytene spermatocytes and round spermatids, respectively (Wu et al, 2012). Notably, although Dicer mRNA levels were not affected in Drosha cKO sperm, Drosha mRNA levels were significantly reduced in Dicer cKO sperm compared with those in WT sperm (Figure 2-1G).

To determine whether Dicer and Drosha cKO sperm are deficient in sncRNAs, we performed sncRNA deep sequencing (sncRNA-Seq) (Figure 2-2). A total of 379 miRNAs were identified in the sperm of WT, Dicer cKO and Drosha cKO mice (Figure 2-2A and Tables 2-S1 – S8). In Dicer cKO sperm, ~47% of miRNAs were dysregulated, including ~15% upregulated and 32% downregulated (Figure 2-2A, -2B and Tables 2-S1 – S3). Among the downregulated miRNAs, 10% were downregulated by 2- to 10-fold (Figure
2-2A, -2B and Table 2-S2), and 22% by >10-fold, which were defined as absent (Figure 2-2A, -2B and Table 2-S3). Interestingly, ∼53% were unchanged (Figure 2-2B and Table 2-S4). Similarly, in Drosha cKO sperm, ∼52% of miRNAs were dysregulated, with ∼19% upregulated and 33% downregulated, compared with WT controls (Figure 2-2A, -2B and Tables 2-S5 – S7). Among the downregulated miRNAs, 14% were downregulated by 2- to 10-fold, whereas 19% were absent (down by >10-fold) (Figure 2-2A, -2B and Tables 2-S6, -S7). Approximately 48% of miRNAs were not altered in Drosha cKO sperm (Table 2-S8).

Due to a lack of endo-siRNA annotation in murine sperm, we performed a search for novel endo-siRNAs in silico using our previous method (Schuster et al, 2015), and identified 711 putative novel endo-siRNAs (Table 2-S9). 65 out of the 711 endo-siRNAs identified were abundantly expressed (>2 copies) in at least one of the sperm samples from WT, Dicer cKO and Drosha cKO mice. In Dicer cKO sperm, 4 of 65 (accounting for ∼6%) endo-siRNAs were significantly downregulated (by >10 fold) and only 1 (accounting for ∼2%) was upregulated (Figure 2-2C, -2D and Table 2-S10). As expected, in Drosha cKO sperm, none of the 65 endo-siRNAs showed any significant changes when compared with the WT controls (Figure 2-2C, -2D and Table 2-S10). miRNA profiles were altered in both Dicer and Drosha cKO sperm, whereas the endo-siRNA profiles were affected only in Dicer cKO sperm, further supporting the current concept that endo-siRNAs are Dicer dependent and Drosha independent (Kim et al, 2009; Wu et al, 2012). Overall, the Dicer or Drosha cKO spermatozoa were not completely devoid of, but rather partially deficient in miRNAs and/or endo-siRNAs.
Eggs fertilized by spermatozoa with aberrant miRNA and endo-siRNA profiles through ICSI display reduced preimplantation developmental potential

A partial depletion of miRNAs and endo-siRNAs in Dicer or Drosha cKO sperm is, in fact, advantageous to our purpose because it allows mature spermatozoa to be produced despite abnormal miRNA and endo-siRNA profiles in these mutant sperm. To test whether sperm with altered sperm miRNA and endo-siRNA contents could support fertilization and early embryonic development, we injected the control (WT) and cKO sperm into WT oocytes (C57BL/6J), and evaluated the developmental potential by counting the number of embryos that reached each of the five developmental stages, including two pronuclei (2PN), 2-cell, 4-cell, morula and blastocyst. Dicer cKO sperm-derived ICSI embryos displayed significantly reduced developmental potential in all five stages evaluated, compared with WT controls (Table 2-1 and Figure 2-3). By contrast, Drosha cKO sperm-derived ICSI embryos only showed a significant reduction in developmental potential starting at the 4-cell stage and thereafter (Table 2-2 and Figure 2-4). In general, Drosha cKO spermatozoa appeared to be more competent than the Dicer cKO spermatozoa in supporting fertilization and preimplantation embryonic development, based on the fact that close-to-normal fertilization rate and zygote to 2-cell transitions were observed in Drosha cKO sperm-fertilized eggs (Figure 2-4A, -4B), but not in the Dicer cKO sperm-derived eggs (Figure 2-3A, -3B).

Supplementation of WT sperm RNA enhances preimplantation development of ICSI embryos derived from spermatozoa with aberrant miRNA or endo-siRNA profiles

The compromised developmental potential of embryos derived from ICSI using Dicer and Drosha cKO sperm might not necessarily reflect the deficiency in snRNA
contents because these spermatozoa were mostly defective as a result of disrupted spermiogenesis. Although we injected only sperm heads with normal-looking morphology, potential structural defects in the injected sperm heads might have also contributed to the reduced developmental potential observed in the ICSI embryos. If the compromised developmental potential were truly caused by the loss of sperm-borne miRNAs/endo-siRNAs, supplementation of normal paternal RNA contents would enhance the developmental potential. Otherwise, the reduced developmental potential might well have been due to other structural or functional defects of the sperm head, which have nothing to do with sperm-borne sncRNAs. Therefore, we injected total, or small RNAs, isolated from WT sperm into eggs fertilized by Dicer and Drosha cKO spermatozoa through ICSI. Injection of the WT sperm small RNA fractions into ICSI eggs fertilized by Dicer cKO spermatozoa slightly improved the embryonic developmental potential at the 4-cell, morula and blastocyst stages, although the overall developmental potential remained lower than that of the WT controls (Table 2-1 and Figure 2-3). Interestingly, when WT sperm total RNA was injected into the ICSI eggs fertilized by Dicer cKO spermatozoa, ~50% of ICSI-derived 2-cell embryos developed into blastocysts. This rate is similar to that of WT controls, suggesting a full rescue can be achieved when Dicer cKO sperm-fertilized eggs are supplemented with WT sperm-borne total RNA (Figure 2-3E).

Using a similar strategy, we studied the effects of injecting WT sperm total RNAs, or small RNAs only, into eggs fertilized by Drosha cKO spermatozoa through ICSI (Table 2-2 and Figure 2-4). Supplementation of WT sperm small RNAs significantly increased the developmental potential from 2-cell to blastocyst stages from
∼26% to ∼41%, which is close to the developmental rates in WT controls (∼50%) (Figure 2-4E). Although at a lower efficiency, supplementation of WT sperm total RNAs also enhanced the developmental potential of Drosha cKO sperm-fertilized ICSI embryos from ∼26% to ∼34% (Table 2-2).

Four types of controls were included to determine whether the rescue truly resulted from sperm-borne RNAs. The first was a vehicle control that monitored the potential inhibitory or activating factors contained in the solution used for diluting sperm RNAs. The second control was sperm RNA treated with RNase A prior to injection, which would indicate whether the effects observed were RNA dependent. The third was a somatic cell RNA control, which contained total or small RNAs isolated from the heart. This control would tell whether the effects observed were sperm RNA specific. The fourth was a control for parthenogenesis. The much-reduced developmental potential in the vehicle control suggests potential damage to developing embryos when injected with vehicle only (Tables 2-1, -2 and Figures 2-3, -4). A lack of rescuing effects in the second and the third controls (i.e., ‘sperm RNA+RNase’ and somatic cell RNA) indicate that the improved developmental potential observed in supplementation with WT sperm total or small RNAs was both RNA dependent and sperm RNA specific. Overall, a lack of rescue in all of the four types of controls suggests that the rescue effects observed were sperm-borne RNA specific and RNA dependent.

**Supplementation of WT sperm RNAs significantly improves the birth rate of ICSI embryos derived from spermatozoa with aberrant miRNA and endo-siRNA profiles**

To evaluate the post-implantation development of the embryos derived from partial miRNA- or endo-siRNA-deficient spermatozoa supplemented with WT sperm-
borne RNA, we transferred 2-cell embryos into the oviducts of recipient mice, and allowed them to develop to term. In the WT control group, ~28% of transferred 2-cell embryos led to live-born pups, whereas only 4-8% of transferred 2-cell embryos from ICSI using Dicer cKO or Drosha cKO sperm developed to term with live-born pups (Table 2-3 and Figure 2-5), suggesting a decreased post-implantation developmental potential in embryos derived from ICSI using partial miRNA- or endo-siRNA-deficient spermatozoa. By contrast, the birth rates were almost doubled (from 8.5% to 14.8%) and tripled (from 4.3% to 12.7%), for transferred 2-cell embryos derived from ICSI using Dicer cKO spermatozoa followed by supplementation of WT sperm total RNA and by Drosha cKO spermatozoa with subsequent supplementation of WT sperm small RNAs, respectively (Table 2-3 and Figure 2-5). Such a significant improvement suggests that early supplementation of normal sperm RNA can drastically enhance not only the preimplantation development, but also the post-implantation development and the birth rate of ICSI embryos derived from sperm partially deficient in miRNAs and endo-siRNAs in mice. Both the male and female offspring, derived from ICSI using both Dicer and Drosha cKO spermatozoa with or without supplementation of WT sperm RNAs, developed normally with normal fertility when they reached adulthood.

Maternal miRNAs are dispensable for both fertilization and preimplantation development

It has been demonstrated that miRNA are dispensable for oocyte maturation, and miRNA-deficient oocytes from Zp3-Cre; Dgcr8<sup>lox/lox</sup> or Zp3; Dros<sup>h</sup>a<sup>lox/lox</sup> females are fertile (Ma et al, 2010; Suh et al, 2010; Yuan et al, 2014). However, endo-siRNAs appear to be essential for oocyte maturation because Dicer cKO oocytes display spindle defects
and are infertile, whereas Dgcr8 cKO or Drosha cKO oocytes are completely normal and fertile (Murchison et al, 2007; Suh et al, 2010; Yuan et al, 2014). To investigate whether a lack of both maternal and paternal miRNAs and/or endo-siRNAs would have additive adverse effects on fertilization and preimplantation development, we injected Dicer and Drosha cKO spermatozoa into Drosha cKO oocytes and evaluated the developmental potential (Table 2-4 and Figure 2-6).

No differences in developmental potential were observed between embryos derived from WT sperm injected into Drosha cKO versus WT oocytes (Tables 2-1, -2 and 4). However, we observed reductions in the fertilization rate and early developmental potential between embryos derived from ICSI using Dicer cKO sperm injected into WT and Drosha cKO oocytes (Tables 2-1 and -4). These results suggest that the absence of maternal miRNAs has little or no effect on fertilization and preimplantation development, and a lack of both maternal and paternal miRNAs does not worsen the preimplantation development. However, the embryos derived from ICSI using Drosha cKO spermatozoa and Drosha cKO oocytes displayed reduced developmental potential in all five stages observed. This drastically reduced preimplantation developmental potential might reflect an absolute requirement for the zygotic or embryonic Drosha expression during preimplantation development because it has been documented that Drosha-null embryos die at ~E6.5 (Yuan et al, 2014). Supplementation of WT sperm RNAs, either total or small, failed to rescue the defective developmental potential (Table 2-S11), supporting the notion that the defects are derived from the oocytes rather than Drosha cKO spermatozoa.
**Paternal miRNAs are indeed delivered to the oocytes during fertilization and can persist beyond the 2PN stage**

To determine whether sperm-borne miRNAs are indeed delivered into oocyte and persist in early embryos, we performed snRNA-Seq using WT oocytes and 2PN embryos derived from ICSI using WT, Dicer cKO and Drosha cKO spermatozoa (Figure 2-7). A total of 82 and 85 miRNAs (>3 reads per million) were identified in WT oocytes and WT 2PN embryos, respectively (Figure 2-7 and Table 2-S12A, -S12B); 36 and 86 miRNAs were identified in Dicer cKO and Drosha cKO sperm-derived 2PN embryos (Table 2-S12A, -S12B). Comparative analyses of the miRNAs profiles among WT sperm, WT oocytes and 2PN embryos derived from ICSI using WT, Dicer cKO and Drosha cKO spermatozoa, identified 14 miRNAs that were present in WT 2PN embryos and sperm, but not in WT oocytes (Figure 2-7 and Table 2-S12C). Among the 14 miRNAs, 5 were present in the Drosha cKO 2PN embryo and none was found in the Dicer cKO 2PN embryo (Figure 2-7 and Table 2-S12C). These data prove that sperm-borne miRNAs can be delivered to the oocytes during fertilization and many can persist beyond the 2PN stage.

**Embryos derived from ICSI using Dicer or Drosha cKO spermatozoa display aberrant gene expression profiles during preimplantation development**

To further explore the underlying molecular mechanism, we analyzed the expression profiles of 96 genes known to be essential for preimplantation development (Guo et al, 2010; Levy, 2001; Messerschmidt et al, 2014; Zheng and Dean, 2007). Using the latest microfluidics-based, high-throughout qPCR system (Fluidigm BioMark HD), we simultaneously analyzed the expression levels of the 96 genes in single WT oocytes, and embryos at 2PN, 2-cell, 4-cell, morula and blastocyst stages derived from ICSI using
WT, Dicer and Drosha cKO spermatozoa. Dysregulation of genes was observed at all stages (Figure 2-8 and Table 2-S13). Although levels of the maternal transcripts were reduced from oocytes to the 2PN stage in WT controls, a similar decrease was observed in Drosha but not Dicer cKO sperm-derived 2PN embryos (Figure 2-8, genes framed in blue). At the 2-cell stage, five genes (Nes, Nodal, Vim, Cdkn2d and Stat5b) were activated and highly expressed in WT embryos, whereas these genes failed to be activated in 2-cell embryos derived from either Dicer or Drosha cKO sperm (Figure 2-8, framed genes). Although numerous genes were dysregulated (up- or down-regulated) in subsequent stages, this most likely reflects secondary effects of the initial disruptions in 2PN and 2-cell embryos, caused by deficiencies in paternal miRNAs and/or endo-siRNAs. In general, the expression profiles of Drosha cKO sperm-derived embryos appeared to be closer to those of the WT controls than the Dicer cKO sperm-derived embryos. This is consistent with the overall better developmental potential of Drosha cKO sperm-derived embryos, compared with Dicer cKO sperm-derived ones.

The significant dysregulation of genes at 2PN and 2-cell stages might represent the primary effects of deficiencies in paternal miRNAs and/or endo-siRNAs. Therefore, we further examined whether those dysregulated genes were targets of those miRNAs or endo-siRNAs that are deficient in Dicer or Drosha cKO spermatozoa. In Dicer cKO sperm-derived 2PN and 2-cell embryos, ~64% (25 out of 39) of dysregulated genes all had their targeting miRNAs significantly (>2-fold) dysregulated in Dicer cKO sperm (Figure 2-9A and Table 2-S14). In Drosha cKO sperm-derived 2PN and 2-cell embryos, ~52% (17 of 33) of dysregulated genes could be targeted by miRNAs that were
significantly dysregulated in Drosha cKO sperm (Figure 2-9B and Table 2-S14). For example, Neurog3 is drastically upregulated, whereas Vim is significantly downregulated in both Dicer and Drosha cKO sperm-derived 2PN and 2-cell embryos; miRNAs targeting Neurog3 (miR-17/17-5p/20ab/20b-5p/93/106ab/427/518a-3p/519d and miR-124/124ab/506) were also significantly dysregulated in Dicer and Drosha cKO sperm (Figure 2-9 and Tables 2-S14, -S15). Similarly, miRNAs that target Vim (miR-124/124ab/506, miR-138/138ab and miR-320abcd/4429) were also dysregulated in Dicer and Drosha cKO sperm (Figure 2-9 and Tables 2-S14, -S15). Neurog3 encodes a protein belonging to a basic helix-loop-helix (bHLH) transcription factor involved in neurogenesis and spermatogenesis (Hong et al, 2008; Stewart and Behringer, 2012). Vim encodes a class III intermediate filament protein widely expressed in the developing embryo and in cells of mesenchymal origin in the adult (Graw, 1996). Premature activation of Neurog3 and failure in activation of Vim reflect the disrupted early developmental program in Dicer and Drosha cKO sperm-derived embryos.

Taken together, these data suggest that the normal paternal miRNA and endo-siRNA profiles might have effects on the proper turnover of maternal transcripts, as well as the timely initiation of the early transcriptional program during fertilization and zygote to 2-cell transition. Moreover, the effects of the paternal miRNAs are most likely mediated through post-transcriptional regulation of the maternal and early zygotic mRNAs, for example, miRNA-mediated mRNA stability control.
Expression patterns of H3K4me2, H3K4me3 and H3K9me3 are normal in 2PN and 2-cell embryos derived from Dicer and Drosha cKO sperm

As an essential epigenetic mechanism, histone modifications regulate gene expression during development, for example, methylation of histone H3 at different lysine residues can either activate (H3K4me, H3K36me, H3K79me) or repress (H3K9, H3K27, H4K20) transcription (Jenuwein and Allis, 2001; Talbert and Henikoff, 2010). Recent reports have shown that noncoding RNAs (ncRNAs) are involved in epigenetic regulation of gene expression either through post-transcriptional regulation of important epigenetic regulators, including DNA methyltransferases, histone methyltransferases and deacetylases (Sato et al, 2011) or by serving as a ‘sequence guide’ and thus directing chromatin modifying machineries to the correct loci for epigenetic modifications (Maruyama et al, 2012). To explore whether an initial lack of sperm-borne snRNAs could affect histone modifications during the 2PN to 2-cell transition, we examined the expression patterns of H3K4me2, H3K4me3 and H3K9me3 in 2PN and 2-cell embryos derived from ICSI using Dicer and Drosha cKO spermatozoa. At the 2PN stage, H3K4me2 was expressed in both male and female pronuclei in all four types of embryos (WT, Dicer cKO sperm derived, Drosha cKO sperm derived and Drosha null). Both H3K4me3 and H3K9me3 were detected only in female pronuclei of the 2PN embryos of all four types (Figure 2-S1). At the 2-cell stage, H3K4me2, H3K4me3 and H3K9me3 were detected in the nuclei of all four types of embryos (Figure 2-S1). No discernable differences were observed among all four types of embryos. Given the limited resolution of the immunofluorescence method used, one cannot preclude epigenetic changes in specific chromatin regions despite the lack any discernable changes in the global histone
methylation patterns. Nevertheless, our data suggest that the altered paternal miRNA/endo-siRNA contents do not have a significant impact on the three specific types of histone modifications in 2PN and 2-cell embryos.

**DISCUSSION**

The incomplete depletion of miRNAs and endo-siRNAs in Dicer and Drosha cKO sperm is consistent with the persistent expression of Dicer and Drosha mRNAs in spermatogenic cells even after Cre-mediated gene deletion (Wu et al, 2012). Several possibilities exist, which might explain why around half of the total sperm-borne miRNAs remain unchanged and a small proportion even get upregulated in cKO sperm. First, although the Dicer or Drosha gene was inactivated by late pachytene or round spermatid stages, their transcripts might have been synthesized, then stabilized and stored in ribonucleoproteins, a mechanism well documented for numerous transcripts that are needed for the haploid phase of spermatogenesis (Iguchi et al, 2006). The stored transcripts can then be translated into proteins even in the absence of the gene in haploid germ cells. Similarly, pre-miRNAs may be synthesized and stored in ribonucleoproteins prior to the Cre-mediated gene inactivation in late pachytene spermatocytes and round spermatids. These pre-miRNAs could then be processed in spermatids and eventually packed into sperm. Second, those unchanged or upregulated miRNAs or endo-siRNAs might be more stable and thus would exist for an extended period of time and eventually could be packaged into sperm. The stabilization of miRNAs can be achieved through binding to circular RNAs or mRNAs (Aravin et al, 2007; Peng and Lin, 2013). Third, the cytoplasm of pachytene spermatocytes and developing spermatids are interconnected
through intercellular bridges (IBs) (Haglund et al, 2011; Hermo et al, 2010). IBs allow for sharing of the cytoplasmic contents, including organelles and mRNAs (Ventela et al, 2003). Therefore, if the Dicer or Drosha gene is not deleted in one cell, the transcripts synthesized by this cell would be able to cross the IBs and reach its neighboring cells. Lastly, those miRNAs and endo-siRNAs might be produced through the non-canonical pathway (Pek et al, 2012), which does not require Dicer or Drosha. Nevertheless, with reduced levels of Dicer or Drosha transcripts, although some spermatids manage to complete spermiogenesis, the spermatozoa in the cKO testes or epididymides are low in number, largely deformed and do not display normal motility, resembling human oligo-astheno-teratozoospermia (OAT). On the basis of testicular histology, sperm morphology, sperm counts and sperm motility, the phenotype of our Dicer and Drosha cKO mice appears to be less severe than that of Dicer1 and Dgcr8 cKO mice, as reported previously (Zimmermann et al, 2014). The discrepancy might well result from different Cre deleter lines (our Stra8-Cre versus their Ddx4-Cre) and different mouse strains used in these two studies. The OAT phenotype implies that the cKO spermatozoa are structurally and functionally compromised, which explains why those cKO males are infertile through natural mating. Because our goal was to show the effects of altered sperm miRNA and endo-siRNA contents on fertilization and early embryonic development, we purposely chose those normal-looking sperm heads for injection for the following reasons: (1) abnormalities in other parts of the sperm (e.g., flagellum) can be ignored; (2) those normal-looking sperm heads should have fewer structural defects, making them closer to the control sperm heads except for altered miRNA and endo-siRNA profiles. Further supporting this notion, the almost complete rescue by the injection of WT sperm total
RNA into Dicer cKO sperm-fertilized eggs implies that the defects largely lie in the RNA contents in Dicer cKO sperm and the rescuing effects are RNA dependent.

Differential effects were observed between supplementation experiments using total versus small sperm RNA fractions, for example, WT sperm total RNA appears to be more efficient than small RNAs in rescuing developmental defects in ICSI embryos derived from Dicer cKO sperm although total RNAs contains small RNAs. This discrepancy can be explained by the fact that total RNAs could contain factors such as large noncoding RNAs and/or mRNAs that are absent in small RNA preparations, which might have contributed the rescuing effects observed. Alternatively, in the total RNAs, small RNAs are diluted and thus, more optimal for the rescue. By contrast, sperm small RNAs seem to be more efficient than sperm total RNAs in rescuing the developmental defects of eggs fertilized with Drosha cKO sperm. Since we injected the same volume of total RNAs or small RNAs at the same concentrations into eggs, the difference in efficacy might result from the dilution of small RNAs in the total RNA samples. Our data suggest that WT sperm total RNAs can rescue the reduced developmental potential in Dicer cKO sperm-derived embryos, and WT sperm small RNAs can enhance the developmental potential of Drosha cKO sperm-derived embryos. These data also suggest that that Dicer cKO sperm have a more profound RNA deficiency (e.g., defective in both large and small RNA species), whereas Drosha cKO sperm are mainly deficient in snRNA. This is also supported by the fact that Drosha cKO sperm, in general, perform better than Dicer cKO sperm in supporting fertilization and preimplantation development prior to the 4-cell stage. Overall, our data demonstrate that sperm RNA supplementation enhances both preimplantation development and birth rate when snRNA-deficient sperm
(e.g., Dicer or Drosha cKO sperm) are used for ICSI. It would be interesting to study whether such a strategy is equally efficient in ICSI using sperm derived from other infertile mice with OAT. If this method can universally enhance the ICSI outcome, it might be considered for future application in assisted reproductive technology (ART) clinics to enhance success rates.

If paternal miRNAs or endo-siRNAs are truly functional, they must act mainly prior to, or soon after, zygotic genome activation, which occurs mainly at late zygote and 2-cell stages (Yuan et al, 2015). Although our data suggest a role of bulk paternal sncRNAs in post-fertilization development, it remains unclear how individual miRNAs and endo-siRNAs function at the molecular levels. A previous study reports that a paternal miRNA, miR-34c, is essential for the first cleavage, based on injection of miR-34c inhibitor into zygotes (Liu et al, 2012); however, this claim is not supported by the fact that Mir34c-null male mice are completely fertile (Wu et al, 2014). In fact, miR-34c belongs to a family of six miRNAs (miR-34a/b/c and miR-449a/b/c) encoded by three miRNA clusters (Mir34a, Mir34b/c and Mir449), which all contain the same seed sequence for 3’UTR recognition and thus are all functionally redundant. Inactivation of one of the three miRNA clusters does not cause any phenotype during either development or in adulthood (Wu et al, 2014). Nevertheless, our profiling analyses on maternal and early zygotic genes suggest that the paternal- or sperm-borne sncRNAs might have a role in regulating proper maternal mRNA turnover during zygote to 2-cell transition. Disruptions in both maternal transcript turnover and zygotic gene activation would impact, in theory, all early events, consequently leading to an arrest mostly at the 2-cell stage. It remains an interesting future topic to investigate how paternal transcripts lead to
degradation of maternal transcripts and how proper maternal transcript turnover affects zygotic gene activation. The fact that the majority of the dysregulated genes in 2PN and 2-cell embryos derived from ICSI using Dicer or Drosha cKO sperm are direct targets of miRNAs deficient or dysregulated in Dicer or Drosha cKO sperm strongly suggests that the paternal miRNAs act on their target mRNAs, which are mostly maternal transcripts, in the 2PN and 2-cell embryos. Many of the paternal miRNAs are also present in oocytes as the maternal miRNAs (Dixon et al., 2011; Hong et al., 2008). It remains puzzling that paternal miRNAs, rather than the same sets of maternal miRNAs, can affect maternal and early zygotic transcripts. However, this is consistent with the earlier findings, showing that maternal miRNAs appear to be functionally suppressed or non-functional during oocyte maturation and fertilization (Dixon et al., 2012; Hong et al., 2008; Leese et al., 2001). The underlying physiological significance of this phenomenon remains an interesting topic for future investigation.

In summary, we demonstrate that aberrant sperm-borne miRNA and endo-siRNA profiles correlate with reduced preimplantation developmental potential, which can be rescued by supplementation of wild-type sperm total RNAs or small RNAs in mice. Our data suggest that paternal miRNAs and endo-siRNAs are important for initiating the normal developmental program during early preimplantation development, especially from fertilization to the 2PN to 2-cell transition.
ACKNOWLEDGEMENTS

W.Y. and S.Y. conceived and designed the study; S.Y., A.S., C.T., T.Y., N.O., J.B. and H.Z. performed the experiments; W.Y. and S.Y. wrote the manuscript. This work was supported by the National Institutes of Health (NIH) [HD060858, HD071736 and HD074573 to W.Y.]. All knockout mouse lines were generated and maintained at the University of Nevada Genetic Engineering Center (UNGEC) supported, in part, by an NIH COBRE grant [1P30GM110767]. The authors declare no competing or financial interests.
MATERIALS AND METHODS

Reagents and media

All reagents used were purchased from Sigma unless otherwise stated. The modified CZB-HEPES medium containing 20 mM HEPES-Na, 5 mM NaHCO3, and 0.1 mg/ml polyvinyl alcohol (cold water soluble) was used for collecting sperm or oocytes. The CZB medium, supplemented with 5.56 mM D-glucose and 4 mg/ml BSA (Fraction V, Calbiochem), was used for culturing oocytes before ICSI, as previously described (Chatot et al, 1990; Kimura and Yanagimachi, 1995; Yanagimachi et al, 2004). The medium used for culturing fertilized embryos after ICSI was EmbryoMax KSOM medium supplemented with amino acids (KSOM+AA; Millipore, MR-121-D).

Generation of postnatal germ cell-specific Dicer or Drosha cKO mice

The Institutional Animal Care and Use Committee (IACUC) of the University of Nevada, Reno, approved all animal work in this study. The Cre-loxP strategy was used to generate the germline conditional knockout mice. Stra8-iCre;Dicer<sup>lox/del</sup> (called Dicer cKO), Stra8-iCre;Drosha<sup>lox/del</sup> (called Drosha cKO) male mice and Zp3-iCre;Drosha<sup>lox/del</sup> female mice were generated as our previously reported (Wu et al, 2012; Yuan et al, 2014).
Oocyte preparation and ICSI

WT and Zp3-icre;Drosha\textsuperscript{lox/del} female mice at 5-10 weeks of age were superovulated by intraperitoneal injection of 5 IU of pregnant mare's serum gonadotropin (PMSG), followed by intraperitoneal injection of 5 IU of human chorionic gonadotropin (hCG) 48 h later. Mature oocytes (MII stage) were collected from the oviducts 14-16 h after hCG injection, and freed from cumulus cells by treatment with 0.1% bovine testicular hyaluronidase in HEPES-CZB at 37°C for 2-3 min. The cumulus-free oocytes were washed and kept in the CZB medium for at least 1 h in an incubator at 37°C with air containing 5% CO2 before ICSI. An extended ICSI protocol can be found in the supplementary Materials and Methods.

Total and small RNA isolation from mouse sperm

The total RNA was isolated using the mirVana miRNA Isolation Kit (Life Technologies) following the manufacturer's instructions with modifications at the lysis stage. In brief, after addition of lysis buffer, the frozen sperm pellets were homogenized at low settings for 90 s, followed by incubation for 5 min at 65°C. Complete lysis of sperm heads was verified by microscopic examination. Once a total lysis of sperm heads was achieved, the samples were then placed on ice and the default protocol was resumed. To determine the quantity and quality, sperm RNA samples were analyzed using the RNA 6000 Nano chips run on an Agilent 2100 Bioanalyzer (Agilent). An RNA integrity number (RIN) of 2-4 indicates good sperm RNA quality.
Injection of sperm total and small RNA into post-ICSI oocytes

The concentration of sperm total or small RNAs was adjusted to 20-100 pg/µl and an aliquot of 1-2 µl of the RNA solution was loaded into a microinjection needle (Eppendorf, 930000043). Zygote microinjection was performed in HEPES-CZB medium ∼2 h after ICSI following standard procedures (Nagy et al, 2003). Each zygote was injected with ∼1-2 pl small or total RNA solution and a successful injection was verified by a visible, minor expansion of the cytoplasmic membrane. To cause RNA degradation, sperm total RNAs were treated with RNase A (Invitrogen, 12091-021) at a molar ratio of 1:10 (sperm RNA:RNase A) at 37°C for 30 min. The final concentration of RNase A in the sperm total RNA degradation products was ∼5-10 pg/µl; given that the injection volume was ∼1-2 pl, ∼5-10×10−6 pg RNase A might have been injected into a zygote. Injected zygotes were then transferred into pre-balanced KSOM+AA medium and cultured in an incubator with air containing 5% CO2 at 37°C.

Evaluation of developmental potential of ICSI-derived preimplantation embryos

ICSI and sperm RNA-supplemented ICSI oocytes were allowed to develop in KSOM+AA medium in an incubator with air containing 5% CO2 at 37°C. The number of 2PN stage embryos was counted 6-8 h after ICSI or 3-4 h after RNA supplementation. Subsequently, the numbers of embryos at 2-cell, 4-cell, morula and blastocyst stages were examined at 24 h, 48 h, 72 h and 96 h, respectively. To evaluate the potential of post-implantation development, two-cell embryos were transferred into the oviducts of
pseudopregnant CD1 females. Cesarean section was performed on day 19 after embryo transfer and live-born pups were counted.

**High-throughput single-cell qPCR**

MII oocytes were collected from the oviducts 14-16 h after hCG injection and treated with hyaluronidase to remove cumulus cells. Single embryos were collected at 6-8 h (2PN stage), 24 h (2-cell stage), 48 h (4-cell stage), 72 h (morula stage), 96 h (blastocyst) after ICSI using WT, Dicer or Drosha cKO sperm. An extended high-throughput qPCR method can be found in the supplementary Materials and Methods.

**Quantitative real-time PCR (qPCR)**

WT, Dicer cKO and Drosha cKO sperm total RNA was subjected to DNA removal using a DNA-free DNase (Ambion), followed by cDNA synthesis using the SuperScript III First-strand Kit (Invitrogen) according to the manufacturer's instructions. cDNA concentrations were measured using a NanoDrop 2000 spectrophotometer, and then diluted to a concentration of 25 ng/μl to serve as cDNA templates. SYBR Green-based real-time quantitative PCR was performed to examine mRNA expression levels. Gapdh was used as an internal control for data normalization.
**Immunohistochemistry**

Embryos were fixed in 4% paraformaldehyde in HEPES-CZB medium for 1 h and washed three times in 0.1 M glycine with 0.3 mg/ml BSA at room temperature (RT). The embryos were permeabilized in 0.2% Triton X-100 in PBS for 15 min and blocked using a blocking solution containing 2% BSA in PBS for 1 h at RT, followed by incubation with the following antibodies diluted in the blocking solution for 1 h at RT: rabbit anti-H3K4me2 polyclonal antibody (Millipore, 07-030, 1:300 dilution), rabbit anti-H3K4me3 polyclonal antibody (Diagenode, pAb-003-050, 1:300) and rabbit anti-H3K9me3 (AbCam, ab8898, 1:500). After washing in the blocking solution, embryos were incubated with fluorescence-conjugated, species-specific secondary antibodies [Alexa Fluor 594 goat anti-rabbit IgG(H+L); Molecular Probes, A11012, 1:2000] for 1 h at RT. Finally, the embryos were counterstained with 4′,6-diamidino-2-phenylindole dilacatate (DAPI, Sigma) for indirect immunofluorescent assays using a fluorescence microscope (Zeiss, HAL100).

**Small noncoding RNA deep sequencing (sncRNA-Seq) and data analyses**

Sperm sncRNA libraries were prepared using the standard protocol of an Ion Total RNA-Seq Kit v2 (Invitrogen) and biological triplicates for each sample type (WT, Dicer cKO, Drosha cKO) were prepared. For low-input sncRNA-Seq, 20-30 WT oocytes or 2PN embryos derived from ICSI using WT, Dicer cKO and Drosha cKO spermatozoa (biological duplicates for each pooled sample) were collected followed by lysis in a buffer containing 1% Salkosyl, 20 mM Tris-HCl pH 8.0, 20 mM KCl, 100 mM DTT. The
extended sncRNA-seq and data analysis processing for sperm, oocyte and 2PN embryos can be found in supplementary Materials and methods. Data have been deposited into the Gene Expression Omnibus (accession number GSE73824).

**Bioinformatic analysis**

miRNA families predicted to target the 96 early genes examined were identified using TargetScan (Berezikov et al, 2005; Lewis et al, 2005). A detailed protocol can be found in the supplementary Materials and methods.

**Statistics**

Data are presented as means±s.e.m. and statistical differences between datasets were assessed by one-way ANOVA or Student's t-test using SPSS16.0 software. P≤0.05 was considered significant and P≤0.01 was considered highly significant. ICSI data were analyzed using χ2 tests, compared with the WT group and P≤0.05 was regarded as significant.
**FIGURES**

![Figure 2-1](image)

**Figure 2-1** – Inactivation of Dicer and Drosha in spermatogenetic cells leads to oligoasthenoteratozoospermia. (A) Gross morphology of adult WT, Dicer cKO (*Stra8-Cre; Dicer<sup>lox/del</sup>*) and Drosha cKO (*Stra8-Cre; Drosha<sup>lox/del</sup>*) testes and epididymides. (B) Testis weight of WT, Dicer cKO and Drosha cKO mice. Sperm counts (C) and motility (D) of human tubal fluid (HTF)-activated epididymal spermatozoa collected from WT, Dicer cKO and Drosha cKO mice. (E) Phase-contrast microscopic images showing epididymal spermatozoa with normal and abnormal morphology in Dicer and Drosha cKO male mice. Scale bars: 25 μm. (F) Quantitative analyses of epididymal spermatozoa with normal and abnormal morphology in WT, Dicer and Drosha cKO male mice. (G) qPCR
analyses of *Dicer, Drosha* and *Spem1* mRNA levels in WT, Dicer cKO and Drosha cKO sperm. All data are based on analyses using biological triplicates and are presented as means±s.e.m. *P<0.05; **P<0.01 (vs WT controls).

**Figure 2-2** – Dicer and Drosha cKO sperm display altered miRNA and endo-siRNA expression profiles, as revealed by snRNA-Seq analyses. (A) Heat map showing a total of 379 miRNAs identified in WT, Dicer cKO and Drosha cKO sperm. (B) Pie charts illustrating the proportions of changed (≥2-fold) and unchanged (<2-fold) miRNAs in
Dicer (left) or Drosha (right) cKO sperm. Adjusted P-values ($P_{adj}$)<0.05 were considered to be significantly up- or down-regulated. (C) Heat map showing 65 known endo-siRNAs identified in WT, Dicer cKO and Drosha cKO sperm. (D) Pie charts illustrating proportions of the changed ($\geq$2-fold) and unchanged (<2-fold) endo-siRNAs in Dicer (left panel) or Drosha (right panel) cKO sperm. Adjusted P-values ($P_{adj}$)<0.05 were considered to be significantly up- or down-regulated. The original sncRNA-Seq data can be found in Tables 2-S1 – S8 and -S10.

**Figure 2-3** – Histogram showing the rate of development of embryos derived by ICSI using WT oocytes and Dicer cKO sperm with or without sperm RNA supplementation. (A) 2PN, (B) 2-cell, (C) 4-cell, (D) morula and (E) blastocyst embryos. Values labeled with different letters (a, b or c) are significantly different, based on $\chi^2$ test ($P<0.05$).
**Figure 2-4** – Histogram showing the rate of development of embryos derived by ICSI using WT oocytes and Drosha cKO sperm with or without sperm RNA supplementation. (A) 2PN, (B) 2-cell, (C) 4-cell, (D) morula and (E) blastocyst embryos. Values labeled with different letters (a-d) are significantly different, based on χ² test (P<0.05).
**Figure 2-5** – Histogram showing the rate of live-born pups derived from ICSI using WT oocytes and WT, Dicer cKO, or Drosha cKO sperm with or without supplementation of sperm-borne RNAs. Bars labeled with different letters (a-c) are significantly different, based on χ² test (P<0.05).
Figure 2-6 – Histogram showing the rate of development of 2PN embryos derived from ICSI using Drosha cKO oocytes and WT, Dicer cKO and Drosha cKO sperm. (A) 2PN, (B) 2-cell, (C) 4-cell, (D) morula and (E) blastocyst embryos. Values labeled with different letters (a, b, or c) are significantly different, based on χ² test (P<0.05).

Figure 2-7 – Heat map showing miRNA profiles in WT oocytes, WT sperm and 2PN embryos derived from ICSI using WT, Dicer cKO and Drosha cKO sperm, as determined by low-input sncRNA-Seq. The expression levels were normalized against those in WT oocytes. Biological duplicates were analyzed for each group. The original sncRNA-Seq data are shown in Table 2-S12.
Figure 2-8 – Heat map showing expression profiles of 96 genes in single oocytes and preimplantation embryos derived from ICSI using WT, Dicer and Drosha cKO sperm at 2PN, 2-cell, 4-cell, morula and blastocyst stages. The high-throughput quantitative real-time PCR (HT qPCR) analyses were performed using the Fluidigm BioMark HD HT qPCR system. *Gapdh* was used as an internal control for data normalization. Biological triplicates were analyzed for each genotype (i.e., WT, Dicer and Drosha cKO). The original HT qPCR data can be found in Table 2-S13.
Figure 2-9 – Fold changes of dysregulated genes with or without significantly altered targeting miRNAs in Dicer and Drosha cKO sperm in 2PN and 2-cell embryos derived from ICSI using Dicer and Drosha cKO sperm. (A) Log2 fold-changes of 96 genes analyzed in Dicer cKO sperm-derived 2PN and 2-cell embryos. Genes targeted by significantly altered miRNAs (fold changes ≥2, \( P_{adj} < 0.05 \)) in Dicer cKO sperm are in red.
font, and those without significantly changed targeting miRNAs are in green font. Transcripts without known targeting miRNAs in the databases are in black font. (B) Log2 fold-changes of 96 genes analyzed in Drosha cKO sperm-derived 2PN and 2-cell embryos. Genes targeted by significantly altered miRNAs (fold changes ≥2, \( P_{adj} < 0.05 \)) in Drosha cKO sperm are in red font, and those without significantly changed targeting miRNAs are in green font. Transcripts without known targeting miRNAs in the databases are in black font. The original data on HT qPCR assays and miRNA target analyses can be found in Tables 2-S14, -S15.
### Table 2-1. Preimplantation development of embryos derived from ICSI using WT oocytes and Dicer cKO sperm with or without sperm RNA supplementation

<table>
<thead>
<tr>
<th>Injected content</th>
<th>Total no. of surviving oocytes (no. of experiments)</th>
<th>Number of embryos at each stage</th>
<th>2PN (% of total)</th>
<th>2-cell (% of 2PN)</th>
<th>4-cell (% of 2-cell)</th>
<th>Morula (% of 2-cell)</th>
<th>Blastocyst (% of 2-cell)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT sperm</td>
<td>171 (10)</td>
<td></td>
<td>159 (87.72)(^a)</td>
<td>121 (80.67)(^a)</td>
<td>109 (80.08)(^a)</td>
<td>97 (80.17)(^a)</td>
<td>61 (50.41)(^a)</td>
</tr>
<tr>
<td>Dicer cKO sperm</td>
<td>226 (15)</td>
<td></td>
<td>165 (73.01)(^b)</td>
<td>113 (68.48)(^b)</td>
<td>71 (62.83)(^b)</td>
<td>57 (50.44)(^b)</td>
<td>34 (30.08)(^b)</td>
</tr>
<tr>
<td>Dicer cKO sperm+WT sperm small RNA</td>
<td>239 (14)</td>
<td></td>
<td>170 (71.13)(^b)</td>
<td>111 (65.29)(^b)</td>
<td>76 (68.47)(^b)</td>
<td>66 (59.46)(^b)</td>
<td>39 (35.14)(^b)</td>
</tr>
<tr>
<td>Dicer cKO sperm+WT sperm total RNA</td>
<td>109 (7)</td>
<td></td>
<td>82 (75.23)(^b)</td>
<td>44 (53.66)(^b)</td>
<td>35 (79.55)(^b)</td>
<td>28 (63.64)(^b)</td>
<td>22 (60.00)(^b)</td>
</tr>
<tr>
<td>Dicer cKO sperm+vehicle</td>
<td>81 (5)</td>
<td></td>
<td>69 (74.07)(^b)</td>
<td>32 (53.39)(^b)</td>
<td>16 (50.00)(^c)</td>
<td>9 (28.13)(^c)</td>
<td>7 (21.88)(^c)</td>
</tr>
<tr>
<td>Dicer cKO sperm+WT sperm total RNA+RNase (1:10 molar ratio)</td>
<td>54 (3)</td>
<td></td>
<td>39 (72.22)(^b)</td>
<td>25 (64.10)(^b)</td>
<td>11 (44.00)(^b)</td>
<td>9 (36.00)(^b)</td>
<td>4 (16.00)(^b)</td>
</tr>
<tr>
<td>Dicer cKO sperm+WT heart total RNA</td>
<td>42 (2)</td>
<td></td>
<td>31 (73.81)(^b)</td>
<td>19 (61.29)(^b)</td>
<td>10 (52.63)(^b)</td>
<td>7 (36.84)(^b)</td>
<td>5 (26.32)(^b)</td>
</tr>
<tr>
<td>None (control for parthenogenesis)</td>
<td>30 (2)</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

2PN, 2 pronuclei. Statistical analyses were conducted using \(\chi^2\) test; values with different superscripts are significantly different (\(P<0.05\)).

### Table 2-2. Preimplantation development of embryos derived from ICSI using WT oocytes and Drosha cKO sperm with or without sperm RNA supplementation

<table>
<thead>
<tr>
<th>Injected content</th>
<th>Total no. of surviving oocytes (no. of experiments)</th>
<th>Number of embryos at each stage</th>
<th>2PN (% of total)</th>
<th>2-cell (% of 2PN)</th>
<th>4-cell (% of 2-cell)</th>
<th>Morula (% of 2-cell)</th>
<th>Blastocyst (% of 2-cell)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT sperm</td>
<td>171 (10)</td>
<td></td>
<td>150 (87.72)(^a)</td>
<td>121 (80.67)(^a)</td>
<td>109 (80.08)(^a)</td>
<td>97 (80.17)(^a)</td>
<td>61 (50.41)(^a)</td>
</tr>
<tr>
<td>Drosha cKO sperm</td>
<td>164 (12)</td>
<td></td>
<td>133 (81.10)(^a)</td>
<td>100 (75.19)(^b)</td>
<td>64 (64.00)(^b)</td>
<td>52 (52.00)(^b)</td>
<td>26 (25.00)(^b)</td>
</tr>
<tr>
<td>Drosha cKO sperm+WT sperm small RNA</td>
<td>149 (10)</td>
<td></td>
<td>119 (73.87)(^b)</td>
<td>88 (73.55)(^b)</td>
<td>68 (77.27)(^b)</td>
<td>61 (69.32)(^b)</td>
<td>36 (40.91)(^b)</td>
</tr>
<tr>
<td>Drosha cKO sperm+WT sperm total RNA</td>
<td>79 (4)</td>
<td></td>
<td>66 (83.84)(^a)</td>
<td>44 (66.67)(^a)</td>
<td>35 (75.55)(^a)</td>
<td>28 (63.64)(^a)</td>
<td>15 (34.09)(^a)</td>
</tr>
<tr>
<td>Drosha cKO sperm+vehicle</td>
<td>97 (8)</td>
<td></td>
<td>69 (71.13)(^b)</td>
<td>45 (65.22)(^b)</td>
<td>27 (70.00)(^b)</td>
<td>26 (57.76)(^b)</td>
<td>11 (24.44)(^b)</td>
</tr>
<tr>
<td>Drosha cKO sperm+WT sperm total RNA+RNase (1:10 molar ratio)</td>
<td>64 (3)</td>
<td></td>
<td>45 (70.31)(^b)</td>
<td>23 (51.11)(^b)</td>
<td>10 (43.47)(^b)</td>
<td>9 (30.13)(^b)</td>
<td>4 (17.39)(^b)</td>
</tr>
<tr>
<td>Drosha cKO sperm+WT heart total RNA</td>
<td>41 (2)</td>
<td></td>
<td>31 (75.61)(^b)</td>
<td>24 (77.42)(^b)</td>
<td>17 (70.83)(^b)</td>
<td>15 (62.50)(^b)</td>
<td>7 (29.17)(^b)</td>
</tr>
<tr>
<td>None (control for parthenogenesis)</td>
<td>30 (2)</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

2PN, 2 pronuclei. Statistical analyses were conducted using \(\chi^2\) test; values with different superscripts are significantly different (\(P<0.05\)).
Table 3. Pups born through transfer of 2-cell embryos derived from ICSI using WT oocytes and WT, Dicer cKO, Drosha cKO sperm with or without supplementation of sperm-borne RNAs

<table>
<thead>
<tr>
<th>Injected content</th>
<th>No. of 2-cell embryos transferred (no. of experiments)</th>
<th>No. of recipients</th>
<th>No. of live-born offspring (%)</th>
<th>Pup genotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT sperm</td>
<td>65 (3)</td>
<td>4</td>
<td>18 (27.69)a</td>
<td>WT</td>
</tr>
<tr>
<td>Dicer cKO sperm</td>
<td>47 (3)</td>
<td>4</td>
<td>4 (8.51)b,c</td>
<td>DicercKO</td>
</tr>
<tr>
<td>Drosha cKO sperm</td>
<td>47 (3)</td>
<td>3</td>
<td>2 (4.26)c</td>
<td>DroshaKO</td>
</tr>
<tr>
<td>Dicer cKO sperm+WT sperm total RNA</td>
<td>54 (3)</td>
<td>4</td>
<td>8 (14.81)b</td>
<td>DicercKO</td>
</tr>
<tr>
<td>Drosha cKO sperm+WT sperm small RNA</td>
<td>71 (4)</td>
<td>5</td>
<td>9 (16.88)c</td>
<td>DroshaKO</td>
</tr>
</tbody>
</table>

Statistical analyses were conducted using $\chi^2$ test; values with different superscripts are significantly different ($P<0.05$).

Table 2-3. Pups born through transfer of 2-cell embryos derived from ICSI using WT oocytes and WT, Dicer cKO, Drosha cKO sperm with or without supplementation of sperm-borne RNAs

Table 4. Preimplantation development of embryos derived from ICSI using Drosha cKO oocytes and WT, Dicer cKO and Drosha cKO sperm

<table>
<thead>
<tr>
<th>Sperm genotype</th>
<th>Total no. of surviving oocytes (no. of experiments)</th>
<th>Number of embryos at each stage</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>2PN (% of total)</td>
</tr>
<tr>
<td>WT</td>
<td>125 (7)</td>
<td>109 (87.20)a</td>
</tr>
<tr>
<td>Dicer cKO</td>
<td>140 (6)</td>
<td>103 (73.57)a</td>
</tr>
<tr>
<td>Drosha cKO</td>
<td>142 (9)</td>
<td>100 (70.42)a</td>
</tr>
</tbody>
</table>

2PN, 2 pronuclei; statistical analyses were conducted using $\chi^2$ test; values with different superscripts are significantly different ($P<0.05$).

Table 2-4. Preimplantation development of embryos derived from ICSI using Drosha cKO oocytes and WT, Dicer cKO and Drosha cKO sperm

SUPPLEMENTARY INFORMATION

Supplemental figures, tables, and methods can be found at:

http://dev.biologists.org/content/143/4/635.supplemental
REFERENCES


Hong et al. 2008. Dicer1 is essential for female fertility and normal development of the female reproductive system. Endocrinology, 149, 6207 – 6212.


Yuan et al. 2015. mir-34b/c and mir-449a/b/c are required for spermatogenesis, but not for the first cleavage division in mice. Biol. Open, 4, 212 – 223.


CHAPTER III

SpermBase – A database for sperm-borne RNA contents

Andrew Schuster¹, Chong Tang¹, Daniel Dailey¹, Yeming Xie¹, Nicole Ortogero¹, Shuiqiao Yuan¹, Wei Yan¹,²,*

¹Department of Physiology and Cell Biology, University of Nevada School of Medicine, 1664 North Virginia Street, MS0575, Reno, NV 89557, USA

²Department of Biology, University of Nevada, Reno, 1664 North Virginia Street, Reno, NV 89557, USA

This chapter is in preparation for submission as a publication.

*Corresponding author:

Wei Yan M.D., Ph.D.
Professor
University of Nevada School of Medicine
Center for Molecular Medicine, Room 207B
1664 North Virginia Street, MS/0575
Reno, NV 89557
Tel: 775 784 7765
Fax: 775 784 4362
Email: wyan@medicine.nevada.edu
URL: http://www.medicine.nevada.edu/yan
ABSTRACT

Since their discovery ~three decades ago, sperm-borne RNAs, both large/small and coding/noncoding, have been reported in multiple organisms, and some have been implicated in spermatogenesis, early development, and epigenetic inheritance. Despite these advances, isolation, quantification and annotation of sperm-borne RNAs remain nontrivial. The yields and subspecies of sperm-borne RNAs isolated from sperm can vary drastically depending on the methods used, and no cross-species analyses of sperm RNA contents have been conducted using a standardized sperm RNA isolation protocol. To address these issues, we developed a simple RNA isolation method that is applicable to sperm of various species, thus allowing for reliable interspecies comparisons. Based on RNA-Seq analyses, we established SpermBase, a database dedicated to sperm-borne RNA profiling of multiple species. Currently, SpermBase contains large and small RNA expression data for human, rat, mouse, and rabbit total sperm and, for some species, sperm heads. By analyzing large and small RNAs for conserved features, we found that many sperm-borne RNA species were conserved across all four species analyzed, and among the conserved small RNAs, sperm-borne tsRNAs and miRNAs can target a large number of genes known critical for early development.
INTRODUCTION

Despite the cessation of transcription and shedding of the cytoplasm at the final stages of spermiogenesis, mature spermatozoa have been shown to possess diverse populations of both small and large RNAs (Casas and Vavouri, 2014; Sendler et al., 2013; Krawetz et al., 2011; Jodar et al., 2013; Yuan et al., 2016). These RNA populations, along with the paternal genome, are delivered to the oocyte during fertilization, where they can persist even beyond zygotic genome activation (Ostermeier et al., 2004; Yuan et al., 2016). Since their discovery in 1989, sperm-borne RNAs have been suggested to function not only in spermatogenesis but also during fertilization and early embryonic development, suggesting that they are not merely the remnants of sperm development (Pessot et al., 1989; Boerke et al., 2007; Jodar et al., 2013; Wu et al., 2014; Yuan et al., 2015a). Sperm lacking a family of five miRNAs (miR-449a/b/c and miR-34b/c) fail to support fertilization and preimplantation embryonic development (Yuan, 2015, Biology Open paper); sperm with aberrant miRNA contents display compromised fertilization rate and preimplantation embryonic development when injected into wild type oocytes (Yuan et al., 2016). In addition, sperm small RNAs have been found to play a role in epigenetic inheritance (Wagner et al., 2008; Rassoulzadegan et al., 2006; Yuan et al., 2015b; Gapp et al., 2014; Rodgers et al., 2015; Chen et al., 2016; Sharma et al., 2016).

The study of sperm RNAs has been challenging, in part, because of the difficulties associated with sperm RNA isolation. Compared to other cell types, sperm contain much fewer RNAs, and some RNAs appear to be localized to the nucleus and associated with the disulfide bond-rich sperm chromatin, which is difficult to lyse (Jodar et al., 2013; Goodrich et al., 2013; Goodrich et al., 2007). Therefore, it is no surprise that
different RNA extraction procedures could lead to hugely variable sperm-borne RNAs contents although the sperm RNAs are isolated from sperm of the same species – this issue is compounded by the inherent heterogeneity of RNA populations amongst sperm samples from individuals of the same species (Mao et al., 2013; Mao et al., 2014; Barragán et al., 2015; Cappallo-Obermann et al., 2011; Lalancette et al., 2009). These problems may explain why very little or no overlaps can be found in sperm RNA datasets reported by several groups (data not shown). Additionally, a sperm RNA isolation protocol that works for one species does not necessarily work for another, due to inter-species differences in sperm morphology and chromatin condensation (Varner and Johnson, 2007; Das et al., 2010; Shafeeque et al., 2014). The aforementioned issues surrounding the study of sperm-borne RNA render comparisons of data obtained using different experimental approaches unreliable. Therefore, it is necessary that a broadly applicable methodology be established for studying the RNA contents of sperm across species. To this end, we developed a simple and effective protocol for sperm RNA isolation, which can be applied to multiple species with only minor modifications. Based on sperm RNA-Seq data, we established SpermBase, a database dedicated to sperm RNA expression profiling for various species. Currently, SpermBase contains expression data of both large and small RNAs in sperm of four species (rat, mouse, human, and rabbit), with plans to expand to more species in the future. In addition to data from the intact spermatozoa (‘total sperm’), SpermBase also contains expression data for sperm head samples. RNAs localized in the sperm head are thought to be more functionally significant than those in other parts of the sperm (Yan et al., 2008; Yanagimachi, 2005; Yan, 2014). To demonstrate the utility of SpermBase, we compared the large and small
RNAs of mouse, rat, rabbit and human, to elucidate the conserved features of mammalian sperm-borne RNA populations.

RESULTS

Complete lysis of sperm nuclei is key to successful sperm RNA isolation

The RNA isolation method-induced bias results mainly from low RNA abundance, susceptibility of sperm RNA to degradation and inter-species differences in sperm chromatin (Mao et al, 2013; Mao et al, 2014; Jodar et al, 2013; Goodrich et al, 2013; Sendler et al, 2013; Gilbert et al, 2007; Varner and Johnson, 2007; Das et al, 2010; Shafeeque et al, 2014). Since many sperm RNAs are embedded inside the nuclei and associated with sperm chromatin, it is critical to lyse sperm heads completely so that sperm RNAs can be totally released. To this end, we established a ‘modular’ RNA isolation protocol that is applicable to the sperm of any species, by only slight modification of the lysis step. Given the enriched disulfide bonds throughout the sperm chromatin, a lysis buffer containing reducing agents is required for a complete lysis of sperm nuclei (Goodrich et al, 2007). In our protocol, we adopted the mirVana total RNA isolation kit, and made modifications only at the initial lysis step (Table 3-1). It should be noted that other, similar kits with reducing agents in their lysis solutions (e.g., RNeasy by Qiagen) have also been demonstrated to be efficacious for sperm RNA extraction and would likely be on par with our methodology with species-appropriate modifications to the lysis stage (Goodrich et al, 2013). As outlined in Table 3-1, human and rabbit sperm heads could be easily lysed by pipetting up and down in the lysis buffer, whereas mouse sperm required mechanical homogenization and rat sperm even needed heating in addition to homogenization.
Electropherograms of the extracted sperm-borne RNAs were used to assess overall sample quality (Figure 3-1). A lack of intact 18S and 28S rRNAs in sperm make traditional RNA integrity number (RIN) conventions inappropriate for assessing sperm RNA quality (Johnson et al., 2011; Cappallo-Obermann et al., 2011). Instead of a RIN of ~10 being ideal for other cell types, our lab has observed that a RIN of ~3 is indicative of high quality sperm RNA, consistent with a previous report (Schroeder et al., 2006). Lower RINs (< 2) typically represent excessive RNA degradation, while higher RINs (> 4) may indicate somatic cell RNA contamination. We also observed that, in addition to a RIN of 2 – 4, sperm RNA electrophoretic profiles across species are characterized by a large population of RNA less than 200 nt in length (Figure 3-1), consistent with previous observation in mouse (Kawano et al., 2012), human (Johnson et al., 2011), horse (Das et al., 2010), bull (Gilbert et al., 2007), and domestic swine (Yang et al., 2009). To our knowledge, this is the first time this common length distribution has been shown for rabbit (Figure 3-1A) and rat (Figure 3-1B) sperm-borne RNA.

**SpermBase is easy to use and will be expanded to cover as many species as possible**

SpermBase is a publically accessible database and will be found at [www.spermbase.org](http://www.spermbase.org). The website is separated into five main pages – Home, Search, Method, Species, and FAQ & Contact. The Home page (Figure 3-S1) provides an introduction to the database itself, touching on the topics discussed above in the introduction of this article. Links can be found throughout the introduction to other parts of the SpermBase website.
At the Search page, users can search SpermBase for expression data for their gene(s) of interest. Each RNA that was found in our data is categorized by its original or, if it was identified by the authors, its given name (‘Gene name’). As discussed below, the reads aligned to tRNA genes were classified as different subclasses (e.g., 5’ halve, tRF-5) of tRNA-derived sncRNAs (tsRNAs). The naming convention for these tsRNAs was to add the subclass of tsRNAs after the original tRNA name, e.g., the 5’ halve of “trna1000-PheGAA” would be named “trna1000-PheGAA-5halve.” Each RNA is then organized by the species it was identified in (‘Species’), the ‘Sample type’ (i.e., total sperm or sperm head), the class of large or small RNA, e.g., mRNA, piRNA (‘Class’), and its expression (‘Expression’). The expression values shown in SpermBase are the normalized read counts, given as RPK (reads per hundred thousand mapped reads) for sncRNA genes and RPKM (reads per kilobase of transcript per million mapped reads) for mRNAs.

The Method page describes the methodology (discussed previously) used for sperm RNA isolation followed RNA-Seq analyses for building SpermBase. As SpermBase expands to include additional species, we will update this page with the lysis stage parameters used for the new species. Users who employ our modular RNA isolation protocol on sperm from animals not described in SpermBase are encouraged to send us information on the lysis parameters that they found to be most optimal for that species.

On the Species page, a list of each species currently available on SpermBase is provided, along with download links for their respective expression data. The expression data for each individual species is available in multiple formats (.txt, .csv, and Excel .xls).
The files can be downloaded by either left clicking the file or right clicking to “Save link as…” if the user wishes to rename the file prior to downloading. A FASTA file containing the most abundant sequence for each individual tsRNAs observed is also provided for download. We will also announce any future additions to SpermBase on this page.

The FAQ & Contact page provides answers to frequently asked questions related to SpermBase and detailed contact information. Users who have questions not addressed in the FAQ are encouraged to email SpermBase. Troubleshooting queries and comments about the design and functionality of SpermBase are also welcome.

**Sperm-borne mRNAs**

RNA-Seq was performed using total RNA isolated from both sperm head (for rabbit and mouse) and total sperm (for rat, rabbit, mouse and human) samples to determine the expression of coding genes within mature sperm. Previous investigations into the mRNA content of mammalian sperm have found diverse populations of coding genes within both mouse and human sperm (Sendler *et al*, 2013; Fang *et al*, 2014). Similarly, we observed several thousand coding genes present in the total sperm and sperm head samples (Figures 3-2A and 3-2B).

Next, we assessed the degree to which sperm-borne mRNAs are conserved across mammals. After removing any genes expressed below 3 RPKM from consideration, we identified over five hundred coding genes that were present in the total sperm of all four species (Figure 3-2A, Table 3-S2A). Using the same cutoff, we observed 3,506 coding
genes that were commonly expressed in both of our sperm head samples (i.e., mouse and rabbit) (Figure 3-2B, Table 3-S2B). A gene ontology (GO) term enrichment analysis was performed on the conserved total sperm mRNAs that were identified (Figure 3-2C, Table 3-S3). “Development” and “morphogenesis” were the most prevalent amongst the significantly enriched biological process terms, raising the possibility that many of these conserved coding genes play a functional role during early embryonic development (Figure 3-2C, Table 3-S3A). One of the conserved genes we identified, Clu (clusterin), is a sperm-specific transcript known to be absent in oocytes, but present in zygotes (Ostermeier et al, 2004).

**Sperm-borne small RNAs**

In addition to mRNAs, we sequenced small noncoding RNA (sncRNA) in human, rat, rabbit, and mouse total sperm samples, as well as rabbit and mouse sperm head fractions. Expression data for miRNAs, piRNAs, small nucleolar RNAs (snoRNAs), small nuclear RNAs (snRNAs), rRNAs, tRNA-derived small RNAs (tsRNAs), and mitochondrial small RNAs (mitosRNAs) are currently available in SpermBase. Consistent with previous reports, we observed diverse sncRNA populations in the sperm of each species surveyed (Krawetz et al, 2011). The most abundant classes of sncRNAs in sperm were miRNAs, tsRNAs, piRNA, and mitosRNAs (Figure 3-3A; Table 3-S4). Interestingly, significant differences were observed in RNA contents between total sperm and sperm heads (Figure 3-3A). The tsRNAs were much more abundant in the sperm heads than in the total sperm suggesting that tsRNAs are mainly localized to the sperm heads. Reduced proportions of mitosRNAs and miRNAs in sperm heads suggest that
these two types of small RNAs, in addition to tsRNAs, are the major small RNA constituents in other compartments of the total sperm. Similarly, in rabbit sperm heads, tsRNAs were the predominant small RNA species followed by mitosRNAs and miRNAs; mitosRNAs were much more abundant in the total sperm, compared to total sperm, suggesting that mitosRNAs are mainly localized to the non-head regions of the total sperm. Considering their consistent abundance in both total sperm and sperm heads, we focused on four small RNA classes (piRNAs, mitosRNAs, miRNAs, and tsRNAs), while investigating the sncRNA conservation in mammalian sperm.

**piRNA** – Previous work has demonstrated that piRNAs are highly expressed in male germ cells, and play an essential role during spermatogenesis (Gou et al., 2014). However, little is known about the purpose of their presence in mature sperm (Casas and Vavouri, 2014). Unlike miRNAs, piRNAs are poorly conserved in their sequences across species (Mani and Juliano, 2013). Therefore, we investigated the general features of the piRNA populations in each species. In pachytene spermatocytes, there is a burst in piRNA expression – the piRNAs expressed in the meiotic phase of spermatogenesis are called ‘pachytene piRNAs,’ while the piRNAs that start to be expressed in spermatogonia are referred to as the ‘pre-pachytene piRNAs’ (Meikar et al., 2011; Gou et al., 2014). Pre-pachytene piRNA are typically 26 – 28 nt in length and have a strong preference towards uracil and adenine at their 1st and 10th nucleotide positions, respectively. In contrast, pachytene piRNAs are typically 30 nt in length and only possess a preference for adenine at their 1st nucleotide (Ortogero et al., 2014). Based on these characteristics, we assessed whether the piRNAs found in total sperm samples were pre-pachytene or pachytene in
origin. The majority of piRNAs in total sperm samples were 29 – 32 nt in length across species (Figure 3-S2A), and our analyses of nucleotide preferences showed a strong bias towards uracil for the 1st nucleotide (Figure 3-S2B) and only a slight preference towards adenine at the 10th nucleotide (Figure 3-S2C). These data suggest that the majority of the rabbit, human, rat, and mouse sperm-borne piRNAs are pachytene piRNAs.

mitosRNA – The majority of the sncRNA-Seq reads aligned to mitochondrial genome were much shorter than their matching full-length transcripts. Many of the sequences aligned to a mitochondrial RNA were a consistent length, suggesting that they are not derived from the degradation of intact transcripts, but rather are mitochondrial genome-encoded small RNAs (mitosRNA) that were purposefully produced, as reported previously (Ro et al, 2013) (Figures 3-S3 - S7). The lack of a consensus sequence length for mitosRNAs has been observed previously (Ro et al, 2013). We also found that sperm-borne mitosRNAs varied in length depending on their origin of the mitochondrial genes, and this feature was conserved across all four species surveyed (Figures 3-S4 - S7). mitosRNAs ranked second in terms of relative abundance among all four major sperm-borne small RNA species, i.e., tsRNAs > mitosRNAs > miRNAs > piRNAs (Figure 3-3A).

miRNA – Several groups have previously investigated the miRNA contents in sperm, for a variety of organisms such as mouse (Kawano et al, 2012; García-López et al, 2015), human (Sendler et al, 2013; Jodar et al, 2013), bull (Fagerlind et al, 2015; Stowe et al, 2014; Du et al, 2014; Govindaraju et al, 2012), and pig (Curry et al, 2011). Despite this
cache of sperm miRNA expression information, cross-species comparisons using this existing data would be, as discussed previously, intrinsically unreliable due to differences in methodology (Mao et al., 2013; Mao et al., 2014; Barragán et al., 2015; Cappallo-Obermann et al., 2011; Lalancette et al., 2009; Varner and Johnson, 2007; Das et al., 2010; Shafeeque et al., 2014). Using the standardized miRNA expression data available in SpermBase, we were able to identify 67 miRNAs that were present in the total sperm samples of all four species (Figure 3-3B; Table 3-S5A). These miRNAs accounted for the majority of all miRNAs expression in every species surveyed (Table 3-S5B). Many of the conserved miRNAs were members of the same clusters (i.e., their genes were within 10kb of one another), for example, one of these clusters contains the miR-34b/c family, which, along with miR-449a/b/c, are essential for normal spermatogenesis and fertility in male mice (Yuan et al., 2015a). Three (miR-34c-3p, miR-19b-3p, miR-148b-3p) of the seven miRNAs that were, in another study, found to be differentially expressed when comparing bulls of moderate and high fertility, were also members of the 67 conserved miRNAs (Fagerlind et al., 2015).

To assess whether the conservation of these miRNAs across species held any functional significance, we predicted the gene targets for all 67 miRNAs in silico using both RNAhybrid and miRanda to compare the miRNAs to 3’ UTR sequences, discarding the gene target predictions not made by both programs (Rehmsmeier et al., 2004; John et al., 2005). The predicted gene targets of each conserved miRNA can be found in Table 3-S6. Many miRNAs are functionally redundant, a phenomenon that is commonly observed in single miRNA knockout mouse models, the majority of which lack any
aberrant phenotype (Wu et al., 2014; Olive et al., 2015). Because of this, we used genes that were targeted by at least two of the 67 conserved miRNAs for our subsequent studies. In order to gauge whether the putative functions of these miRNAs are conserved, for every species, we performed a gene ontology (GO) term enrichment analysis on the top 2,000 redundantly targeted genes, ranked by the number of matching conserved miRNAs, then by the average p-value of each predicted gene target. The results of the GO term enrichment analysis are summarized in Table 3-S7. Interestingly, ‘development’ and ‘morphogenesis’ were two of the most prevalent terms for every species surveyed (Figure 3-4A). Likewise, of the top 20 biological process (BP) terms identified in our interspecies comparison, six were related to development (Table 3-S7G). These data suggest not only that these 67 miRNAs are potentially active as gene silencers during early development, but that their role is conserved across mammals.

To further assess whether the conserved sperm miRNAs are active during early development, we compared the list redundantly targeted murine genes (i.e., targeted by at least two miRNAs) to genes that are known to be expressed during the first stages of development (i.e., oocyte to the four-cell stage). We found that the numbers of predicted gene targets that matched to genes expressed in early development were significantly higher than the expected number of random matches (Table 3-2), supporting the putative role of the 67 conserved sperm miRNAs as regulators of gene expression during early development.

tsRNAs – While annotating the sncRNA-Seq data for SpermBase, we observed that tRNAs accounted for more reads than any other sncRNAs, with the exception of human
total sperm, in which mitosRNAs were more abundant (Figure 3-3A; Table 3-S4). A closer look at these tRNA-aligned reads revealed that the majority of them were ~30 nt, indicating that they actually represented tsRNAs and not intact mature tRNA species (Figure 3-S8). These tsRNAs have previously been identified in a myriad of organisms and cell types including sperm, where they were found to be highly abundant (Megel et al, 2015; Peng et al, 2012). Similar to piRNAs, tsRNAs are divided into further subgroups based on their length as well as their origins. The tsRNA species of 27 nt or longer were classified as 5’ or 3’ halves depending on whether they were derived from the 5’ or 3’ half of the mature tRNA (split at the anticodon), while those 19 – 26nt in length were classified as tRF-5’s or tRF-3’s depending on their half preference (Gebetsberger and Polacek, 2013). We found that 5’ halves were the most abundant tsRNAs in every species and sample type surveyed, especially sperm heads, consistent with previous findings in mouse sperm (Figure 3-3C) (Peng et al, 2012). The predominance of 5’ halves in sperm-borne small RNA populations, in addition to speculation that they may play an important role in epigenetic inheritance, led us to focus our attention on this particular class of tsRNAs (Peng et al, 2012; Kiani and Rassoulzadegan, 2013; Kiani et al, 2013; Sharma et al, 2016; Schuster et al, 2016; Chen et al, 2016).

To assess the similarity of the 5’ halve populations between the species on SpermBase, we sorted the total sperm 5’ halves by the amino acid of their precursor tRNA. In every organism, tRNA^{Gly} species accounted for the most 5’ halves, followed by tRNA^{Glu}, tRNA^{Val}, tRNA^{Met}, and tRNA^{Lys} (Figure 3-3D). This finding is in agreement
with previous studies on murine sperm-borne 5’ halves, which found that 5’ halves from tRNA\textsuperscript{Gly} and tRNA\textsuperscript{Glu} were the most abundant tsRNA in sperm (Peng \textit{et al}, 2012). The observation that the majority of the 5’ halves in each species surveyed are derived from the same group of tRNAs also indicates that the production and retention of these sncRNAs is conserved across mammalian sperm, similar to miRNAs (Figure 3-3B).

Several studies have found that 5’ halves may possess the ability to act as post-transcriptional gene regulators, similar to miRNAs (Elbarbary \textit{et al}, 2009; Ivanov \textit{et al}, 2011; Wang \textit{et al}, 2013; Yamasaki \textit{et al}, 2009). In order to evaluate whether the 5’ halves that we found in the SpermBase data were capable of complementary sequence-based gene regulation, we performed an unbiased gene target prediction analysis, matching the 5’ halves against the 3’ UTR, as well as the 5’ UTR and CDS sequences available for each species. Instead of using both miRanda and RNAhybrid as we did for the conserved sperm miRNAs, we solely used RNAhybrid. This is because RNAhybrid accounts for the length of the target sequence, ensuring that the longer 3’ UTR and CDS sequences would not receive a higher number of random matches than the 5’ UTR (Rehmsmeier \textit{et al}, 2004). The predicted gene targets for each species are provided in Table 3-S8. We observed that, on average, the 5’ UTR was targeted the most, compared to the 3’ UTR and CDS sequences, and that this was true for all four species (Figure 3-3E; Table 3-S8). The proportion of genes that were targeted by more than one 5’ halve was also consistently higher on average for the 5’ UTR-based analysis (~88% targeted by at least two 5’ halves) when compared to the CDS- and 3’ UTR-based analyses (~68% and ~78%, respectively), suggesting that there were fewer random target predictions
when the 5’ UTR sequences were analyzed. Based on these findings, we utilized the 5’ UTR-based putative gene targets for our subsequent analyses.

From the 5’ UTR-based gene target predictions, we selected the top 4,000 redundantly targeted genes, which were ranked according to the number of 5’ halves that targeted each gene, then by the average p-value for each prediction, to perform a GO term enrichment analysis for each species. The results of this analysis are summarized in Table 3-S9. Terms relevant to development and morphogenesis were common in the results for every species, similar to what was observed for the conserved sperm miRNAs (Figures 3-4A and 3-4B, Tables 3-S7G and 3-S9G). Of the top 20 BP terms for the 5’ halves, three were directly related to development (Table 3-S9G). Interestingly, the (5’ halve) BP term ranked second on Table 3-S9G was “Wnt signaling pathway,” a pathway vital for proper early embryonic development across the animal kingdom (Petersen and Reddien, 2009). Additionally, ‘catabolic’ and ‘metabolic’ were prevalent in the enriched terms (Figure 3-4B). This is not surprising, as sperm-borne 5’ halves have previously been linked to altered expression of metabolic genes in the early embryo (Chen et al, 2016). Overall, it seems that, like the conserved sperm miRNAs, sperm-borne 5’ halves may also play a role in regulating gene expression during early development. To investigate this further, we compared the redundantly targeted murine genes (i.e., targeted by at least two 5’ halves) identified in the 5’ UTR-based analysis to genes known to be expressed throughout early development. As also seen with the putative miRNA targets, the number of predicted 5’ halve targets that matched these early development expressed genes was significantly higher than number of anticipated random matches, providing
additional evidence that 5’ halves may play a functional role after fertilization (Table 3-2). The 5’ halves were predicted to target a much larger percentage of the early development genes (~88%) relative to the miRNAs (~25%) (Table 3-2). This observation is consistent with another recent study, which determined that sperm-borne 5’ halves targeted ~80% of the 8-cell embryo transcriptome (Chen et al, 2016).

DISCUSSION

Sperm-borne RNAs are thought to play important roles in spermatogenesis, early development, and epigenetic inheritance (Pessot et al, 1989; Boerke et al, 2007; Jodar et al, 2013; Wu et al, 2014; Yuan et al, 2015a; Wagner et al, 2008; Rassoulzadegan et al, 2006; Yuan et al, 2015b; Gapp et al, 2014; Rodgers et al, 2015; Chen et al, 2016; Sharma et al, 2016). Unfortunately, the unique structural features of sperm (i.e., highly condensed sperm chromatin and chromatin-associated RNA contents) and the relatively low abundance of sperm RNA makes the study of sperm RNA technically challenging (Jodar et al, 2013; Goodrich et al, 2013; Sendler et al, 2013; Goodrich et al, 2007; Mao et al, 2013; Mao et al, 2014; Barragán et al, 2015; Cappallo-Obermann et al, 2011; Lalancette et al, 2009; Varner and Johnson, 2007; Das et al, 2010; Shafeeque et al, 2014). Because of this, we established SpermBase, a repository of sperm-borne large and small RNA expression data. Currently, data are available for four mammalian species (human, mouse, rat, and rabbit), and in some cases, both total sperm and sperm head fractions. To ensure the comparability of multi-species sperm RNA data in SpermBase, we conducted sperm RNA-Seq on sperm RNA samples isolated using our standard sperm RNA isolation protocol. The key to successful isolation of high quality sperm RNAs is to
dissolve the sperm heads completely, which requires reducing agents, mechanical disruptions, and increased temperature during lysis, depending on the extent to which the sperm chromatins are compacted. In general, the more compact the sperm heads are, the more vigorous the treatment needs to be during lysis. In our protocol, the rat sperm requires longer homogenization (90 seconds) plus heating (65°C for 5min) for complete lysis of the sperm chromatin, whereas human and rabbit sperm only need up and down pipetting inside the lysis buffer. Interestingly, rat sperm appear to be much more compact than human and rabbit sperm.

Another interesting observation is the difference in RNA contents between total sperm and sperm heads. The fact that the sperm heads contain many more tsRNAs than the total sperm, and that the proportions of mitosRNAs, miRNAs and piRNAs are also reduced in some heads, suggest that these small RNAs are localized to the sperm nucleus and may be associated with sperm chromatin, as suggested in previous reports (Peng et al, 2012). This is also supported by the fact that RNA yields could be increased dramatically when sperm heads are completely lysed. Because of the considerable difference in RNA contents in sperm heads and total sperm, it is critical to verify that all sperm heads are completely lysed though microscopic observation after the lysis step. It is our hope that, in addition to using the data on SpermBase, other groups will utilize our RNA extraction method and quality control metric (i.e., RIN of 2 – 4 as an indicator of ideal sperm RNA quality) when performing their own sperm RNA studies. While we are confident in the efficacy of our method, we do caution other groups to make sure that they initially test, and if necessary, optimize the lysis step for their own samples, due to the sensitive nature of the extraction process. In the future, we hope to develop an RNA
extraction method that can be applied to sperm from any species, without making any alterations to the protocol. Other future improvements to SpermBase will include the addition of more sperm head fraction data, as well as the inclusion of other species, expanding the utility of SpermBase for the scientific community. Currently, we are working on adding large and small RNA expression data for the sperm of zebrafish, horse, monkey, and bull.

Using SpermBase data, we investigated the conserved general features of mammalian sperm mRNA and small RNA populations. We limited our analyses of the conservation of sperm-borne small RNAs to four classes – piRNAs, mitosRNAs, miRNAs, and tsRNAs, because they represent the most abundant small RNA species in both total sperm and sperm heads. The piRNA populations in the total sperm of all four mammals appeared to be pachytene in origin. As the sequences of individual piRNAs are poorly conserved across species, it is difficult at this time to say whether the piRNA present in sperm serve some biological purpose or are just the random remnants of the pachytene piRNA expression burst (Gou et al., 2014). In every species, mitosRNAs appear to be the second most abundant small RNA species in both total sperm and sperm heads; however, their physiological role remains unknown. Given that mitosRNAs are mainly confined to mitochondria and the mitochondria are essential for sperm motility, it would not be surprising if mitosRNAs differ in abundance or subspecies between fertile and infertile sperm (Sousa et al., 2011). A total of 67 conserved miRNAs account for the majority of all miRNAs expressed in every species examined. The finding that the predicted gene targets of these conserved miRNA are all related to early development,
suggests that sperm-borne miRNAs may have a role in fertilization and early development. Indeed, our lab recently reported that oocytes fertilized with sperm deficient in sperm-borne miRNAs due to *Dicer* or *Drosha* conditional knockout (cKO) displayed a significantly reduced developmental potential (Yuan et al., 2016). Together, our results strongly suggest that these 67 conserved miRNAs are important for successful reproduction, and should be investigated in future studies on male infertility.

Our analyses indicate that tsRNAs are the most abundant sperm small RNAs in every species, with 5’ halves as the most dominant subclass, consistent with a previous report (Peng et al., 2012). In every species surveyed, the bulk of these 5’ halves originated from the same precursor mature tRNAs, namely tRNA\textsubscript{Gly}, tRNA\textsubscript{Glu}, tRNA\textsubscript{Val}, tRNA\textsubscript{Met}, and tRNA\textsubscript{Lys}. Unlike miRNAs, which preferentially bind to the 3’ UTR of their target genes (Lewis et al., 2005), sperm-borne 5’ halves preferentially target the 5’ UTR (Figure 3-3E). Interestingly, the genes that are targeted by these 5’ halves are mostly related to early development, suggesting the sperm-borne tsRNAs may also have a role in regulating early development. Previous studies have demonstrated that 5’ halves have the potential to act as post-transcriptional regulators by binding to the 5’ UTR in a reverse complimentary manner in cell lines (Wang et al., 2013; Ivanov et al., 2011; Elbarbary et al., 2009; Yamasaki et al., 2009). In another study, 5’ halves inhibited translation via the displacement of eIF4G/eIF4A from the 5’ ends of mRNA (Ivanov et al., 2011). Our findings, together with these prior studies, suggest that 5’ halves may regulate gene expression in early development by binding mRNA through complementary sequences in their 5’ UTRs, displacing translational initiation machinery. In a similar finding, Chen et al determined that putative 5’ halve target sites in the CDS
were less frequent than the number observed in promoter regions (i.e., 2 kb upstream of the transcriptional start site) (Chen et al, 2016). The data housed at SpermBase will undoubtedly be a boon to future investigations into the mechanism behind the proposed regulatory functions of sperm-borne 5’ halves.

Sperm-borne RNAs have been implicated as potential mediators of epigenetic inheritance (Wagner et al, 2008; Rassoulzadegan et al, 2006; Yuan et al, 2015b; Gapp et al, 2014; Yan, 2014; Rodgers et al, 2015; Chen et al, 2016; Sharma et al, 2016). Of particular interest are the sperm-borne tsRNAs, as a methyltransferase involved with tsRNA production, Dnmt2, was found to be necessary for the transmission of two epigenetically inherited phenotypes (Kiani et al, 2013). While the effects of Dnmt2 deficiency on the tsRNA populations in sperm have yet to be determined, in other cell types, the production of 5’ halve species was found to increase in the absence of Dnmt2 (Schaefer et al, 2010). It is therefore possible that changes in sperm-borne 5’ halve levels can have an effect on the early embryo; this is in agreement with the results of our analysis of the predicted gene targets of the 5’ halves, which suggested that 5’ halves interact with genes related to development. Recently, in three separate epigenetic inheritance models, sperm-borne 5’ halve expression was observed to be affected by an altered diet (i.e., high fat or low protein) or vinclozolin (a common agricultural fungicide) exposure (Chen et al, 2016; Sharma et al, 2016; Schuster et al, 2016). Future investigations into other examples of epigenetic inheritance could potentially benefit from the tsRNA expression data and consensus sequences housed in SpermBase.
Since their discovery, research on sperm-borne RNAs has revealed that they are not simply the remnants of spermatogenesis – to the contrary, the likelihood that sperm RNAs play a functional role, in many different contexts (e.g., post-transcriptional regulation during early development and epigenetic inheritance), has only increased in recent years. The difficulties associated with sperm RNA isolation and sperm RNA contents in various species have no doubt slowed the advance of efforts to elucidate additional functions and to assess the conservation of these RNAs across evolution. These issues will be alleviated by SpermBase, which provides public expression data as well as a simple and effective RNA extraction methodology, thereby making the study of sperm-borne RNA more accessible to other labs. We hope that, with this contribution, we will see an acceleration of our understanding of the physiological roles of sperm-borne RNAs.

ACKNOWLEDGEMENTS

The authors would like to thank Qi Chen for his review and comments to the manuscript. Author contributions: A.S. and W.Y. conceived and designed the experiments, A.S., C.T., S.Y., and N.O. performed the experiments, A.S. and C.T. analyzed the data, C.T. and Y.X. created the database, A.S. and W.Y. wrote the manuscript. This work was supported, in part, by grants from the NIH (HD060858, HD071736 and HD076495 to W.Y.) and the Templeton Foundation (PID: 50183 to WY). Sequencing was conducted in the Single Cell Genomics Core supported, in part, by a NIH grant (P20RR18751).
MATERIALS AND METHODS

Sperm collection

Mouse total sperm and sperm head collection were conducted as follows: Mouse epididymides were placed in 37 °C HEPES buffered human tubule fluid (HEPES-HTF) and dissected into smaller fragments. The sperm were allowed to escape from the epididymis during a 30 minute incubation at 37°C. The sperm-containing supernatants were collected and centrifuged for 5 minutes at 700xg. After removing the supernatants, fresh HEPES-HTF was added gently on top of the sperm pellet. During a 30 minute incubation at 37°C, sperm were allowed to ‘swim-up’ from the pellet. The swim-up sperm were collected with the supernatant. After pelleting the sperm at 700xg for 5 minutes, the sperm were washed three times with 1XDPBS and centrifugation at the same speed. After the third wash, the supernatant was removed and the sperm pellets were snap frozen in liquid nitrogen. For sperm head isolation, sperm were added to a 4.5 mL 83.5% sucrose cushion and centrifuged at 100,000 g for one hour (Beckman SW41Ti rotor); afterwards, the sperm head pellet was collected.

Rat total sperm were isolated as follows: Rat epididymides were dissected and minced in F12 culture medium containing 0.1% bovine serum albumin followed by a 30 min incubation at 37°C. The sperm-containing supernatants were collected and washed with 1XPBS by centrifugation (800-1,000xg for 5min). Sperm pellets were re-suspended in 200µL NIM medium (121.6 mM KCl, 7.8 mM Na2HPO4, 1.4 mM KH2PO4, 0.1% polyvinyl alchol, and 10 mM EDTA), 100 µL collagenase, and 100 µL hyaluronidase followed by an incubation at 37°C for 1h, with occasional mixing. Rat total sperm were then washed three times using 500µL NIM though centrifugation (4,000xg for 3 min)
before snap frozen in liquid nitrogen and frozen sperm pellets were stored at -80°C until RNA isolation.

Rabbit ejaculates were collected using the artificial vagina method, as described (Bredderman et al., 1964). Rabbit sperm were washed three times with HEPES-HTF medium and the sperm pellets were snap frozen in liquid nitrogen stored at -80°C until RNA isolation. De-identified human sperm samples were purchased from California CryoBank. The cryopreserved human sperm were thawed by incubation in a waterbath at 37°C followed by three washes with HEPES-HTF and the sperm pellets were subjected to RNA isolation using the method described below.

**Sperm RNA isolation**

Total RNA was isolated using the mirVana miRNA Isolation Kit (Life Technologies) following the manufacturer’s instructions with modifications at the lysis step. The same procedures were used for total sperm and sperm head samples of each species. For mouse sperm, after the addition of the lysis buffer, the frozen sperm pellets were homogenized using a homogenizer at low settings for 1 minute on ice. For human and rabbit samples, after the addition of the lysis buffer, the frozen sperm pellets were pipetted up and down using a hand-held pipette (Eppendorf) on ice, until the pellet dissolved. For rat samples, after the addition of the lysis buffer, the frozen sperm pellets were homogenized using a homogenizer at low settings for 90 seconds, followed by a five minute incubation at 65°C. After lysis, a small aliquot (~5-10ul) was placed on a glass slides followed by covering with a coverslip for microscopic examination to ensure sperm nuclei were completely lysed. The lysis step was repeated until sperm nuclei were
completely lysed. The completely lysed sperm samples were subjected to the remaining default protocol for RNA isolation.

For quality control, sperm RNA samples were analyzed using the RNA 6000 Nano chips on an Agilent 2100 Bioanalyzer (Agilent). The RNA integrity number (RIN) was used to estimate the sperm RNA quality and a RIN of 2 – 4 was regarded as good sperm RNA quality.

**RNA-Seq**

Library preparation was performed using the Ion Total RNA-Seq Kit v2 (Life Technologies). Large RNA libraries were prepared using the whole transcriptome protocol provided with the library kit, while small RNA libraries were prepared using the small RNA protocol. For rat sperm, prior to small RNA library construction, total RNA samples were enriched for small RNA using the protocol provided with the aforementioned library kit (Life Technologies). For rabbit sperm, to obtain higher library yields, 22 cycles were used instead of 14 (the default) during the second stage of cDNA amplification. To remove excessive primer dimers, aliquots of 40 ng of rabbit small RNA libraries were size selected using 4% E-Gel EX, set on program 8 (size selected library fragments ~90 – 150 bp) (Life Technologies). Rat large and small RNA libraries, and the rabbit small RNA libraries were barcoded using Ion Xpress Barcode Adapters (Life Technologies). Quality control was performed using Agilent High Sensitivity chips (Agilent) and Experion DNA 1K kits (BioRad). Mouse and human small RNA libraries were loaded onto Ion 318 v2 chips (one sample per chip) using the Ion PGM Template OT2 200 Kit and Ion PGM Sequencing 200 Kit v2, and sequenced on the Ion PGM
system (Life Technologies). Barcoded rat and rabbit small RNA libraries were loaded onto Ion PI chips (one chip per species) via the Ion PI Template OT2 200 v3 and Ion PI Sequencing 200 v3 kits, and sequenced on an Ion Proton Sequencer (Life Technologies).

Mouse and human large RNA libraries were loaded onto Ion PI chips (one chip per sample) via the Ion PI Template OT2 200 v2 and Ion PI Sequencing 200 v2 kits, and sequenced on an Ion Proton Sequencer (Life Technologies). Rat and rabbit large RNA libraries were loaded onto Ion PI chips (three samples per chip for rat, two samples per chip for rabbit) via the Ion PI Template OT2 200 v3 and Ion PI Sequencing 200 v3 kits, and sequenced on an Ion Proton Sequencer (Life Technologies).

**Bioinformatics**

The large RNA-Seq data was annotated as follows: Reads were trimmed using fastx_trimmer and fastq_quality_trimmer ($t = 30$), and the resulting trimmed reads were mapped to the genome of each respective species (mouse, mm10; rat, rm5; human, hg19; rabbit, oryCun2) with TopHat (v2.09; default settings plus --b2-very-sensitive -r 200 and --mate-std-dev to 100) (Trapnell et al., 2009). Illumina iGenome references (ENSEMBL) were used for rat, mouse, and human, and generated via TopHat for rabbit (ENSEMBL, release 76) (Flicek et al., 2014; Trapnell et al., 2009). The aligned reads were then assembled using Cufflinks (v2.1.1; default settings plus --frag-bias-correct, --max-bundle-length 1e7, and --multi-read-correct) using the same genome reference versions and mask GTF files containing all known RNA for each species (ENSEMBL) (Flicek et al., 2014). Expressed genes in the data with the ‘protein_coding’ biotype (ENSEMBL) were extracted for further study and are housed on SpermBase.
The sncRNA-Seq data was annotated as follows: Reads <15nt and >50nt were
discarded. The remaining reads were matched to known sncRNA, consisting of, when
available for each species, mature miRNA (miRBase release 21), tRNA (Genomic tRNA
Database; mm10, mouse; rn5, rat; oryCun2, rabbit; hg19, human), piRNA (piRNABank),
rRNA (ENSEMBL, release 76), snoRNA (ENSEMBL, release 76), snRNA (ENSEMBL,
release 76), and mitochondrial RNA (ENSEMBL, release 76), using Sequery (0 – 2
mismatches allowed) (Kozomara and Griffiths-Jones, 2014; Chan and Lowe, 2009;
Lakshmi and Agrawal, 2008; Flicek et al, 2014; Ortogero et al, 2013). Unmatched reads
were matched to mouse testis (Song et al, 2011) and sperm endo-siRNA (Yuan et al,
2016) with Sequery (0 mismatches allowed). The remaining unmatched reads were
aligned to the genome of each respective species (human, hg19; mouse, mm10; rat, rn5;
rabbit, oryCun2) via Bowtie (settings -n 2 -k 3 --best --al -q) (Langmead et al, 2009).
Aligned reads were matched to the genomic coordinates of known mature miRNA,
tRNA, rRNA, snRNA, snoRNA, and mitochondrial RNA, when available for each
species (same databases used previously). Additional annotation of the rabbit sncRNA-
Seq data was performed, due to the incomplete annotation of the rabbit sncRNA
transcriptome. Unmatched, aligned reads 18 – 25 nt in length were matched to known
rat, human, and mouse mature miRNA (miRBase release 21) via Sequery (0 – 2
mismatches allowed) (Kozomara and Griffiths-Jones, 2014; Ortogero et al, 2013).
Unmatched, aligned reads 26 – 32 nt in length were matched to known rat, human, and
mouse piRNA (piRNABank) via Sequery (0 – 2 mismatches allowed) (Lakshmi and
Agrawal, 2008; Ortogero et al, 2013). Remaining unmatched, aligned reads 26 – 32nt in
length were analyzed by piRNA Predictor, to detect novel rabbit piRNA (Zhang et al., 2011). The remaining unmatched, aligned reads 26 – 32 nt in length were then matched to the novel rabbit piRNA via Sequery (0 – 2 mismatches) (Ortogeró et al., 2013). Read counts were obtained by in-house Python scripts. Reads were normalized as reads-per-hundred thousand aligned reads (RPK). Genes with fewer than 1 RPK were not included in the expression tables of SpermBase; when a gene had greater than 1 RPK in one sample type (e.g., total sperm), but not the other (e.g., sperm head), the expression of the gene was included for both sample types.

The sncRNA-Seq reads that initially matched to tRNA genes were extracted and matched to the 5’ and 3’ halves of the full length tRNA (split at the 3’ end of the anticodon), via Sequery (0 – 2 mismatches allowed) (Ortogeró et al., 2013). Reads ≥ 27 nt and ≤ 26 nt were named halves and tRFs, respectively, and further classified as 5’ halves, tRF-5’s, 3’ halves, and tRF-3’s based on whether the read aligned to the 5’ or 3’ half of the tRNA.

The conserved sperm miRNA gene targets were predicted using RNAhybrid (settings -n 50 -m 50000 -c -d xi,θ -p 0.05 -e -20) and miRanda (default settings; score ≥ 140 and energy ≤ -20) against the 3’ UTR of their respective species (ENSEMBL, release 76), discarding predictions not made by both programs (Flicek et al., 2014; Rehmsmeier et al., 2004; Rehmsmeier and Krueger, 2006; Enright et al., 2005). Human miRNA were used to match against rabbit 3’ UTR, due to the lack of confirmed mature miRNA sequences for the conserved miRNA identified. The 5’ halve gene targets were predicted using RNAhybrid (settings -n 50 -m 50000 -c -d xi,θ -p 0.01 -e -20) against the 5’ UTR, CDS, and 3’ UTR sequences of their respective species’ (ENSEMBL, release 76). The xi and θ values for each analysis were determined by RNAcalibrate (settings -n 50 -m
using randomly selected mature miRNA (miRBase, release 21) and known 5’ halve sequences (tRFDB, human and mouse) (Table 3-S1) (Kozomara and Griffiths-Jones, 2014; Rehmsmeier et al, 2004; Kumar et al, 2015).

Gene ontology term enrichment analysis was performed using the g:Profiler suite (Reimand et al, 2007; Reimand et al, 2011). Input gene lists were ordered by the number of miRNA or 5’ halves that targeted each gene (highest to lowest), then by the average p-value (lowest to highest). The top 2000 and 4000 genes of the miRNA and 5’ halve ordered lists, respectively, were analyzed in g:GOST (settings: significant only, ordered query, no sorting or hierarchical sorting - moderate, functional categories with 3 – 1000 terms only) (Reimand et al, 2007; Reimand et al, 2011). Word clouds were generated using Genes2WordCloud (http://www.maayanlab.net/G2W/) using the biological process terms found to be significantly enriched (p ≤ 0.05) using the above settings (no sorting), and the aforementioned ordered lists (miRNA and 5’ halve putative gene targets) or the total sperm coding genes (RPKM ≥ 3) found to be conserved across species (Baroukh et al, 2011).

The significance of the number of observed matches between the predicted conserved miRNA and 5’ halve gene targets and genes present in early development (≥ 5 FPKM) was determined using a Chi-square test. Early development gene expression was obtained from the Database of Transcriptome in Mouse Early Embryos (DBTMEE) (Park et al, 2015). The observed number of matches were compared to the number of expected random matches, which was defined as the number of early development genes for each stage multiplied by the percentage of all known mouse coding genes that were targeted by the conserved miRNA or 5’ halves.
**Figure 3-1** – Sperm RNA quality control. Representative electropherograms generated from the sperm-borne total RNA of (A) rabbit, (B) rat, (C) human, and (D) mouse are provided along with their respective RNA integrity number (RIN).
Figure 3-2 – (A) Conserved total sperm and (B) sperm head coding genes. The number of genes identified in each species, and more than one species, are provided. Common gene names were used to determine conservation across species. (C) Common words in enriched GO terms. Biological process (BP) terms significantly enriched \((p \leq 0.05)\) amongst the total sperm conserved coding genes (Table 3-3A, no filtering) were analyzed for common word usage. The size of each word is based on its frequency within the enriched BP terms.
Figure 3-3 – Sperm small RNA contents in four mammalian species including the mouse, rat, rabbit and human.  (A) Pie charts showing the proportional distribution of
each of the eight sncRNA populations in total sperm (TS) or sperm heads (SH) based on sncRNA datasets currently available in SpermBase. (B) Venn diagram showing sperm miRNAs conserved among the four mammalian specie based on sncRNA datasets currently available in SpermBase. (C) Pie charts showing the proportional distribution of each of the five tsRNA subclasses in total sperm (TS) and sperm heads (SH) in the four mammalian species. “Other” refers to tRNA-aligned reads that were not classified as tsRNAs. (D) Histogram showing conservation of 5’ halve origins. The tRNA genes from which 5’ halves were derived were grouped by amino acid, and ranked. “Misc” refers to amino acids not shown on the table. (E) Relative number of 5’ halve gene targets by target site. For each species, the number of predicted gene targets for each 5’ halve when matched against 5’ UTR, CDS, and 3’ UTR sequences were normalized to the number of 3’ UTR-based gene targets observed.
Figure 3-4 – Common words in the biological process (BP) terms enriched in the predicted gene targets of (A) conserved sperm miRNA and (B) 5’ halves. BP terms significantly enriched (p ≤ 0.05) amongst the predicted gene targets for both sets of sncRNA were analyzed for common word usage. Words that were determined to be significantly prevalent in multiple species were used to generate both (A) and (B). The size of each word reflects its relative frequency. (C) Example of 5’ halve binding to the 5’ UTR of an mRNA (H1foo).
Figure 3-S1 – SpermBase home page.
Figure 3-S2 – (A) Length distribution of sncRNA-Seq reads that aligned to piRNA genes. For the total sperm samples of each species, relative expression of piRNAs was
separated by piRNA species length. (B, C) piRNA nucleotide preferences. The distribution of nucleotides at the 1\textsuperscript{st} (B) and 10\textsuperscript{th} (C) positions (starting at the 5’ end) of sperm-borne piRNA were determined.

**Figure 3-S3** – Length distribution of sncRNA-Seq reads that aligned to mitochondrial genes. The relative expression of mitosRNAs for each length is provided for the total sperm samples of the four mammalian species studied.
Figure 3-S4 - Length distribution of abundant mouse total sperm mitosRNAs. Reads that aligned to the ten most abundant mitochondrial RNA in mouse total sperm were separated by length. The relative number of reads for each length is provided for the top ten mitochondrial genes.
Figure 3-S5 - Length distribution of abundant human total sperm mitoRNA. Reads that aligned to the ten most abundant mitochondrial RNA in human total sperm were separated by length. The relative number of reads for each length is provided for the top ten mitochondrial genes.
Figure 3-S6 - Length distribution of abundant rat total sperm mitosRNAs. Reads that aligned to the ten most abundant mitochondrial RNA in rat total sperm were separated by length. The relative number of reads for each length is provided for the top ten mitochondrial genes.
Figure 3-S7 - Length distribution of abundant rabbit total sperm mitosRNAs. Reads that aligned to the ten most abundant mitochondrial RNA in rabbit total sperm were separated by length. The relative number of reads for each length is provided for the top ten mitochondrial genes.
**Figure 3-S8** – Length distribution of sncRNA-Seq reads aligned to tRNA genes. The relative number of reads for each length is provided for each species, for both total sperm (TS) and sperm head (SH), when available.
TABLES

<table>
<thead>
<tr>
<th>Species</th>
<th>Modification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human &amp; Rabbit</td>
<td>Pipet sperm gently up and down on ice until sperm pellet dissolves</td>
</tr>
<tr>
<td>Mouse</td>
<td>Homogenize sample on a low setting for one minute on ice</td>
</tr>
<tr>
<td>Rat</td>
<td>Homogenize sample on a low setting for 90 seconds on ice, followed by a five minute incubation at 65 °C</td>
</tr>
</tbody>
</table>

Table 3-1 – Summary of the lysis procedure used for each of the four species studied.

<table>
<thead>
<tr>
<th>Stage</th>
<th>No. of genes present at stage (≥ 5 FPKM)</th>
<th>miRNA</th>
<th>5’ halves</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>No. of expected random matches</td>
<td>No. of actual matches</td>
</tr>
<tr>
<td>Oocyte</td>
<td>7374</td>
<td>1681</td>
<td>1818</td>
</tr>
<tr>
<td>1-cell</td>
<td>7111</td>
<td>1621</td>
<td>1779</td>
</tr>
<tr>
<td>2-cell</td>
<td>7309</td>
<td>1666</td>
<td>1818</td>
</tr>
<tr>
<td>4-cell</td>
<td>7273</td>
<td>1658</td>
<td>1804</td>
</tr>
<tr>
<td>P-value</td>
<td></td>
<td>1.60E-11</td>
<td></td>
</tr>
</tbody>
</table>

Table 3-2 – Predicted miRNA and 5’ halve gene targets are significantly present in early development stages. The predicted gene targets of the sperm-borne 5’ halves and conserved miRNAs were matched against genes known to be present (≥ 5 FPKM) in early development stages and compared to the number of expected random matches. The significance of the matching, determined by a Chi-square test, is provided (P-value).
Table 3-S1 – ξ and θ values used for RNAhybrid. The ξ and θ values that were used in our RNAhybrid gene target predictions are provided. The values were determined using RNAcalibrate. Randomly selected mature miRNA or known 5’ halves were matched against 5’ UTR, CDS, or 3’ UTR sequences in RNAcalibrate, and the average ξ and θ values were determined.

Table 3-S2 – Conserved sperm-borne coding genes. The total sperm (A) and sperm head (B) coding genes that were found to be conserved across species (Figures 3-2A and 3-2B) are provided.
[see attached file]

Table 3-S3 – Enriched GO terms (conserved total sperm coding genes). The GO Term ID, term description, and p-value of each biological process (A), cellular component (B), and molecular function (C) are provided. Murine GO terms were used for the analysis. Results without filtering and moderate hierarchical filtering are provided on each table.

<table>
<thead>
<tr>
<th>sncRNA class</th>
<th>Human TS (%)</th>
<th>Mouse TS (%)</th>
<th>Mouse SH (%)</th>
<th>Rat TS (%)</th>
<th>Rabbit TS (%)</th>
<th>Rabbit SH (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>miRNA</td>
<td>15.6</td>
<td>23.7</td>
<td>12.2</td>
<td>22.2</td>
<td>7.6</td>
<td>4.3</td>
</tr>
<tr>
<td>piRNA</td>
<td>11.4</td>
<td>6.5</td>
<td>6.2</td>
<td>29.2</td>
<td>19.1</td>
<td>2.2</td>
</tr>
<tr>
<td>tsRNA</td>
<td>20.4</td>
<td>46.5</td>
<td>75.5</td>
<td>40.6</td>
<td>42.1</td>
<td>82.5</td>
</tr>
<tr>
<td>snoRNA</td>
<td>0.2</td>
<td>1.1</td>
<td>1</td>
<td>0</td>
<td>0.8</td>
<td>0.3</td>
</tr>
<tr>
<td>snRNA</td>
<td>0.8</td>
<td>0.6</td>
<td>0.7</td>
<td>0</td>
<td>0.4</td>
<td>0.7</td>
</tr>
<tr>
<td>rRNA</td>
<td>3.3</td>
<td>0.8</td>
<td>0.5</td>
<td>2.4</td>
<td>0.7</td>
<td>1.4</td>
</tr>
<tr>
<td>endo-siRNA</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0</td>
</tr>
<tr>
<td>mitochondrial RNA</td>
<td>48.2</td>
<td>20.7</td>
<td>3.6</td>
<td>5.5</td>
<td>29.3</td>
<td>8.5</td>
</tr>
</tbody>
</table>

Table 3-S4 – Small RNA distribution. The relative expression of each class of sncRNA is provided as percentages for each sncRNA dataset available on SpermBase.

[see attached file]

Table 3-S5 – Conserved sperm miRNA. (A) Conserved sperm miRNA. This list contains the miRBase ID, without the species abbreviation, of all 67 conserved sperm miRNA identified. (B) Relative abundance of conserved sperm miRNA. The amount of total miRNA expression that the 67 conserved miRNA accounted for in each species and sample type is presented as a percentage.
Table 3-S6 – Conserved miRNA gene target predictions. The predicted gene targets of the 67 conserved miRNA observed, determined by miRanda (score ≥ 140 and energy ≤ -20) and RNAhybrid (p ≤ 0.05 and energy ≤ -20), are organized by species. The gene targets are provided as ENSEMBL Gene IDs.

Table 3-S7 – Enriched GO terms (conserved miRNA predicted targets). The GO Term ID, term description, and p-value of each enriched biological process (BP) (A, D), cellular component (B, E), and molecular function (C, F) term are provided. Results without any filtering (A – C) and with moderate hierarchical filtering (D – F) applied are provided. (G) The hierarchically filtered BP terms identified in the lists of predicted gene targets for the conserved sperm miRNA (D) were ranked by the number of species they were enriched in, then by the averaged p-value for each term. The top 20 BP terms are shown in the table (first to last).
Table 3-S8 – 5’ halve gene target predictions. The predicted gene targets of the 5’ halves observed in total sperm, determined by RNAhybrid (p ≤ 0.05 and energy ≤ -20), are summarized. The RNAhybrid predictions in each table are separated by the target sequence (5’ UTR, left; CDS, middle; 3’ UTR, right) and separate tables are provided for (A) human, (B) mouse, (C) rat, and (D) rabbit total sperm.

Table 3-S9 – Enriched GO terms (5’ halve predicted targets). The GO Term ID, term description, and p-value of each enriched biological process (BP) (A, D), cellular component (B, E), and molecular function (C, F) term are provided. Results without any filtering (A – C) and with moderate hierarchical filtering (D – F) applied are provided. (G) The hierarchically filtered BP terms identified in the lists of predicted gene targets for the 5’ halves (D) were ranked by the number of species they were enriched in, then by the averaged p-value for each term. The top 20 BP terms are shown in the table (first to last).
REFERENCES


Varner and Johnson. 2007. From a sperm’s eye view – revisiting our perception of this intriguing cell. *Proc. 53rd Annual Conv. Am. Assoc. Equine Practitioners* (Orlando, FL USA)


Yuan et al. 2015a. mir-34b/c and mir-449a/b/c are required for spermatogenesis, but not for the first cleavage division in mice. Biol. Open, 4, 212 – 223.


Yuan et al. 2016. Sperm-borne miRNAs and endo-siRNAs are important for fertilization and preimplantation embryonic development. Development, 143, 635 – 647.

CHAPTER IV

Ancestral vinclozolin exposure alters the epigenetic transgenerational inheritance of sperm small noncoding RNAs

Andrew Schuster¹, Michael Skinner², Wei Yan¹*

¹Department of Physiology and Cell Biology, University of Nevada School of Medicine, 1664 North Virginia Street, MS0575, Reno, NV 89557, USA

²Center for Reproductive Biology, School of Biological Sciences, Washington State University, Pullman, WA 99164-4236, USA

This chapter can be found in the following publication:

Schuster et al. 2016. Ancestral vinclozolin exposure alters the epigenetic transgenerational inheritance of sperm small noncoding RNAs. Environ. Epigenet., 2, dvw001

*Corresponding author:

Wei Yan M.D., Ph.D.  
Professor  
University of Nevada School of Medicine  
Center for Molecular Medicine, Room 207B  
1664 North Virginia Street, MS/0575  
Reno, NV 89557  
Tel:  775 784 7765  
Fax:  775 784 4362  
Email: wyan@medicine.nevada.edu  
URL: http://www.medicine.nevada.edu/yan
ABSTRACT

Exposure to the agricultural fungicide vinclozolin during gestation promotes a higher incidence of various diseases in the subsequent unexposed F3 and F4 generations. This phenomenon is termed epigenetic transgenerational inheritance and has been shown to in part involve alterations in DNA methylation, but the role of other epigenetic mechanisms remains unknown. The current study investigated the alterations in small noncoding RNA (sncRNA) in the sperm from F3 generation control and vinclozolin lineage rats. Over 200 differentially expressed sncRNAs were identified and the tRNA-derived sncRNAs, namely 5’ halves of mature tRNAs (5’ halves), displayed the most dramatic changes. Gene targets of the altered miRNAs and tRNA 5’ halves revealed associations between the altered sncRNAs and differentially DNA methylated regions. Dysregulated sncRNAs appear to correlate with mRNA profiles associated with the previously observed vinclozolin-induced disease phenotypes. Data suggest potential connections between sperm-borne RNAs and the vinclozolin-induced epigenetic transgenerational inheritance phenomenon.
INTRODUCTION

Over a decade ago, the Skinner lab observed that gestating rats (F0 generation) transiently exposed during gonadal sex determination (i.e., embryonic day E8.5–E14.5) to the endocrine disrupter vinclozolin produced offspring (F1 generation) with higher incidences of various diseases, such as increased spermatogenic cell apoptosis and kidney abnormalities, through a mechanism that did not involve genetic inheritance nor relied on mutations of DNA sequences (Anway et al., 2005; Anway et al., 2006). Interestingly, despite the exposure occurring solely in the parent generation (F0), the disease phenotype and frequency persisted in rats of the subsequent generations, including the great and great–great grand offspring (F3 and F4, respectively) (Anway et al., 2005). Although the F1 generation embryo and germ cells of the grand offspring (F2 generation) had been exposed to vinclozolin in utero, the F3 generation rats and their progeny were never directly exposed, but maintained the transgenerational phenotype (Anway et al., 2005; Nilsson and Skinner, 2015; Yan, 2014). This phenomenon is termed “epigenetic transgenerational inheritance,” as the phenotype is transmitted across generations through altered epigenetic information in the germline, without continued direct environmental exposure (transgenerational inheritance) (Anway et al., 2005; Nilsson and Skinner, 2015). Since this initial discovery, a large number of environmental exposures and other endocrine disruptors [e.g., dioxin, bisphenol A (BPA), pesticides DDT and methoxychlor, hydrocarbons, and tributyltin] have also been shown to promote the epigenetic transgenerational inheritance of disease phenotypes (Bruner-Tran and Osteen, 2011; Manikkam et al., 2012; Manikkam et al., 2013; Rissman and Adli, 2014; Manikkam et al., 2014). Although germline DNA methylation has been investigated as a transgenerational
epigenetic mechanism, the role of other epigenetic processes such as ncRNA have not been thoroughly investigated (Yan, 2014).

The vinclozolin-induced transgenerational inheritance of abnormal phenotypes is primarily paternally transmitted (Anway et al, 2005). Therefore, the sperm of vinclozolin lineage rats transmit the altered epigenetic information between generations (Anway et al, 2006). Differential DNA methylated regions (DMRs) have previously been identified in the sperm and male primordial germ cells (PGCs; isolated at E13 and E16) of F3 generation vinclozolin lineage rats (Guerrero-Bosagna et al, 2010; Skinner et al, 2013). A number of different epigenetic mechanisms have been suggested to have a role in epigenetic transgenerational inheritance (Skinner et al, 2014). In addition to transmitting the paternal genome to the zygote, sperm are known to contain a diverse population of RNAs, including small noncoding RNAs (sncRNAs), which are also delivered to the developing embryo following fertilization (Casas and Vavouri, 2014; Sendler et al, 2013; Krawetz et al, 2011; Ostermeier et al, 2004; Jodar et al, 2013). Previous studies have demonstrated that RNAs derived from the sperm of mice with certain phenotypes (e.g., paramutation induced white tail tip), when injected into a fertilized wildtype oocytes, were sufficient to induce the same phenotype in the next generation (Wagner et al, 2008; Rassoulzadegan et al, 2006; Yuan et al, 2015; Gapp et al, 2014). A similar approach was used to show the role of ncRNAs in the transgenerational inheritance of behavioral phenotypes (Gapp et al, 2014). These previous studies have suggested a possible role for sperm-borne ncRNAs in epigenetic transgenerational inheritance (Yan, 2014).

To investigate the relevance of sperm-borne ncRNAs in the context of vinclozolin-induced epigenetic transgenerational inheritance, sncRNA sequencing
(sncRNA-Seq) using sperm of both F3 generation control lineage and vinclozolin lineage rats was performed. In this study, we identified >200 significantly differentially expressed sncRNAs in sperm, which belong to several different classes of sncRNAs. Notably, tRNA 5′ halves, a class of tRNA-derived sncRNAs (tsRNAs), were dramatically up-regulated. The gene targets of the altered miRNAs and tRNA 5′ halves were determined in silico and compared to previously observed mRNA transcriptomes associated with vinclozolin-induced disease phenotypes, as well as the genes proximal to DMRs found in F3 vinclozolin lineage male PGCs and sperm. Both the altered miRNAs and tRNA 5′ halves were predicted to target genes relevant to the vinclozolin transgenerational disease phenotypes, as well as a significant number of genes proximal to the DMRs. Observations suggest a correlation between sperm-borne ncRNA and the vinclozolin-induced epigenetic transgenerational inheritance phenomenon.
RESULTS

Transgenerationally altered sncRNAs in F3 generation sperm

Both the control lineage and vinclozolin lineage F3 generation sperm possessed diverse populations of sncRNAs, consisting mainly of miRNAs, tsRNAs, mitochondrial genome-encoded small RNAs (mitosRNAs), and piRNAs (Figure 4-1a). Using DESeq2, a differential expression analysis software, we compared control lineage and vinclozolin lineage sncRNA expression levels and identified 222 sncRNAs with significantly ($P_{adj} \leq 0.1$) altered expression (Figure 4-1b, Table 4-1). The magnitude of the changes (ratio) is presented in Table 4-S1. Although mostly unchanged, 21 of 251 miRNA observed in sperm ($\sim 8\%$) displayed either up- or down-regulated expression. Similarly, both piRNAs and rRNA-derived sncRNAs appeared to be predominantly unaltered, with only 16% and 14% either up- or down-regulated, respectively (Table 4-1). In male germ cells piRNAs can be classified as “pre-pachytene” piRNAs or “pachytene” piRNAs by their length, timing of expression, and 1st and 10th nucleotide preferences. Pre-pachytene piRNAs are typically 26–28 nt in length and prefer uracil and adenine at their 1st and 10th nucleotides, respectively. The pachytene piRNAs, however, expressed after the pre-pachytene piRNAs, are typically 30 nt long and only have a preference for uracil at their first nucleotide (Ortogero et al., 2014). The majority of the dysregulated piRNAs displayed pachytene piRNA characteristics, indicating that they originate during the later stages of spermatogenesis (Figure 4-S1). Interestingly, 19 out of 24 (79%) mitosRNAs were up-regulated in the vinclozolin lineage sperm. Unlike most other sncRNAs, mitosRNAs do not have a consensus length, and range from 12 to 137 nt depending on
the individual species (Ro et al., 2013). The length distribution of the mitosRNAs was compared to determine whether the processing of full-length mitochondrial RNAs in sperm was altered, but no significant differences were found (data not shown). Although sperm tails are generally removed from the preparation, the possibility that vinclozolin lineage sperm possess more mitochondria than the typical sperm may explain the uniform up-regulation of the mitosRNAs, but remains to be investigated.

Numerous tsRNAs were identified in sperm (Table 4-1), which is consistent with a previous report showing tsRNAs are abundant relative to other sncRNA classes in sperm (Peng et al., 2012). To investigate this further, each sequencing read that aligned to a tRNA gene during annotation was matched to either the 5′ or 3′ half of the each known tRNA. The majority of reads that aligned to the 5′ half of the tRNA was 30–33 nt long, whereas the reads which aligned to the 3′ half had a much wider length distribution (Figure 4-1c). The fragments that were 27 nt or longer were termed 5′ or 3′ halves depending on their origin within the mature tRNA, whereas those less than 26 nt in length were termed tRNA fragments from the 5′ end (tRF-5s) or tRNA fragments from 3′ end (tRF-3s) depending on their side preference (i.e., 5′ or 3′) (Gebetsberger and Polacek, 2013) (Figure 4-1d). Differential expression analysis via DESeq2 revealed that the 5′ halves and tRF-5s were exclusively up-regulated in F3 generation vinclozolin lineage sperm, whereas the 3′ halves and tRF-3s were predominantly down-regulated (Table 4-1). Overall, levels of numerous sncRNAs were changed in vinclozolin lineage F3 sperm, with tsRNAs and mitosRNAs being the most altered ones.
miRNA and 5′ halve tsRNA gene target prediction

Due to their potential relevance to the epigenetic transgenerational inheritance phenomenon (Kiani et al., 2013), we mainly focus further investigations on the 16 altered 5′ halves, as well as the 21 altered miRNAs. The miRNAs are known to regulate gene expression by acting as post-transcriptional regulators (Bartel, 2004). Although less established, various types of tsRNAs, including 5′ halves, have also shown post-transcriptional regulatory abilities (Kumar et al., 2014; Ivanov et al., 2011; Elbarbary et al., 2009; Yamasaki et al., 2009; Wang et al., 2013; Sobala and Hutvagner, 2013). To assess whether the targets of these sncRNAs were relevant to any of the effects of ancestral vinclozolin exposure that had been reported previously (Anway et al., 2005; Anway et al., 2006), the putative sncRNA targets were analyzed in silico. The transgenerationally altered sperm miRNAs were matched to the 3′ untranslated region (UTR) sequences available for rat, utilizing both RNAhybrid and miRanda (Flicek et al., 2014; Rehmsmeier et al., 2004; Kruger and Rehmsmeier, 2006; John et al., 2004). Without an established tRNA halve-mediated silencing mechanism available for 5′ halves, the search was expanded to 5′ UTRs and coding sequences (CDS) in addition to 3′ UTRs. Only RNAhybrid was used, as it accounts for the length of the target sequence, which ensured that there would not be a bias toward the longer (and therefore more prone to random matches) 3′ UTR and CDS over the 5′ UTR (Rehmsmeier et al., 2004). The total number of targets for each transgenerationally altered sperm miRNA and 5′ halves is provided in Table 4-S2. The down-regulated miRNAs had significantly more predicted targets than the up-regulated miRNA, suggesting that they may have more functional significance. The 5′ UTR sequences were targeted by the 5′ halves much more frequently on average.
than the 3’ UTRs and CDS, with the CDS typically having the fewest putative targets (Table 4-S2; Figure 4-2). Around two-thirds of the predicted gene targets identified in the 5’ halve–5’ UTR RNAhybrid analysis were targeted by more than one 5’ halve, whereas roughly 50% and 25% of the 5’ halve–3’ UTR and 5’ halve–CDS gene target predictions, respectively, were targeted more than once (Table 4-S3). This suggests that if sperm 5’ halves do possess silencing capabilities, they likely exercise this ability through the 5’ UTRs of their target mRNAs. As little is known about the potential function of 5’ halves, the gene target predictions based on the 3’ UTR and CDS were still considered in our subsequent analyses.

**Gene Category and Pathway Analysis**

To investigate the putative targets of the transgenerationally altered sperm miRNAs and 5’ halves, a gene ontology (GO) term enrichment analysis was performed on the genes that were targeted by more than one snRNA. The miRNA target genes that had at least two miRNAs were split into redundant up- and down-regulated miRNA target lists. Searching the Database for Annotation, Visualization and Integrated Discovery (DAVID) for enriched biological process (BP), molecular function (MF), cellular component (CC), and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways revealed several significant (Benjamini ≤ 0.05) results for the redundant down-regulated miRNA targets, but none for the up-regulated miRNA (Table 4-S4). An enrichment map of the significant BP genes (Figure 4-3a) shows that the redundantly targeted genes are relevant to apoptosis, early development, kinase activity, regulation of
transcription, and the nervous system (i.e., its development and synaptic signaling).

Significant CC genes also revealed that many of the genes localize in various nervous system cells (e.g., “synapse part”) and the only significant KEGG pathway identified was “axon guidance” (Table 4-S4).

A similar analysis was performed on the predicted gene targets of the transgenerationally altered 5′ halve tsRNAs using genes that were targeted by at least three 5′ halves. The results of the 5′ UTR−, CDS−, and 3′ UTR−5′ halve analyses were analyzed separately. For every GO term category analyzed, the 5′ UTR-derived redundant gene targets had the largest number of significantly enriched terms (Table 4-S5). Apoptosis, neuronal development, regulation of metabolic processes, phosphorylation, cell cycle, and transport were the predominant gene categories of the significantly enriched BP terms (Figure 4-3b). In terms of significance, endocytosis, neurotrophin signaling pathway, and oocyte meiosis were the top enriched KEGG pathways (Table 4-S5). A portion of the enriched terms for the redundant targets of the down-regulated miRNAs were also related to apoptosis and nervous system development, suggesting that the alterations of miRNAs and 5′ halves may play similar roles in vinclozolin-induced epigenetic transgenerational inheritance.

**Correlation with Vinclozolin-Induced Transgenerational Sperm DNA Methylation Alterations**

A correlation between the transgenerationally altered sperm miRNAs and 5′ halves observed was made with the F3 generation vinclozolin sperm differential DMRs previously described (Guerrero-Bosagna *et al.*, 2010; Skinner *et al.*, 2015). The majority
of transgenerational DMR is not associated with genes and is intergenic (Skinner et al, 2015). Of the 21 miRNAs and 16 5′ halves that were found to be differentially expressed, 16 and 14, respectively, originated within 5 Mb of a sperm DMR (Table 4-2A, -2B). Previously the DMR and regulated genes have been shown to exist within 2–5 Mb regions termed an epigenetic control region (ECR) (Skinner et al, 2012), so a 5 Mb region was investigated. The genes proximal (i.e., up to 100 kb up and downstream) to the promoter associated DMRs identified previously in F3 generation vinclozolin lineage sperm were also investigated (Guerrero-Bosagna et al, 2010; Skinner et al, 2015). The correlation between the predicted targets of the altered miRNAs and 5′ halves (5′ UTR and 3′ UTR matching results only) and genes proximal to the DMRs was significant ($P$ value < 0.1) (Table 4-3). However, the number of actual matches of predicted gene targets to DMR proximal genes was consistently lower than the number of expected random matches for the dysregulated miRNA, signifying a negative correlation (Table 4-S6). In contrast, the majority of actual matches observed for the putative 5′ halve targets (i.e., 5′ UTR and 3′ UTR only) was above the number of random matches, suggesting that 5′ halves possess the potential to post-transcriptionally regulate genes proximal to DMRs present in the sperm of F3 vinclozolin lineage rats (Table 4-S6).

An additional experiment examined the correlations with the transgenerational DMRs in the PGCs, which are the stem cells for the differentiated sperm cell. The previously identified DMRs in PGCs isolated from F3 generation vinclozolin lineage fetal testis at embryonic days E13 and E16 of development were analyzed (Skinner et al, 2013). Although the transgenerational DMRs in the E13 and E16 PGCs would not be
directly relevant to the sperm DMRs, they are associated with the precursor cells of the sperm so correlations may be present. Initially, we investigated whether any of the transgenerationally altered miRNAs or 5’ halves originated near the E13 or E16 DMRs, but found very few correlations, relative to the sperm DMRs (Table 4-2A, -2B). Next, we compared the genes proximal to the DMRs and the predicted gene targets of the altered miRNA and 5’ halves. The genes proximal to the E13 PGC DMRs significantly correlated (P value < 0.1) with the predicted 5’ halve targets, for the predictions from both the 5’ UTR and 3’ UTR-based matching (Tables 4-4A and 4-S7A). When we analyzed the E16 PGC DMRs, we again observed a significant correlation (P value < 0.1) between the 5’ halves and proximal genes, this time just for the results of the 5’ UTR-based matching (Tables 4-4B and 4-S7B). There was a less significant correlation (P value of 0.070 vs. 0.015 for the 5’ halves) between the predicted targets of the down-regulated miRNAs and E16 DMR proximal genes as well (Tables 4-4B and 4-S7B). Combined observations suggest an association between the transgenerationally altered DNA methylation regions in F3 generation male PGCs and sperm with the sperm-borne sncRNA expression.
DISCUSSION

Our data demonstrate that many sncRNAs were transgenerationally altered in the sperm of F3 generation vinclozolin lineage rats, tsRNAs in particular. The potential gene targets of both the altered miRNAs and 5’ halves were identified in order to determine whether the expression changes of these sncRNAs had any functional significance to the previously observed pathology of ancestral vinclozolin exposure.

Perhaps, the most striking vinclozolin-induced transgenerational effect on the sperm sncRNA transcriptome was the alteration of the majority of observed tsRNAs. Approximately 82% of the tsRNAs in the vinclozolin lineage F3 generation sperm were altered and ~93% of the reads were attributed to 5’ halves. Comparing this to the other altered tsRNAs, which accounted for 81% (3′ halves), 65% (tRFs-5), and 49% (tRFs-3) of their respective species, we see that 5′ halves were dramatically altered in the transgenerational vinclozolin lineage sperm. This is especially noteworthy, given that 5′ halves are known to be highly enriched in mature sperm (Peng et al., 2012). In the vinclozolin lineage sncRNA-Seq results, 5′ halves were the most abundant sncRNAs and accounted for ~25% of all the annotated sncRNAs.

The gene categories that were significantly enriched in the lists of redundantly targeted genes for the down-regulated miRNA and the 5′ halve–5′ UTR targets were correlated to several of the transgenerational disease phenotypes. In the enrichment maps of both the altered miRNAs and 5′ halves, apoptosis was one of the largest clusters of enriched terms (Figure 4-3a, -3b). An increase in spermatogenic cell apoptosis was observed previously in 90% of vinclozolin males, regardless of generation (Anway et al,
Gene categories and KEGG pathways related to the development and regulation of neurons were highly enriched within the lists of redundantly targeted genes for both the miRNAs and 5’ halves (Figure 4-3a, -3b; Tables 4-S4 and 4-S5). Several of the phenotypes associated with vinclozolin-induced epigenetic transgenerational inheritance of disease involve the nervous system and brain development, despite no discernable morphological changes to the brain (Anway et al., 2006; Skinner et al., 2014; Skinner et al., 2008; Gillette et al., 2015; Crews et al., 2012). In a 2008 study by the Skinner lab, many neural-pathways were found to contain differentially expressed genes in the amygdala and hippocampus of F3 vinclozolin lineage male and female rats, including axon guidance, which was the only significant KEGG pathway identified for the down-regulated miRNA (Table 4-S4) (Skinner et al., 2014; Skinner et al., 2008). In the same study, F3 generation vinclozolin male and female rats exhibited an increase and decrease in anxiety-like behaviors, respectively (Skinner et al., 2008). Other changes in behavior and the brain transcriptome of F3 vinclozolin rats have been observed in subsequent studies (Skinner et al., 2014; Gillette et al., 2015; Crews et al., 2012). It is possible that aberrant gene targeting by the altered sperm miRNA and 5’ halves during early development could cause alterations in the developing brain and nervous system, ultimately manifesting as the transgenerational disease phenotypes described above.

Interestingly, an increase in 5’ halve expression has previously been linked to altered neuron survival and anxiety-related behaviors. Mice deficient in NSun2 (NSun2−/−), an RNA methyltransferase known to modify tRNAs, possessed elevated levels of 5’ halves in their skin and brain. This correlated with embryonic brain
defects (of the cerebral cortex, hippocampus, and striatum) and, as was also seen in F3 generation vinclozolin lineage male rats, a reduction in anxiety-related behaviors (Skinner et al, 2008; Blanco et al, 2014). *NSun2*−/− male mice are infertile as well, due to the arrest of spermatogenesis prior to the meiotic progression into the pachytene stage (Hussain et al, 2013). Therefore, it should come as no surprise that up-regulation of 5′ halves would correlate to the similar, albeit less severe, vinclozolin lineage transgenerational disease phenotype.

The ability of 5′ halves to act as translational suppressors has been demonstrated in other organisms (Ivanov et al, 2011; Elbarbary et al, 2009; Yamasaki et al, 2009; Wang et al, 2013). In one study, certain 5′ halves inhibited translation by displacing eIF4G/eIF4A from the 5′ end of mRNAs (Ivanov et al, 2011). Based on the results of the RNAhybrid analysis, 5′ halves appear to preferentially bind to the 5′ UTRs of mRNAs, relative to the CDS and 3′ UTRs. The ability of 5′ halves to bind complementary sequences within mRNAs, and subsequently silence their expression has also been demonstrated (Elbarbary et al, 2009; Wang et al, 2013). Therefore, in some cases, the altered expression of 5′ halves could potentially regulate gene expression by binding to the 5′ UTRs and displacing cap-bound complexes and hindering translation of the target mRNA.

A correlation between the previously identified differential DMRs in the vinclozolin-induced transgenerational sperm with the altered sncRNA identified in this study demonstrated that ~88% and ~76% of the differentially expressed 5′ halves and miRNAs, respectively, originate within 5 Mb of a sperm DMR (Table 4-2A, -2B). Our correlation analysis also revealed that of the genes predicted to be targeted by the altered
5′ halves via their 5′ or 3′ UTRs, a significant number were within at least 100 kb of a sperm DMR (Table 4-3). Previously, ECRs were identified in somatic tissues of vinclozolin lineage rats, spanning 2–5 Mb (Skinner et al, 2012). Based on these findings, it seems plausible that the altered 5′ halves are components of various ECRs, and that their differential expression is either due to the presence of the DMRs themselves, in response to other genes proximal to the DMRs, or a combination of both. In addition to sperm, a correlation between the differentially expressed sncRNAs and DMRs previously found in the E13 and E16 PGCs of vinclozolin lineage rats was made (Skinner et al, 2013). Although relatively fewer 5′ halves (Table 4-2A) and miRNAs (Table 4-2B) were expressed within 5 Mb of a PGC DMR, we did observe correlations between the gene targets of the sncRNAs and genes proximal to the DMRs. There were a significant number of putative 5′ halve gene targets that were within 100 kb of an E13 (Table 4-4A) or E16 (Table 4-4B) PGC DMR. A significant number of predicted gene targets of the down-regulated miRNA also matched genes proximal to the E16 PGC DMRs (Table 4-4B). It is possible that the altered expression of these miRNAs and 5′ halves during spermatogenesis is also influenced by dysregulation of the genes proximal to DMRs in PGCs. However, based on our observations, it seems that epigenetic alterations present at later stages of male germ cell development (i.e., sperm) play a potentially greater role in the differential expression of sperm-borne miRNA and 5′ halves. Future studies that assess the methylation and gene expression of vinclozolin lineage male germ cells, at several different stages of development, will be informative.

Observations show that ancestral exposure to vinclozolin causes transgenerational (i.e., F3 generation) alterations in the sncRNA expression levels in mature sperm. The
altered miRNAs and 5′ halve tsRNA potential target genes were correlated to several of the previous observed vinclozolin transgenerational disease phenotypes. Interestingly, a dramatic up-regulation of 5′ halves was observed in the vinclozolin lineage sperm (Table 4-1). In two other examples of epigenetic inheritance, the Kit and Sox9 paramutations, it was determined that the methyltransferase Dnmt2 was necessary for the transmission of the phenotype associated with the paramutation (Kiani et al., 2013). Dnmt2, like NSun2, confers stability to mature tRNA via cytosine-5 methylation and its loss affects 5′ halve production (Blanco et al., 2014; Schaefer et al., 2010; Tuorto et al., 2012; Hussain et al., 2013). Because of this, tsRNAs, specifically 5′ halves, have been postulated to be effectors of epigenetic inheritance (Kiani et al., 2013; Kiani and Rassaoulzadegan, 2013). Our discovery that 5′ halves were abnormally enriched in transgenerational vinclozolin lineage sperm, and potentially correlated to the vinclozolin lineage transgenerational disease phenotype, provides additional evidence to the idea that sperm-borne 5′ halves act as epigenetic regulators. A new study in progress will determine whether similar observations involving an up-regulation of 5′ halves develop in another model of environmentally induced (e.g., DDT-induced) epigenetic transgenerational inheritance of disease. Overall, our findings highlight the potential importance of sperm-borne sncRNAs in the mechanism of vinclozolin-induced epigenetic transgenerational inheritance, as well as support the proposed role of 5′ halves as active epigenetic regulators during early development.
ACKNOWLEDGEMENTS

This work was supported, in part, by grants from the NIH (ES012974 to M.K.S.) and the Templeton Foundation (PID: 50183 to M.K.S. and W.Y.). Sequencing was conducted in the Single Cell Genomics Core supported, in part, by a NIH Grant (P20RR18751). MKS and WY conceived the study; AS conducted the experiments; AS, WY and MKS analyzed the data; AS, WY and MKS wrote the manuscript. MKS and WY contributed equally to the study. Correspondence and requests for materials should be addressed to skinner@wsu.edu or wyan@medicine.nevada.edu. The authors declare no competing financial interests.
MATERIALS AND METHODS

Animal studies and breeding

Female and male rats of an outbred strain Hsd:Sprague Dawley (Harlan) at 70–100 days of age were fed *ad lib* with a standard rat diet and *ad lib* tap water for drinking. To obtain time-pregnant females, the female rats in proestrus were pairmated with male rats. The sperm-positive (Day 0) rats were monitored for diestrus and body weight. On Days 8 through 14 of gestation, the six different females per group were administered daily intraperitoneal injections of vinclozolin (100 mg/kg BW/day) or dimethyl sulfoxide (vehicle). The vinclozolin was obtained from Chem Service Inc. (West Chester PA, USA) and was injected in a 200 µl DMSO/sesame oil vehicle as previously described (Yan, 2014). Treatment lineages are designated “control” or “vinclozolin” lineages. The gestating female rats treated were designated as the F0 generation. The offspring of the F0 generation rats were the F1 generation. Nonlittermate females and males aged 70–90 days from F1 generation of control, or vinclozolin lineages were bred to obtain F2 generation offspring. The F2 generation rats were bred to obtain F3 generation offspring. As epigenetic transgenerational inheritance involves parent of origin germline epigenetic information, the optimal phenotype is obtained through breeding within the lineage, so the control and vinclozolin lineages used intrabreeding within the lineage to obtain the F3 generation. No sibling or cousin breeding was used to avoid any inbreeding artifacts. Only the F0 generation gestating female was directly treated transiently with vinclozolin. The control and vinclozolin lineages were housed in the same room and racks with lighting, food and water as previously described (Skinner *et al*, 2015). The epididymal
sperm was collected from 1-year-old adult males and three pools of three males \( (n = 9) \) used for the analysis, with all males being from different litters for both the control and vinclozolin lineages. All experimental protocols for the procedures with rats were preapproved by the Washington State University Animal Care and Use Committee (IACUC approval # 02568).

**RNA isolation**

The F3 generation vinclozolin and control lineage male epididymal sperm were collected and processed as previously described and stored at \(-70^\circ \text{C}\) until use (Guerrero-Bosagna *et al.*, 2010). The total RNA was isolated using the mirVana miRNA Isolation Kit (Life Technologies) following the manufacturer’s instructions with modifications at the lysis stage. In brief, after addition of lysis buffer, the frozen sperm pellets were homogenized at low settings for 90 s, followed by a 5-min incubation at 65°C. Samples were then placed on ice, and the default protocol was resumed. For quality control, RNA integrity numbers (RIN) were obtained by RNA 6000 Nano chips run on an Agilent 2100 Bioanalyzer (Agilent). A RIN of 2–4 indicates good sperm RNA quality.

**Sequencing**

Prior to library preparation, total sperm RNA samples were enriched for small RNAs using the protocol provided in the Ion RNA-Seq Kit v2 (Life Technologies). Small RNA-enriched samples were used for small RNA library preparation, using the same kit, and barcoded with Ion Xpress Barcode Adapters (Life Technologies). Quality control was performed using Agilent High Sensitivity chips (Agilent). Libraries were loaded onto
the same Ion PI chip via the Ion PI Template OT2 200 v3 and Ion PI Sequencing 200 v3 kits, and sequenced on an Ion Proton Sequencer (Life Technologies).

Bioinformatics

The sncRNA-Seq data were annotated as follows: reads shorter than 15 nt were discarded. The remaining reads were matched to known rat sncRNA, consisting of mature miRNA (miRBase, release 21), tRNA (Genomic tRNA Database, rn5), piRNA (piRNABank), rRNA (ENSEMBL, release 76), and mitochondrial RNA (ENSEMBL, release 76) using Sequery (0–2 mismatches allowed) (Flicek et al, 2014; Kozomara et al, 2014; Chan and Lowe, 2009; Sai and Agrawal, 2008; Ortogero et al, 2013). Unmatched reads were matched to mouse testis (Song et al, 2011) and sperm endo-siRNA (Yuan et al, 2016) with Sequery (0 mismatches allowed). The remaining unmatched reads were aligned to the rat genome (rn5) via Bowtie (settings -n 2 -k 3 –best -S –al -q) (Langmead et al, 2009). Aligned reads were matched to the genomic coordinates of known rat mature miRNA, tRNA, rRNA, and mitochondrial RNA (same databases used previously). Read counts were obtained by in-house Python scripts. Unnormalized read counts were used for differential expression analysis via DESeq2 ($P_{adj} \leq 0.1$) (Love et al, 2014).

Reads initially matched to tRNA were extracted and matched to the 5’ and 3’ halves of the full length tRNA (split at the 3’ end of the anticodon), via Sequery (0–2 mismatches allowed) (Ortogero et al, 2013). Reads $\geq 27$ nt and $\leq 26$ nt were referred to as halves and tRFs, respectively, and further classified as 5’ halves, tRF-5s, 3’ halves, and tRF-3s based on whether the read aligned to the 5’ or 3’ half of the tRNA (Figure 4-1d).
Transgenerationally altered miRNA gene targets were predicted using RNAsHybrid (settings -n 50 -m 50000 -c -d 2.29,0.18 -p 0.05 -e -20) and miRanda (default settings; score ≥ 140 and energy ≤ −20) against rat 3′ UTR sequences (ENSEMBL, release 76), discarding predictions not made by both programs (Flicek et al., 2014; Rehmsmeier et al., 2004; Kruger and Rehmsmeier, 2006; John et al., 2004). Dysregulated 5′ halve gene targets were predicted using RNAsHybrid (settings -n 50 -m 50000 -c -d xi, θ -p 0.01 -e -20) against rat 5′ UTR, CDS, and 3′ UTR sequences (ENSEMBL, release 76) using xi,θ values (3.07,0.28), (3.00,0.27), and (2.85,0.25), respectively. The xi,θ values were obtained using RNAcalibrate and known 5′ halve sequences (tRFdb) or mature miRNA (miRBase, release 21) (Rehmsmeier et al., 2004; Kozomara et al., 2014; Kumar et al., 2015).

GO term enrichment analysis was performed using DAVID and the categories GOTERM_BP_FAT, GOTERM_MF_FAT, GOTERM_CC_FAT, and KEGG pathways (Dennis et al., 2003). Benjamini score ≤0.05 was used as the significance cutoff.

The significance of the number of observed matches between the predicted gene targets and genes proximal to DMRs (Tables 4-2 and -3, Table 4-S5) was determined using a Chi-square test. The observed number of matches was compared to the expected number of random matches, defined as the number of proximal genes multiplied by the percentage of all known rat coding genes targeted by the dysregulated sncRNA. The adjusted “\( P_{\text{adj}} \)” values are from DESeq2, the differential expression analysis program which adjusts the \( P \)-values to account for multiple testing using the Benjamini–Hochberg procedure, keeping the amount of Type I errors under control. The significance threshold was set at 0.1 for the DESeq2 differential expression and Chi-square analyses.
Figure 4-1 – Differentially expressed sncRNAs in vinclozolin lineage F3 sperm. (a) The relative amount of each sncRNA class found in the control lineage (CL; left) and vinclozolin lineage (VL; right) F3 sperm. Sequencing reads which aligned to tRNA genes were classified as 5’ halves, tRF-5, 3’ halves, or tRF-3 based on their length and tRNA half preference. Those tRNA-matching reads that did not fall into the aforementioned categories were classified as “Other.” (b) Heatmap of the sncRNA expression changes in VL F3 sperm relative to CL F3 sperm. The log\(_2\) fold changes shown were obtained from DESeq2. Genes with \(P_{\text{adj}}\) values of “NA” were not included in the heatmap. (c) tsRNA length distribution. Fragments were separated into the groups 5’ half or 3’ half depending on which side they matched to on their mature tRNA of origin. (d) tsRNA types. tsRNA
that were approximately half the length of the mature tRNA (left schematic side derived from Gebetsberger and Polacek, 2013) were classified as 5’ halves or 3’ halves, depending on the side of the anticodon loop to which they matched. Shorter fragments (>27 nt) were classified as tRF-5s or tRF-3s depending on their origins in the mature tRNA.
Figure 4-2 – Relative numbers of predicted 5′ halve gene targets in 5′ UTR, CDS, and 3′ UTR sequences. The number of predicted gene targets for each differentially expressed 5′ halve species within the 5′ UTR, CDS, and 3′ UTR was normalized to the number of targets within the 5′ UTR, for each 5′ halve species. Relative to CDS and 3′ UTR, 5′ UTR sequences had the highest incidence of RNAhybrid predicted 5′ halve binding sites. A \textit{P}-value ≤0.01 was used as a cutoff.
Figure 4-3 – GO term analyses of dysregulated sncRNAs in the vinclozolin lineage F3 sperm. (a) Enrichment map of the BP terms associated with redundantly predicted gene targets of down-regulated miRNAs. The genes which were predicted targets of at least two down-regulated miRNAs were analyzed via DAVID for significantly enriched BP terms. The results of the DAVID analysis were used to generate an enrichment map, grouping related BP terms into clusters. The text overlaying each cluster describes the unifying “theme” of that cluster (chosen by the authors). The individual BP terms are described in more detail in Table 4-S4. (b) Enrichment map of the BP terms associated with redundantly predicted gene targets of 5’ halves. The genes which were predicted targets of at least three dysregulated 5’ halves were analyzed via DAVID for significantly enriched BP terms. A Benjamini value ≤0.05 was used to define significance. The results of the DAVID analysis were used to generate an enrichment map, grouping related BP
terms into clusters. The text overlaying each cluster describes the unifying “theme” of that cluster (chosen by the authors). The individual BP terms are described in more detail in Table 4-S5.
### TABLES

<table>
<thead>
<tr>
<th>RNA type</th>
<th>Up-regulated genes (no. of genes)</th>
<th>Down-regulated genes (no. of genes)</th>
<th>Unchanged genes (no. of genes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mature miRNA</td>
<td>13</td>
<td>8</td>
<td>230</td>
</tr>
<tr>
<td>rRNA</td>
<td>1</td>
<td>8</td>
<td>54</td>
</tr>
<tr>
<td>piRNA</td>
<td>16</td>
<td>68</td>
<td>513</td>
</tr>
<tr>
<td>Mitochondrial RNA</td>
<td>19</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>5’ halves</td>
<td>16</td>
<td>0</td>
<td>14</td>
</tr>
<tr>
<td>tRF-5</td>
<td>26</td>
<td>0</td>
<td>25</td>
</tr>
<tr>
<td>3’ halves</td>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>tRF-3</td>
<td>1</td>
<td>43</td>
<td>37</td>
</tr>
</tbody>
</table>

**Table 4-1.** Differential expressed sncRNAs in vinclozolin lineage F3 sperm. The number of sncRNA species, classified by RNA type, that were significantly \( P_{\text{adj}} \leq 0.1 \) up- or down-regulated in the VL F3 sperm sncRNA-Seq data, relative to the CL data. SncRNA species with a \( P_{\text{adj}} \leq 0.1 \) were termed “unchanged”; the species with “NA” \( P_{\text{adj}} \) values were not included in this Table.
<table>
<thead>
<tr>
<th>DMRs (cell type)</th>
<th>No. of DMRs</th>
<th>Distance from DMR (Mb)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>(A) 5′ halves</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F3 E13 PGC</td>
<td>23</td>
<td>0</td>
</tr>
<tr>
<td>F3 E16 PGC</td>
<td>13</td>
<td>0</td>
</tr>
<tr>
<td>F3 Sperm</td>
<td>188</td>
<td>0</td>
</tr>
<tr>
<td>(B) miRNA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F3 E13 PGC</td>
<td>23</td>
<td>0</td>
</tr>
<tr>
<td>F3 E16 PGC</td>
<td>13</td>
<td>0</td>
</tr>
<tr>
<td>F3 Sperm</td>
<td>188</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 4-2. Proximity of differential expressed sncRNAs to DMRs. The genomic coordinates of the dysregulated (A) 5′ halves and (B) miRNA were compared to the locations of DMRs previously identified in F3 sperm, E13 PGCs, and E16 PGCs. The genes of sncRNA that were within 1–5 Mb of at least one of the DMR in its respective cell type were counted above. The sncRNA whose genes were contained within a DMR (0 Mb) were also counted.
<table>
<thead>
<tr>
<th>Distance from DMR (kb)</th>
<th>No. of proximal genes</th>
<th>miRNA–3’ UTR</th>
<th>5’ halves–5’ UTR</th>
<th>5’ halves–CDS</th>
<th>5’ halves–3’ UTR</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>65</td>
<td>14&lt;sup&gt;a&lt;/sup&gt;</td>
<td>28</td>
<td>15</td>
<td>28</td>
</tr>
<tr>
<td>1</td>
<td>74</td>
<td>15&lt;sup&gt;a&lt;/sup&gt;</td>
<td>34</td>
<td>19</td>
<td>29</td>
</tr>
<tr>
<td>2</td>
<td>83</td>
<td>16&lt;sup&gt;a&lt;/sup&gt;</td>
<td>36</td>
<td>21</td>
<td>32</td>
</tr>
<tr>
<td>5</td>
<td>109</td>
<td>17&lt;sup&gt;a&lt;/sup&gt;</td>
<td>40</td>
<td>30</td>
<td>37</td>
</tr>
<tr>
<td>10</td>
<td>129</td>
<td>20&lt;sup&gt;a&lt;/sup&gt;</td>
<td>48</td>
<td>38</td>
<td>43</td>
</tr>
<tr>
<td>20</td>
<td>181</td>
<td>26&lt;sup&gt;b&lt;/sup&gt;</td>
<td>66</td>
<td>52</td>
<td>56</td>
</tr>
<tr>
<td>50</td>
<td>313</td>
<td>40&lt;sup&gt;b&lt;/sup&gt;</td>
<td>101</td>
<td>85</td>
<td>90</td>
</tr>
<tr>
<td>100</td>
<td>532</td>
<td>66&lt;sup&gt;c&lt;/sup&gt;</td>
<td>150</td>
<td>141</td>
<td>139</td>
</tr>
<tr>
<td></td>
<td>&lt;i&gt;P&lt;/i&gt;-value</td>
<td>0.011</td>
<td>0.064</td>
<td>0.66</td>
<td>0.061</td>
</tr>
</tbody>
</table>

**Table 4-3.** Correlation between genes proximal to F3 sperm DMRs and predicted gene targets. The genes up and downstream of the previously identified F3 sperm DMRs were compared to the predicted gene targets of the dysregulated miRNA and 5’ halves. <i>P</i>-values were determined by a Chi-square test.

<sup>a</sup>Two of the matching genes were predicted to be targeted by both up and down-regulated miRNA.

<sup>b</sup>Three of the matching genes were predicted to be targeted by both up and down-regulated miRNA.

<sup>c</sup>Four of the matching genes were predicted to be targeted by both up and down-regulated miRNA.
<table>
<thead>
<tr>
<th>Distance from DMR (kb)</th>
<th>No. of proximal genes</th>
<th>miRNA–3′ UTR</th>
<th>5′ halves–5′ UTR</th>
<th>5′ halves–CDS</th>
<th>5′ halves–3′ UTR</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A) F3 E13 PGC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>15</td>
<td>2</td>
<td>9</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>1</td>
<td>18</td>
<td>2</td>
<td>10</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>24</td>
<td>4</td>
<td>12</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>5</td>
<td>31</td>
<td>5</td>
<td>18</td>
<td>11</td>
<td>12</td>
</tr>
<tr>
<td>10</td>
<td>37</td>
<td>5</td>
<td>19</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>20</td>
<td>51</td>
<td>8&lt;sup&gt;a&lt;/sup&gt;</td>
<td>24</td>
<td>16</td>
<td>19</td>
</tr>
<tr>
<td>50</td>
<td>94</td>
<td>15&lt;sup&gt;a&lt;/sup&gt;</td>
<td>43</td>
<td>22</td>
<td>37</td>
</tr>
<tr>
<td>100</td>
<td>148</td>
<td>28&lt;sup&gt;a&lt;/sup&gt;</td>
<td>67</td>
<td>38</td>
<td>62</td>
</tr>
<tr>
<td>P-value</td>
<td>0.97</td>
<td>6.79E−05</td>
<td>0.53</td>
<td>0.019</td>
<td></td>
</tr>
<tr>
<td>(B) F3 E16 PGC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>7</td>
<td>2</td>
<td>5</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>9</td>
<td>3</td>
<td>5</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>9</td>
<td>3</td>
<td>5</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>16</td>
<td>6</td>
<td>8</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>10</td>
<td>22</td>
<td>9</td>
<td>12</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>20</td>
<td>32</td>
<td>9</td>
<td>16</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>50</td>
<td>57</td>
<td>11</td>
<td>23</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>100</td>
<td>93</td>
<td>19</td>
<td>38</td>
<td>23</td>
<td>26</td>
</tr>
<tr>
<td>P-value</td>
<td>0.070</td>
<td>0.015</td>
<td>0.90</td>
<td>0.97</td>
<td></td>
</tr>
</tbody>
</table>

**Table 4.4.** Correlation between genes proximal to F3 PGC DMRs and predicted gene targets. The genes up and downstream of the previously identified F3 E13 (A) and E16 (B) PGC DMRs were compared to the predicted gene targets of the dysregulated miRNA and 5′ halves. *P*-values were determined by a Chi-square test.

<sup>a</sup>One of the matching genes was predicted to be targeted by both up and down-regulated miRNA.
SUPPLEMENTAL INFORMATION

Supplemental figures and tables can be found at:

http://eep.oxfordjournals.org/content/suppl/2016/03/09/dvw001.DC1
REFERENCES


Kumar et al. 2014b. Meta-analysis of tRNA derived RNA fragments reveals that they are evolutionarily conserved and associate with AGO proteins to recognize specific RNA targets. *BMC Biol.*, 12:78.


CHAPTER V
Conclusions

Over the years since their discovery, our understanding of sperm-borne RNAs has undergone several revisions. The initial discovery of RNA in sperm was met with skepticism due to the transcriptionally quiescent state of the cell from elongation onward (Rathke et al, 2014; Kierszenbaum and Tres, 1975; Krawetz, 2005; Miller, 2014). After several years of research, the existence of true sperm-borne RNAs was not only confirmed, but found to be much more complex than originally thought (Krawetz, 2005; Krawetz et al, 2011; Miller and Ostermeier, 2006; Sendler et al, 2013; Jodar et al, 2013). With the controversy surrounding the presence of RNAs in the transcriptionally inert sperm largely behind us, there is now a debate over the relative significance of these RNAs (e.g., small noncoding RNAs, sncRNAs) during fertilization and early embryogenesis. The question of whether the presence of paternal sncRNAs in the embryo holds any physiological significance has shifted to discussions over just how important they are in early development. The research described in this dissertation seeks to further elucidate the putative roles of sperm-borne sncRNAs in the early embryo.

Sperm-borne miRNA and endo-siRNA were found to be non-essential during fertilization, yet critical for preimplantation embryonic development (Chapter II). By utilizing Drosha and Dicer conditional knockout (cKO) mice, sperm that were deficient in miRNA and miRNA / endo-siRNA, respectively, were generated for development studies. In addition to examining how the absence of paternally-delivered miRNA and
endo-siRNA affected embryogenesis, populations of wild type (WT) sperm RNA were tested for their ability to rescue developmental potential in Drosha or Dicer cKO fertilized oocytes. The improvement in developmental outcomes observed in WT sperm RNA-supplemented embryos further supports the significance of these two classes of sncRNA during early development (Yuan et al, 2016).

Before studying the potential functions of sperm-borne endo-siRNA in the early embryo, novel sperm endo-siRNA species had to be identified in silico (Schuster et al, 2015, see Chapter SI). The lack of endo-siRNA annotation in sperm is not surprising – sperm RNAs are notoriously difficult to work with, due to their morphology and relatively low levels of RNA compared to somatic cell types (Jodar et al, 2013; Goodrich et al, 2013; Goodrich et al, 2007). To address the issues currently hindering the study of sncRNAs in sperm, we developed an RNA isolation methodology that is applicable to many different animals; the lysis stage is the only part of the protocol that may necessitate changes to suit a particular species (Chapter III; Table 3-1). This method was used to generate the large and small RNA expression data for SpermBase, thus providing a resource for other labs to utilize in their own studies. Employing a consistent methodology further permitted a cross-species comparison (i.e., mouse, rat, human, and rabbit), revealing some conserved features of mammalian sperm-borne sncRNA (Mao et al, 2013; Mao et al, 2014; Jodar et al, 2013; Goodrich et al, 2013; Sendler et al, 2013; Gilbert et al, 2007; Varner and Johnson, 2007; Das et al, 2010; Shafeeque et al, 2014).

One of the conserved characteristics of sperm-borne sncRNA populations observed in Chapter III was the abundance of tRNA-derived small RNAs (tsRNAs), particularly 5’ halves. Previous studies have implicated 5’ halves and miRNA in paternal epigenetic
inheritance, prompting our investigation into whether they play any role in vinclozolin-induced epigenetic transgenerational inheritance (ETI) (Chapter IV) (Kiani et al., 2013). Altered 5’ halve and miRNA profiles were observed in the sperm of rats indirectly-exposed (i.e., F3 generation) to vinclozolin. Coupled with the putative, developmentally-relevant functions established for these altered 5’ halves and miRNAs, it appears that sperm-borne sncRNA may be able to stably transmit phenotypes across several generations (Schuster et al., 2016).

While our understanding of sperm-borne sncRNA in the context of early development has improved dramatically over the past decade, many questions still remain. If sperm-borne sncRNAs are delivered to the oocyte, how long do they persist and remain active (Ostermeier et al., 2004)? What is the mechanistic basis, if any, for their apparent importance in epigenetic inheritance (Sharma et al., 2016; Chen et al., 2016; Schuster et al., 2016; Liebers et al., 2014; Rassoulzadegan and Cuzin, 2015)? What factors determine which sncRNAs are retained in the sperm during spermiogenesis (Krawetz et al., 2011; Miller and Ostermeier, 2006; Sendler et al., 2013; Jodar et al., 2013)? To address these questions, we must continue to improve our experimental approaches to sperm RNA research (e.g., more efficient RNA isolation, single sperm sequencing, etc). The relatively late discovery (Peng et al., 2012) of the overwhelmingly dominant presence of 5’ halves in sperm underscores the need for further research and emphasizes the importance of thoroughly exploring unconventional and novel features of sperm-borne sncRNA.
REFERENCES


Varner and Johnson. 2007. From a sperm’s eye view – revisiting our perception of this intriguing cell. Proc. 53rd Annual Conv. Am. Assoc. Equine Practitioners (Orlando, FL USA)

Yuan et al. 2016. Sperm-borne miRNAs and endo-siRNAs are important for fertilization and preimplantation embryonic development. Development, 143, 635 – 647.
SUPPLEMENTARY CHAPTER I

In Silico Identification of Novel Endo-siRNAs

Andrew Schuster¹, Grant W Hennig¹, Nicole Ortogero¹, Dickson Luong¹, Wei Yan¹,*

¹Department of Physiology and Cell Biology, University of Nevada School of Medicine, 1664 North Virginia Street, MS0575, Reno, NV 89557, USA

This chapter can be found in the following publication:


*Corresponding author:

Wei Yan M.D., Ph.D.
Professor
University of Nevada School of Medicine
Center for Molecular Medicine, Room 207B
1664 North Virginia Street, MS/0575
Reno, NV 89557
Tel: 775 784 7765
Fax: 775 784 4362
Email: wyan@medicine.nevada.edu
URL: http://www.medicine.nevada.edu/yan
ABSTRACT

Many classes of small noncoding RNAs (sncRNAs), such as microRNAs (miRNAs) and endogenous small interfering RNAs (endo-siRNAs), have been identified as important regulators of gene expression. Endo-siRNAs represent an integral part of the endogenous RNAi pathway and have been identified in multiple organisms and cell types. Wide adoption of the next-generation deep sequencing (NGS)-based sncRNA profiling has made the identification of novel sncRNA species more accessible. However, it remains a challenge to identify novel endo-siRNAs that are not collected in the current endo-siRNA databases. We have developed an in silico method for identification of novel endo-siRNAs using small RNA NGS data. Here, we describe our protocol in detail.
INTRODUCTION

Since the discovery of microRNAs (miRNAs) in 2001 (Lagos-Quintana et al., 2001; Lau et al., 2001), many more non-coding RNA species, e.g., PIWI-interacting RNAs (piRNAs) (Aravin et al., 2006), endogenous small interference RNAs (endo-siRNAs) (Ambros et al., 2003; Llave et al., 2002), piRNA-like RNAs (piRNAs) (Ro et al., 2007), and mitochondrial small RNAs (mitosRNAs) (Lung et al., 2006; Ro et al., 2013) have been identified and characterized. Endo-siRNAs appear to function in the endogenous RNA interference (RNAi) pathway found in numerous organisms (e.g., C. elegans, plants, D. melanogaster, mice, and humans) (Ambros et al., 2003; Hamilton et al., 2002; Czech et al., 2008; Watanabe et al., 2008; Chen et al., 2012). Endo-siRNAs are ~21 nucleotides (nt) in length and are derived from much longer, naturally occurring double-stranded RNAs (dsRNAs) (Song et al., 2011). Precursors for endo-siRNA production are derived either from double-stranded RNAs consisting two complementary single-stranded RNAs (ssRNAs) or from the stem regions of single hairpin transcripts. The ssRNAs can be transcripts from the same genomic loci (cis) or loci on different chromosomes (trans) (Song et al., 2011; Okamura and Lai, 2008; Nilsen, 2008). Single-stranded mature endo-siRNAs are generated from these dsRNA precursors through a DICER-dependent, DROSHA-independent mechanism (Song et al., 2011). Mature endo-siRNAs are then loaded onto the RNA-induced silencing complex (RISC) containing Argonaute proteins, whereby endo-siRNAs “guide” the RISC to a transcript containing a sequence completely complementary to the sequence of the endo-siRNA, leading to
target transcript degradation and thus posttranscriptional silencing (Okamura and Lai, 2008; Golden et al, 2008).

Although numerous endo-siRNAs were initially identified in both male and female germ cells (Song et al, 2011; Babiarz et al, 2008; Tam et al, 2008) and embryonic stem cells (Babiarz et al, 2008), they appear to be expressed in other cell types and organisms (Fagegaltier et al, 2009; Ghilidiyai et al, 2008). Next-generation sequencing (NGS) is becoming a routine in biomedical research. In particular, sncRNA deep sequencing (sncRNA-Seq) has been widely used to profile sncRNA levels in specific cell types or organs. sncRNA-Seq data usually contain millions of sequence reads, which cover literally all small RNA species, including endo-siRNAs. However, identification of endo-siRNAs using sncRNA NGS data remains challenging because it requires a bioinformatic pipeline using computer programs to accomplish sequence matching at both genomic and transcriptomic levels. We have developed such a pipeline through which endo-siRNAs can be distinguished from other known and similar sncRNA species based on their structural features (e.g., length) and sources of their precursors (Song et al, 2011). The length of endo-siRNAs (~21 nt) differentiates them from piwi-interacting RNA (piRNA) and small nucleolar RNA, which range from 28 to 32 and 60 to 220 nt in length, respectively (Grivna et al, 2006; Holley and Topkara, 2011). However, some sncRNAs, such as miRNAs, are roughly the same length (~22 nt) as endo-siRNAs (Bartel, 2004). Therefore, to distinguish between miRNAs and endo-siRNAs, other differences must be taken into consideration. Unlike endo-siRNAs, miRNAs do not typically possess complete sequence complementarity to their target transcripts (Brennecke et al, 2005). Also, miRNAs are derived from short hairpin precursors, instead
of from long dsRNA precursors like endo-siRNAs (Bartel, 2004; Denli et al, 2004). By exploiting the differences between endo-siRNAs and other sncRNAs at the genomic level, it is possible to identify novel endo-siRNAs in silico. In this chapter, we describe a computer-assisted method for the identification of potential novel endo-siRNAs using small RNA NGS data. A flowchart describing our method is provided in Figure S1-1. The key to this method is a computer software called Sequery (version 1.0), which was developed in our lab at the University of Nevada, Reno, and is publically available at no cost (Ortogero et al, 2013). Sequery (version 1.0) is a custom program written in OpenGL and C with a graphical user interface that recognizes input sequences as 1-byte character strings. Sequery can be used to compare NGS reads to sets of DNA (e.g., genomic sequences) and RNA sequences (e.g., transcriptomes) by either complete or partial matching (Ortogero et al, 2013).

**MATERIALS**

Sequery v1.0 is a free program and available online for download at [http://www.medicine.nevada.edu/yan/sequery.html](http://www.medicine.nevada.edu/yan/sequery.html) (left click “Download Sequery”). Sequery is currently only available for Mac OS X. Instructions on how to install Sequery are provided within the Sequery Manual, which is available for download on the same Web page as Sequery v1.0 (left click “Download Sequery Manual”). Files that are loaded onto Sequery must be in FASTA format (Sequery is also compatible with some tab-delimited files; however, FASTA should be used if possible). Details on accepted file types are available in the Sequery manual.
METHODS

Filter Out Known sncRNAs

Initially, all known sncRNAs and all novel miRNAs should be filtered out of your dataset. This helps to prevent misclassification and, subsequently, reduces the computational demands of the analyses by limiting the search to only unclassified sncRNAs. Instructions on how to remove (and annotate) known sncRNAs, and how to identify novel miRNAs, are provided in Chapter 22 of the same volume by Ortogero et al. Only sncRNAs that are between 19 and 23 nt in length should be considered for further endo-siRNA identification analysis.

Identify Potential Endo-siRNAs

1) After a set of unclassified sncRNAs is obtained, the next step is to match each sncRNA to the transcriptome of its organism of origin. First, the reverse complementary sequences of each sncRNA must be generated. Load the file containing the sncRNA sequences being analyzed into Sequery by navigating to the correct directory in the FileMenu window and selecting the correct file. Detailed instructions on how to open files are available under “Opening a file” in the Sequery manual.

2) Once your file has been loaded into the Sequery v1.0 window, right click within the window to display the functions menu.
3) Right click on “Search&Gen” and left click “Generate Rev.Complementary.”

After you click “Generate Rev.Complementary,” Sequery should display the sense and antisense sequences of each sncRNA in the associated Terminal window.

4) Go back to the Sequery v1.0 window and again bring up the “Search&Gen” menu option, and left click “Swap Sense ↔ RC or C.” Then, right click to bring up the functions menu, and right click on “Load&Save.” Left click on “Save as FASTA.” This should bring up the FileMenu, where you can name the new file in the upper left corner (directly above the Folder column). After naming the file, press Enter or Return to save. Close Sequery.

5) The next step in identifying endo-siRNAs is to match the RC sncRNA sequences to the appropriate transcriptome (i.e., the complete transcriptome for the organism from which you obtained the sncRNAs). Start a new session in Sequery. Load the RC sncRNA sequence file into the Sequery v1.0 window, then, right click on “Load&Save” and right click on “Load File.” This will bring up the “FileMenu” again, where you can select the transcriptome file. Both files should be in the same format (i.e., FASTA). Within the Sequery v1.0 window, on the right of both file names are a set of boxes with the letters T, X, and C in each box (left to right). Left click on the “T” box next to the transcriptome file, changing it to an “R”—this sets the transcriptome file as the reference file and the sncRNA file as the test file.
6) An example of what the Sequery v1.0 window should look like is provided in Figure S1-2. Within the Sequery v1.0 window, right click to open the functions menu, then right click “Compare.” Left click on “3M T-R All Exact=” to match the two files (this may take some time, depending on your computer and the size of the two files). After matching has completed (the terminal window should display “Comp ok”), right click to open the functions menu, then right click on “Output” and left click on “3M Tests in Ref + U.” Enter the name of the output file in the FileMenu and hit “Enter” or “Return” to save the file. The RC sncRNAs that matched to at least one transcript should be considered as putative endo-siRNA; the RC sncRNA that did not match to the transcriptome should be removed from any further endo-siRNA analysis. The exact RC sncRNA—transcript matches are also the putative targets for each endo-siRNA.

Predict the dsRNA Precursors of Putative Endo-siRNAs

This method outlines an in silico method for predicting the class (i.e., hairpin, trans, cis) and identity of the dsRNA precursors for each putative endo-siRNA (see Notes 1 and 2).

1) First, the original sncRNA sequences (forward sequences) must be matched to the same transcriptome from the previous section. The matching procedure for the forward sncRNA and transcriptome sequences is identical to the one described in the previous section (for matching the RC sncRNA to the transcriptome). Next,
the RC sncRNA sequences must be compared to the transcriptome again, but with mismatches allowed. The procedure is identical to the protocol described in the previous section, but instead of selecting “3M T-R All Exact=,” left click “3M T-R All Sub1+=,” which allows for a certain number of mismatches between the sncRNA and transcript, depending on the number after the “=” (we have had success allowing up to two mismatches). The ssRNAs that comprise the dsRNA precursor do not have to be perfectly complementary; therefore, the search for matches between the RC sncRNA sequence and transcript needs to allow for mismatches (Song et al., 2011).

2) After the lists of transcript matches for both the RC (with mismatches allowed) and forward putative endo-siRNA sequences have been generated via Sequery, the genomic origins of the transcripts must be compared.

3) When the RC and forward endo-siRNA sequences match to transcripts originating from the same locus, but on opposing sides of the genome, that pair of ssRNA is considered as nat-cis-dsRNA (“nat” refers to the fact that the ssRNA are naturally occurring) precursors.

4) When the RC and forward endo-siRNA sequences match to transcripts from different loci, that pair of ssRNA is considered a nat-trans-dsRNA precursor.
5) Hairpin precursors occur when the RC and forward endo-siRNA match to the same transcript (assuming that transcript possesses a hairpin secondary structure).

6) A graphical example of all three precursor types is provided in Figure S1-3. Unfortunately, categorization of each endo-siRNA cannot currently be performed in Sequery. One straightforward, albeit labor-intensive approach is to manually compare the two lists for each endo-siRNA. Another option is to write a custom script that compares the two lists of transcript matches—we have prepared and uploaded such a script (only for use in mice and with ENSEMBL databases) onto the same page as the Sequery materials (http://www.medicine.nevada.edu/yan/sequery.html).

To resolve precursor structure manually, compare the sequences of the transcript matches for the RC (mismatches allowed) and forward endo-siRNA sequences one by one in BLAST (http://blast.ncbi.nlm.nih.gov/) using the “Align two or more sequences” option (Altschul et al, 1990). The dsRNA from the proper transcript matches typically range from 100 to 1,400 bp in length, with 80–100% complementarity (Song et al, 2011). The structures of suspected hairpin precursors can be confirmed using mFold (http://mfold.rna.albany.edu/?q=mfold/) or another folding program (Zuker, 2003). Examples of this manual approach for identifying nat-cis-dsRNA, nat-trans-dsRNA, and hairpin precursors are provided in Figures S1-4, -5, and -6 respectively.
NOTES

1) It is important to treat the endo-siRNAs identified in your analysis as putative before their expression is confirmed. This dramatically reduces the risk of misclassifying unknown sncRNAs as endo-siRNAs, a common issue with in silico identification analyses.

2) Confirmatory assays are beyond the scope of this chapter, however, Ro and Yan and several others have published detailed methods on the subject (Ro and Yan, 2010).
Figure S1-1 – Flowchart outlining our novel endo-siRNA identification method.
Figure S1-2 – Sequery v1.0 window example.
Figure S1-3 – Graphical examples of dsRNA precursors for endo-siRNA. (a) Putative nat-cis-dsRNA precursor for endo-siRNA-T32. (b) Putative nat-trans-dsRNA precursor for endo-siRNA-T19. (c) Putative hairpin precursor for endo-siRNA-T18, and corresponding mFold predicted hairpin structure.
Figure S1-4 – Example of putative nat-cis-dsRNA precursor for endo-siRNA-T32 (described in Figure S1-3a). The transcripts composing the dsRNA are highlighted in both the forward endo-siRNA—transcript match output file (top) and (mismatch allowed) RC endo-siRNA—transcript match file (middle). The endo-siRNA sequence is highlighted in the BLAST results (bottom); complete BLAST results are not shown due to space constraints.
Figure S1-5 – Example of putative nat-trans-dsRNA precursor for endo-siRNA-T19 (described in Figure S1-3b). The transcripts composing the dsRNA are highlighted in both the forward endo-siRNA—transcript match output file (top) and (mismatch allowed) RC endo-siRNA—transcript match file (middle). The endo-siRNA sequence is highlighted in the BLAST results (bottom).
Figure S1-6 – Example of putative hairpin precursor for endo-siRNA-T18 (described in Figure S1-3c). The transcript forming the hairpin precursor is highlighted in the both the forward endo-siRNA—transcript match output file (top) and (mismatch allowed) RC endo-siRNA—transcript match file (bottom). The mFold secondary structure prediction is shown in Figure S1-3c.
REFERENCES


