EPIGENOMICS OF POST-TESTICULAR SPERM MATURATION

A Dissertation Presented

By

CAROLINA GALAN

Submitted to the Faculty of the
University of Massachusetts Graduate School of Biomedical Sciences, Worcester

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

August 26, 2021

Molecular Biology and Epigenomics
EPIGENOMICS OF POST-TESTICULAR SPERM MATURATION

A Dissertation Presented

By

CAROLINA GALAN

This work was undertaken in the Graduate School of Biomedical Sciences

Interdisciplinary Graduate Program

Under the mentorship of

__________________________________________________________
Oliver J. Rando, M.D., Ph.D., Thesis Advisor

__________________________________________________________
Manuel Garber, Ph.D., Member of Committee

__________________________________________________________
Merav Socolovsky, M.B.B.S., Ph.D., Member of Committee

__________________________________________________________
Paul Thompson, Ph.D., Member of Committee

__________________________________________________________
Brett Nixon, Ph.D., External Member of Committee

__________________________________________________________
Thomas Fazzio, Ph.D., Chair of Committee

__________________________________________________________
Mary Ellen Lane, Ph.D.,
Dean of the Graduate School of Biomedical Sciences

August 26, 2021
ACKNOWLEDGMENTS

I would like to express my sincerest gratitude to my mentor for giving me this opportunity. Ollie, none of this would have been possible without your support, encouragement, and sense of humor. And never mind all the emotional support that came with writing the F31 fellowship – I have never worked so hard for a high-five in my life! Thank you for giving me the freedom to grow into an independent scientist. Although it will never be the same as being in your lab, at least I know you are looking forward to my lengthy, weekly emails. All this to say, if I had to do this all over again, there is no one I would rather have as a mentor for the better half of a decade.

To my committee – Tom, you have been incredibly helpful throughout this whole process – I wish you could be my Life Chair not just my Committee Chair - thank you! Manny, I have learned so much from you and I am grateful for all your computational and statistical advice throughout. Merav, thank you for your support along the way as well as experimental suggestions. To my new defense committee members, Paul – how fortuitous to have the PAD expert one floor down when things took a strange turn – Ronak, Ari, and Santanu were most helpful in providing me reagents and advice. And Brett, we have long admired your work and I am so grateful that you have agreed to participate - thank you!

To the current members of the Rando Lab: Fengyun, Tobias, Shweta, Vera, Ebru, Marina, and Nils – I miss you all more than you know. And to all the former Rando Lab members who shaped my early time lab – Welles, David, Stanley, Tiffany, Lina, Upasna,
Ana (I miss you everyday Ana - thank you for being such a great mentor to me), Colin (for teaching me to stop overthinking everything and just sequence it and for being an awesome mentor overall too), Morten (I miss you Morten!), Caitlin (and Peter!), and finally Ryan for introducing me to life inside the Rando Lab and all of its magic.

And to those in my former science life, thank you to Dan, Jeff, Gretchen, Jun, Mo, Jason, Elke, Andrew, Maria, Hannah, Jon, Danny, Allie, and Sam (and countless others) - I would not be the scientist - or person - I am today without your influence, support, and friendship. And to my those in my former, former science life, Lynn and Mike - thank you for opening my eyes to the vast world of science - all through the tiny world of termites.

Lastly, I have a deepest appreciation for my friends and family who have tirelessly supported me through this process and, my husband, for his unwavering love and patience through all of it!
ABSTRACT

Beyond the haploid genome, mammalian sperm carry a payload of epigenetic information with the potential to modulate offspring phenotype. Morphologically mature sperm exit the testes, but cannot swim or interact with the oocyte without extensive remodeling during epididymal transit; this includes modifications to the lipid composition of the sperm membrane, gain of necessary proteins, and a dramatic shift in sperm RNA content. Epididymal maturation has also been linked to changes in the sperm methylome suggesting that the epididymis might play a broader role in shaping the sperm epigenome. First, we characterized the genome-wide methylation landscape in seven germ cell populations from throughout the male reproductive tract. Our data emphasize the stability of cytosine methylation in mammalian sperm, and identify a surprising, albeit transient, period during which sperm are associated with extracellular DNA. Second, given our interest in the small RNA repertoire of sperm we set out to address known bias in sequencing protocols by comparing several small RNA cloning protocols. We found a protocol recently developed by Kathleen Collins’ lab (OTTR) to be superior to commercially available kits in providing an accurate representation of tRNA fragment levels as compared to Northern blotting. These results not only provide a more accurate representation of tRNA fragments, but also more complexity than previously seen allowing us to reassess the true sperm small RNA content. Taken together, these results provide significant insight into the mechanisms and factors modulating sperm epigenomics during post-testicular sperm maturation.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>TITLE</td>
</tr>
<tr>
<td>REVIEWER PAGE</td>
</tr>
<tr>
<td>ACKNOWLEDGMENTS</td>
</tr>
<tr>
<td>ABSTRACT</td>
</tr>
<tr>
<td>TABLE OF CONTENTS</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
</tr>
<tr>
<td>PREFACE</td>
</tr>
</tbody>
</table>

## CHAPTER I. INTRODUCTION

1.1 Introduction ................................................................. 2
1.2 Overview of paternal environmental conditions .......................... 5
  1.2.1 Paternal dietary exposures ........................................... 5
  1.2.2 Ancestral exposure to traumatic experiences ...................... 6
  1.2.3 Ancestral toxin exposure .............................................. 7
1.3 Epigenetic information carriers in the germline .......................... 8
  1.3.1 Chromatin .............................................................. 8
  1.3.2 DNA Modifications .................................................. 10
  1.3.3 RNAs ..................................................................... 14
1.4 The epididymis is critical for male fertility ............................... 19
1.5 Sperm are extensively remodeled during epididymal transit
  providing a mechanistic “window of opportunity” ........................ 22
1.6 Overview of this dissertation ............................................... 23

## CHAPTER II. STABILITY OF THE CYTOSINE METHYLOME
DURING POST-TESTICULAR SPERM MATURATION IN MOUSE ............ 25

2.1 Abstract ........................................................................... 26
2.2 Introduction ....................................................................... 27
2.3 Results ............................................................................. 30
2.4 Discussion ........................................................................ 59
2.5 Methods ........................................................................... 68
## Chapter 3. Sperm Small RNAs Three Ways

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1 Abstract</td>
<td>77</td>
</tr>
<tr>
<td>3.2 Introduction</td>
<td>78</td>
</tr>
<tr>
<td>3.3 Results</td>
<td>79</td>
</tr>
<tr>
<td>3.4 Discussion</td>
<td>90</td>
</tr>
<tr>
<td>3.5 Methods</td>
<td>119</td>
</tr>
</tbody>
</table>

## Chapter 4. Discussion

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A.1 Abstract</td>
<td>124</td>
</tr>
<tr>
<td>A.2 Introduction</td>
<td>133</td>
</tr>
<tr>
<td>A.3 Results</td>
<td>134</td>
</tr>
<tr>
<td>A.4 Discussion</td>
<td>135</td>
</tr>
<tr>
<td>A.5 Methods</td>
<td>140</td>
</tr>
</tbody>
</table>

**Appendix I. Impact of Aging on the Sperm Epigenome**

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A.1 Abstract</td>
<td>134</td>
</tr>
<tr>
<td>A.2 Introduction</td>
<td>135</td>
</tr>
<tr>
<td>A.3 Results</td>
<td>140</td>
</tr>
<tr>
<td>A.4 Discussion</td>
<td>163</td>
</tr>
<tr>
<td>A.5 Methods</td>
<td>165</td>
</tr>
</tbody>
</table>

**References**

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A.1 Abstract</td>
<td>167</td>
</tr>
</tbody>
</table>
LIST OF TABLES

PAGE

Table 2.1 Sequences of pyrosequencing primers ......................................................... 73

Table 3.1 Sequential mapping to determine percent of reads mapping ......................... 116

Table 3.2 Sequential mapping to determine percent of reads mapping after rRNA removal .......................................................... 118

Table A.1 Enriched pathways in DEGs for Caput Epididymis .................................... 148

Table A.2 Enriched pathways in DEGs for Cauda Epididymis ..................................... 156
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Epigenetics in action</td>
<td>3</td>
</tr>
<tr>
<td>1.2</td>
<td>The epididymis</td>
<td>21</td>
</tr>
<tr>
<td>2.1</td>
<td>Whole genome cytosine methylation in seven germ cell populations</td>
<td>31</td>
</tr>
<tr>
<td>2.2</td>
<td>Correlations between all seven germ cell methylation datasets</td>
<td>34</td>
</tr>
<tr>
<td>2.3</td>
<td>CpG island hypermethylation in caput and vas deferens sperm</td>
<td>37</td>
</tr>
<tr>
<td>2.4</td>
<td>GO enrichments for genes associated with aberrant CpG island methylation in caput sperm</td>
<td>39</td>
</tr>
<tr>
<td>2.5</td>
<td>Genomic context for targeted pyrosequencing followup</td>
<td>41</td>
</tr>
<tr>
<td>2.6</td>
<td>Pyrosequencing validation of WGBS data</td>
<td>42</td>
</tr>
<tr>
<td>2.7</td>
<td>The caput methylation profile is stable for at least ten months in unmated animals</td>
<td>42</td>
</tr>
<tr>
<td>2.8</td>
<td>Caput sperm methylation status is stable under many different buffer conditions</td>
<td>45</td>
</tr>
<tr>
<td>2.9</td>
<td>The unusual methylation program in caput sperm is lost after mating</td>
<td>47</td>
</tr>
<tr>
<td>2.10</td>
<td>Enriching for the first wave of testicular sperm</td>
<td>48</td>
</tr>
<tr>
<td>2.11</td>
<td>Cell-free DNA in the caput epididymis of virgin males is responsible for the caput methylome</td>
<td>50</td>
</tr>
<tr>
<td>2.12</td>
<td>DAPI-staining rim associated with the apical region of the epithelium of caput and vas deferens in virgin males</td>
<td>52</td>
</tr>
<tr>
<td>2.13</td>
<td>Cell-free DNA is responsible for the abnormal caput methylome</td>
<td>54</td>
</tr>
<tr>
<td>2.14</td>
<td>DNase-sensitive aberrant methylation is consistently observed in multiple strain backgrounds</td>
<td>55</td>
</tr>
<tr>
<td>2.15</td>
<td>Resolution of the caput methylome by DNase I treatment</td>
<td>56</td>
</tr>
</tbody>
</table>
Figure 2.16 Cell-free DNA is produced via a NETosis-like process

Figure 2.17 WGBS data for selected imprinted genes

Figure 2.18 Loci studied in Ariel et al, 1994.

Figure 2.19 Mating does not impact the small RNA payload of caput sperm

Figure 3.1 Sperm RNAs have a distinct size distribution as compared to the tissue from which it is generated

Figure 3.2 Examples of RNA modifications present of tRNAs

Figure 3.3 T4 Polynucleotide Kinase (T4 PNK)

Figure 3.4 Reads mapping to tRF-GlyGCC

Figure 3.5 Reads mapping to tRF-ValCAC

Figure 3.6 Reads mapping to tRF-GlyGCC after T4 PNK treatment

Figure 3.7 Reads mapping to tRF-ValCAC after T4 PNK treatment

Figure 3.8 Barplots illustrating diversity of tRNA-mapped reads by species

Figure 3.9 tRNA Coverage plots illustrating alignment position of tRNA by species

Figure 3.10 Barplot illustrating diversity of tRNA-mapped reads by species for OTTR

Figure 3.11 Barplot illustrating diversity of tRNA-mapped reads by species for Illumina Truseq

Figure 3.12 Barplot illustrating diversity of tRNA-mapped reads by species for NEBNext

Figure 3.13 Coverage plot for OTTR protocol by tRNA

Figure 3.14 Coverage plot for Illumina Truseq by tRNA

Figure 3.15 Coverage plot for NEBNext by tRNA

Figure 3.16 Reads mapping to LeuCAG
Figure 3.17 Reads mapping to AspGTC ..........................................................109
Figure 3.18 Read length of insert mapped to tRNAs .........................................111
Figure 3.19 Read length of insert after adaptor removal ...............................113
Figure 3.20 Relative rRNA ratios mapping in each condition and library build ....115
Figure A.1 Schematic of RNA-seq and small RNA experimental samples ..........141
Figure A.2 Confirmation of Expression of Prm1, Prm2, Tnp1, and Tnp2 in 3 week, 6 week, and 40-week-old testes samples ............................................142
Figure A.3 Scatterplot analysis of normalized gene expression of testes ..........144
Figure A.4 Heat map of top ten most variably expressed genes in testes ..........145
Figure A.5 Scatterplot and heatmap analysis of normalized gene expression of caput ..........................................................147
Figure A.6 Heat map of top ten most variably expressed genes in caput .............149
Figure A.7 Decreasing expression of Clu, Gpx5, Adam7, Slc7a5, Slc38a5 and Tmem258 in 40-week-old caput epididymis as compared to 6-week-old ..........150
Figure A.8 Increasing expression of mt-Co1, Ccl21a, Apod, Pecam, Aqp1 and Ly6a in 40-week-old caput epididymis as compared to 6-week-old ...................152
Figure A.9 Scatterplot and heatmap analysis of normalized gene expression of cauda ..........................................................155
Figure A.10 Heat map of top ten most variably expressed genes in cauda ..........157
Figure A.11 Decreasing expression of Crisp1, Cd52, RNase9, RNase12, Adam7 and Itm2b in 40-week-old cauda epididymis as compared to 6-week-old ..........158
Figure A.12 Increasing expression of mt-Cytb, Cxcl12, Apod, Pecam, Ccl21a and Ccl8 in 40-week-old cauda epididymis as compared to 6-week-old ..........160
Figure A.13 Small RNAseq from 6 week and 40-week-old sperm .................162
PREFACE

The chapters included in this dissertation have been modified from manuscripts published in separate publications.

The work presented in Chapter I of this thesis has been adapted from:


* Authors contributed equally to this work

The work presented in Chapter II of this thesis has been adapted from:

CHAPTER I.

INTRODUCTION
1.1 Introduction

Broadly, epigenetics refers to the study of heritable changes in phenotype without changes to the underlying DNA sequence. Forms of epigenetic regulation including modifications to DNA and RNA (i.e., methylation), posttranslational modification of histones, all the instructive chromatin modifiers (readers, writers, erasers), and of course, ncRNAs. This epigenetic regulation is critical for proper development as well as distinct cell lineages. For example, the many differentiated cell types you have in multicellular organisms – and their heritability. This inheritance of cell state means that a liver cell will divide into two liver cells - not a spleen and a heart cell – despite sharing identical genomes (Figure 1.1). (Holoch and Moazed, 2015; Margueron and Reinberg, 2010; Jaenisch and Bird, 2003). Transmission of epigenetic information is stable through cell divisions with the ability of the same genome to produce multiple distinct, yet stable, phenotypes - but what about from one generation to the next? What about through the germline?
Figure 1.1 Illustrating Epigenetics in Action
A) Epigenetic Regulation of cell fate refers to the potential of differentiated cell types emerging from the same genome B) Inheritance of cell state refers to epigenetic regulation dictating the outcome of a specific, heritable cell type.

*Illustrations in this figure created with BioRender.com*
Although biological inheritance is primarily driven by transmission of the genome from one generation to the next, additional “epigenetic” information can be passed on to future generations. Classic examples of epigenetic inheritance include relatively stable inheritance of gene silencing, as in the case of paramutation in maize (Arteaga-Vazquez and Chandler, 2010), as well as short-term cases of programmed epigenetic inheritance which are erased each generation, as in the case of imprinted gene regulation in mammals and plants (Hackett and Surani, 2013; Feng et al., 2013). Beyond such largely environmentally-insensitive cases of epigenetic inheritance, it is increasingly clear that various perturbations can influence epigenetic modifications in germ cells, which can thereby transmit information about prevailing environmental conditions to future generations. This process is essentially a modern reappraisal of the once-discredited idea of the “inheritance of acquired characters”, with subtle quantitative traits responding, over limited numbers of generations, to ancestral environments.

In mammals, it is now well-established that parental dietary challenges and other stressors can induce germline epigenetic alterations and thereby affect metabolic phenotypes in the next generation. However, the mechanistic basis by which these epigenetic modifications ultimately program altered metabolism in offspring remains obscure. Here, we will discuss parental dietary paradigms and their effects on the germline epigenome. We will then focus on potential mechanisms linking these germline epigenetic changes to the physiological changes observed in offspring. And finally, we turn to investigate the role the epididymis plays as a “window of opportunity” for epigenomic reprogramming through the germline.
1.2 Overview of paternal environmental conditions effects on offspring phenotypes

Here we will briefly survey paternal exposure paradigms that have been linked to changes in offspring phenotype, focusing on studies in inbred rodent model systems. In general, ancestral exposure studies typically focus on one of three broad environmental paradigms: altered diet/nutrition, toxin exposure, and stress.

1.2.1 Paternal dietary exposures

A large number of studies have investigated the effect of paternal diets on F1 and F2 offspring in mice and rats. Perhaps the best studied dietary perturbation is consumption of a high-fat diet, which programs a coherent pattern of phenotypes in the next generation, including abnormalities in glucose tolerance, body weight, fat distribution, and reproductive health (Su and Patti, 2019; Ng et al., 2010; Stanford et al., 2018; Fullston et al., 2012; McPherson et al., 2014; Mitchell et al., 2011; Wei et al., 2014; Huypens et al., 2016; de Castro Barbosa et al., 2016). Interestingly, many of these phenotypes, including effects on glucose tolerance, glucose uptake, and weight gain, can be ameliorated when fathers on high fat diets are also forced to exercise (Stanford et al., 2018; McPherson et al., 2017). The next-most common dietary paradigms used in paternal effect studies are related to undernutrition: paternal consumption of a low protein diet has been shown to program changes in cholesterol and lipid metabolism, glucose control, and other cardiovascular parameters in F1 offspring (Ly et al., 2019; Watkins et al., 2018; Carone et al., 2010; Watkins and Sinclair, 2014). Other undernutrition
paradigms with documented effects on progeny phenotypes include caloric restriction (Govic et al., 2016) and intermittent fasting (Anderson et al., 2006).

1.2.2 Ancestral exposure to traumatic experiences

Exposure of fathers to traumatic situations, generally experienced early in life, has been repeatedly documented to drive various behavioral and even metabolic alterations in unexposed offspring. Typical paternal stress paradigms include social defeat stress (Dietz et al., 2011), unpredictable maternal separation and maternal stress (MSUS – (Gapp et al., 2018; Gapp et al., 2014), social instability (Dickson et al., 2018; Saavedra-Rodriguez and Feig, 2013), and chronic variable stress in which males are subjected to unpredictable daily stressors ranging from moist bedding material, to social defeat, to placement of novel items in the cage (Rompala et al., 2018; Rodgers et al., 2015; Rodgers et al., 2013; Morgan and Bale, 2011). Males subjected to MSUS in childhood sired offspring that produced lower levels of the stress-responsive hormone corticosterone and engaged in fewer anxiety-like behaviors as measured by elevated plus maze, light dark box, and forced swim test (Rodgers et al., 2013). Beyond these anxiety-related behaviors, MSUS offspring may also inherit deficits in memory formation processes, as Bohacek et al. observed that MSUS offspring could not differentiate between a familiar and novel object 24 h after initial exposure, and spent less time frozen in response to a conditioned fear stimulus (Bohacek et al., 2015). Furthermore, long-term potentiation was impaired in MSUS offspring, whereas long-term depression was improved over controls, although this phenotype lasted for only one generation. Of note, not only does paternal stress
influence stress-related behaviors in offspring, but there is some evidence that F1 offspring also exhibit metabolic phenotypes, such as increased insulin sensitivity (Gapp et al., 2018) or altered dietary preferences (Dietz et al., 2011).

1.2.3 Ancestral toxin exposure

Finally, we will briefly note the wide range of studies reporting intergenerational effects of a variety of bioactive small molecules, ranging from endocrine-disrupting fungicides to drugs of abuse such as alcohol and nicotine. Most notably, toxin exposure as an epigenetic stressor has been extensively studied in the context of endocrine-disrupting chemicals used in pesticides, fungicides, and herbicides. In one of the earliest documented paternal effect studies, exposure of pregnant rats to high levels of vinclozolin or methoxychlor during fetal gonadal development was reported to cause decreased sperm number and motility in F1, F2, F3, and even F4 generations, with eight percent of males, especially older males, developing infertility (Anway et al., 2005). Other toxins, such as dioxins, bisphenol A, carbon tetrachloride, and pollutants contained in jet fuel have also been linked to offspring phenotypes ranging from reproductive health to metabolism (Nilsson and Skinner, 2015; Susiarjo et al., 2015; Zeybel et al., 2012). Beyond the various environmental contaminants studied, a number of efforts have focused on the effects of drugs of abuse on future generations. For example, paternal ethanol consumption affects a range of phenotypes, including HPA axis responsiveness and alcohol preferences, in F1 offspring (Rompala et al., 2017). Paternal exposure to nicotine, cocaine, and THC have all also been linked to offspring phenotypes including
alter behaviors (Hawkey et al., 2019; Levin et al., 2019; Vassoler et al., 2013), resistance to xenobiotics (Vallaster et al., 2017), and altered glucose control (Vallaster et al., 2017).

1.3 Epigenetic information carriers in the germline

Not only is epigenetic information distributed to both daughter cells during mitotic division (as seen in cell state inheritance in multicellular organisms), but some epigenetic marks are inherited through the germline (reviewed in Bošković, and Rando, 2018); below, we will briefly cover the unusual aspects of the germ cell epigenome.

1.3.1 Chromatin

The nucleoprotein packaging of the sperm genome is remodeled extensively to ensure proper compaction. In many species, this compaction process involves replacing canonical histones with histone variants, transition proteins, and finally the small basic proteins known as protamines (Hoghoughi et al., 2018; Bao and Bedford, 2016). However, this global replacement process is not entirely exhaustive, with ~1–15% of histones being retained in the sperm of various mammals. The locations of retained histones in mammalian sperm remain somewhat unresolved, as many reports have mapped sperm histones to GC-rich promoters of early developmental genes (Yoshida et al., 2018; Erkek et al., 2013; Hammoud et al., 2009), but with at least two studies instead finding histones retained primarily in large gene-poor regions (Yamaguchi et al., 2018; Carone et al., 2014). This discrepancy appears to be at least partly explained by
differences in stability between the bulk of sperm histones located in intergenic regions, and a subset of unusually-stable nucleosomes present at CpG islands. Importantly, evidence from both camps suggests that it is unlikely that a given nucleosome is ever retained uniformly at a particular locus. In other words, even in the case of histone retention at developmental promoters, ChIP-Seq detects signal for protamines as well. As it is highly unlikely that histones and protamines co-occupy the same locus simultaneously (and there is no reason to think that histone and protamine antibodies cross-react), this finding is best explained by a mix of sperm bearing histones at the promoter of interest, and sperm which have successfully replaced the histones at that location. As will also be discussed in the case of cytosine methylation changes, it is difficult to reconcile how epigenetic marks present in only a fraction of sperm could explain paternal effects which typically appear to be highly penetrant in the next generation. As a concrete example, if a paternal perturbation affects histone modifications at a locus where, say, 20% of sperm retain histones rather than protamines, the modifications in question can only be meaningful for 20% of offspring. Although counterarguments to this concern can be envisioned – perhaps the 80% of sperm lacking histones at that promoter are incapable of fertilization and thus irrelevant, or perhaps offspring phenotype integrates information from multiple independent loci – the issue of penetrance must be dealt with when considering the role of the sperm epigenome in paternal effects.

Nonetheless, there is some evidence supporting the idea that manipulating histone retention or covalent modification status can affect genomic functions in the zygote and
beyond. In one prominent example, overexpressing the H3K4 demethylase KDM1A (LSD1) during spermatogenesis was shown to reduce global H3K4me2 levels in sperm, and led to impaired offspring development in three subsequent generations (Siklenka et al., 2015). More recently, Lesch et al. showed that deleting the H3K27 demethylase KDM6A in the paternal germline resulted in increased tumor incidence in offspring, persisting through two generations (Lesch et al., 2019). While these and other related efforts (Ihara et al., 2014) link chromatin-related mutations to offspring phenotypes, the relevant molecular carrier in sperm remains unclear – while cis-acting changes in histone levels or modifications at specific genomic loci in sperm could potentially influence local gene expression in the early embryo, it is also clear that deletion or overexpression of histone-modifying enzymes can affect other epigenetic modifications including sperm RNA levels (Siklenka et al., 2015) and cytosine methylation patterns (Lesch et al., 2019). It is therefore important to be aware that these or other molecular changes in mutant sperm could certainly drive offspring phenotypes even if sperm chromatin changes do not exert any direct effect on zygotic gene regulation.

1.3.2 DNA modification

Methylation of the C5 position of cytosine is the most common DNA modification in mammals, where it plays important roles in processes ranging from transposon silencing to imprinting, X-chromosome inactivation, and genomic stability (reviewed in Greenberg and Bourc'his, 2019; Deniz et al., 2019). Although cytosine methylation patterns can in principle be copied in perpetuity through the action of the maintenance
methyltransferase Dnmt1, this is prevented by two waves of nearly-complete DNA methylation reprogramming that occur during primordial germ cell development, and again in the early embryo (Hackett and Surani, 2013). In the embryo, active demethylation of the paternal genome occurs soon after fertilization, contrasting with the apparently passive demethylation events on the maternal genome that occur over several cell divisions (SanMiguel J.M., Bartolomei, 2018; Okada and Yamaguchi, 2017).

Nonetheless, a subset of genomic loci escape the reprogramming process, including imprinting control regions and certain evolutionarily young transposons (such as intracisternal A-particles, or IAP elements, in the mouse) (Smith and Meissner, 2013). The mechanistic basis by which these loci are protected from demethylation is an active area of investigation: in the case of imprinted genes, a pair of zinc finger DNA-binding proteins – ZFP57 and ZFP445 – appear to play essential roles in preventing demethylation of many imprinting control regions (Takahashi et al., 2019; Li et al., 2008).

As sperm were long believed to carry no functional RNAs, and the very few histones escaping replacement were also generally viewed as unlikely to be functional in the early embryo. The vast majority of paternal effect studies have therefore focused on the sperm cytosine methylome, given the well-established role for cytosine methylation as epigenetic information carrier in mammals. We briefly highlight several examples to illustrate commonalities in the current literature on environmental regulation of the sperm methylome. First, in the prominent paternal effect paradigm based on in utero undernutrition, low-resolution genome-wide analyses identified ~100 loci, primarily in
intergenic regions, hypomethylated in sperm of undernourished males (Radford et al., 2014). As is typical for reported environmental effects on sperm methylation, quantitative followup revealed ~10–20% changes in methylation (e.g. a decrease from 40% to 20% methylation) at between one and nine neighboring CpGs. Importantly, only a small subset of these methylation changes persisted into adulthood, or were correlated with changes in expression of nearby genes, further underlining the question of how these modest methylation changes could influence offspring phenotype. Effects of similar magnitude are quite abundant in the literature, with 10–20% methylation changes identified in sperm obtained from males subject to dietary (Wei et al., 2014), toxin-induced (Skinner et al., 2018), or even odorant-paired stress (Dias and Ressler, 2014) paradigms.

The common theme of modest methylation changes in sperm means that environmental conditions are driving relatively subtle shifts at any one locus in each sperm. This presents a mechanistic challenge based on the “digital” nature of sperm DNA methylation. Haploid sperm carry a single copy of each cytosine in the genome, and only one sperm contributes its genome to any given offspring. As a single cytosine can only occur in one of only two states – methylated or unmethylated – a given CpG cannot be 65% methylated in a single sperm. This means that a 10% methylation change in a sperm sample indicates a change from, say, 2 of 10 sperm bearing a methyl group to 3 of 10 sperm bearing a methyl group at that cytosine. Thus, modest methylation changes should only affect the penetrance of a phenotype within a litter, rather than affecting the majority of offspring across litters as observed in many paternal effect paradigms.
Given this “digital sperm” problem, studies reporting cytosine methylation changes at various repeat elements seem more promising for future investigation. For instance, sperm from mice subjected to a low protein diet from gestation through pre-weaning age exhibited subtle changes in DNA methylation at ribosomal DNA (rDNA) (Holland et al., 2016), with rDNA copies bearing a particular polymorphism (CpG-133^A^) exhibiting more significant methylation changes than other rDNA copies. Related findings include reports of high fat diet driving increased DNA methylation at satellite repeats (centromeres/telomeres) in rats (Youngson et al., 2016), a global decrease in repeat element cytosine methylation in low protein sperm (Watkins et al., 2018), as well as human epidemiological studies linking phthalate exposure to DNA methylation at LINE-1 elements in sperm (Tian et al., 2019).

Countering these studies, we note that extensive whole genome bisulfite sequencing in our own laboratory revealed no significant impact of either low protein or high fat diet on repeat element methylation status (Shea et al., 2015). Rather, we identified a significant confounding influence of repeat copy number differences between even closely-related animals, which initially lead to the artifactual identification of dietary effects on rDNA and other repeat methylation.

Nonetheless, environmental effects on repeat element methylation status bear further investigation both because repeat elements represent some of the rare genomic loci that escape methylation erasure upon fertilization, and because repeat element methylation provides a plausible mechanism for small methylation changes to affect offspring with high penetrance.
1.3.3 RNAs

Finally, germline RNAs are central to the best-understood transgenerational epigenetic inheritance paradigms in model organisms from plants to worms, and have emerged as primary candidates for mediators of paternal effect paradigms in mammals (Chen et al., 2016B). As with other epigenetic information carriers, small RNA populations are reprogrammed throughout development and maturation of germ cells. During testicular spermatogenesis, microRNAs, PIWI-interacting RNAs (piRNAs), and endogenous small interfering RNAs (endo-siRNAs) are dynamically expressed and play critical roles in normal development (Luo et al., 2016). Most notably, the largely germline-specific piRNAs play well-known roles in protecting the germline from transposable element mobilization (Ozata et al., 2019). The sperm RNA payload is then extensively remodeled following the completion of testicular spermatogenesis but prior to mating: piRNAs, which are the major small RNA species in testicular sperm, are almost completely absent in ejaculated sperm, which instead carry an RNA payload comprised primarily of tRNA fragments (tRFs) derived from mature tRNAs (Sharma et al., 2018; Sharma et al., 2016; Chen et al., 2016A; Peng et al., 2012). Sperm also carry a variety of microRNAs, and several recent studies have revealed functional roles for sperm-delivered microRNAs in the early embryo (Conine et al., 2019; Conine et al., 2018; Yuan et al., 2016).

In addition to these populations of small RNAs, which have functional roles in the zygote, sperm carry a variety of other RNA species of unknown significance. For
example, the majority of mRNAs and rRNAs present during the process of spermatogenesis exhibit varying degrees of degradation in mature sperm (Sendler et al., 2013); whether rRNA fragments and other partially-degraded RNAs have biological functions such as those attributed to tRNA fragments largely remains to be seen. In addition to the various digestion products of longer RNAs, circular RNAs (circRNAs) – circular species formed by back-splicing reactions (Patop et al., 2019) – are resistant to exonucleases and may therefore remain relatively intact throughout spermatogenesis and sperm maturation. Indeed, one of the first circRNAs discovered was a highly abundant testis-specific circRNA generated from the sex-determining region Y (Sry) gene (Capel et al., 1993). CircRNAs have been linked to multiple functions, most notably acting in several contexts as microRNA “sponges” which effectively knock down microRNA function. For instance, Sry circRNA has been shown to reduce miR-138 expression (Hansen et al., 2013), while circNAPEPLD, present in both mouse and human mature sperm, was shown to physically interact with multiple oocyte microRNAs (Ragusa et al., 2019).

The central role for small RNAs in well-established transgenerational epigenetic inheritance paradigms in other organisms, has motivated an increasing focus on small RNA levels in mammalian sperm. A key feature driving the recent surge of interest in sperm RNAs in paternal effects is the relative ease of functional testing of small RNAs as causal agents of paternal effect paradigms: microinjection of purified or synthetic small RNAs into control zygotes is far more tractable than modulating histone retention or
cytosine methylation in a locus-specific manner (although CRISPR-based targeting does make the latter approach more feasible than in the past).

Focusing first on dietary paradigms, numerous studies have documented dietary effects on the sperm RNA payload, although as in the case of methylation profiling, these studies often disagree in detail. For example, in the case of high fat paradigms, dietary intervention has been reported to result in: 1) increased expression of several microRNAs and 5’ tRFs (most notably let-7c, tRF-GluCTC and tRF-GluTTC) in rats (de Castro Barbosa et al., 2016); 2) increased levels in tRF-GlyGCC and miR-10a/b, accompanied by a decrease in tRF-GluCTC (Cropley et al., 2016); 3) increased levels of tRF-GlyGCC, tRF-GlyCCC, and tRF-HisGTG (Stanford et al., 2018); 4) altered levels of miR-503, miR-456b, miR-542, and miR-652 (McPherson et al., 2015); 5) increased levels of miR-19b and miR-29a (Grandjean et al., 2015); and 6) widespread changes in tRF and microRNA levels, including increased miR-10a/b, miR-122, and decreased let-7c/f/a/b, as well as changes in the covalent nucleotide modifications associated with tRFs (Chen et al., 2016A). These discrepancies presumably reflect some combination of known and currently-enigmatic technical and biological differences, ranging from differences in RNA extraction and cloning protocols (size range of gel-purified RNAs, for example) to differences in the endemic microbiota across animal facilities.

Early life stressors have also been linked to diverse changes in the sperm RNA payload in many studies: 1) MSUS was reported to drive changes in both microRNA levels (miR-375, miR-200b, miR-672, and miR-466) as well as levels of several mRNAs and lincRNAs (Gapp et al., 2018; Gapp et al., 2014); 2) paternal chronic stress was
shown to affect expression of nine microRNAs (miR-193, miR-204, miR-29c, miR-30a, miR-30c, miR-32, miR-375, miR-532, and miR-698) (Rodgers et al., 2013); 3) cortisol injections led to altered levels of nearly 200 small RNAs, including ~100 microRNAs and ~60 tRFs (Short et al., 2016); 4) social instability in mice led to decreased levels of miR-34c and miR-449a in sperm, and, intriguingly, the same microRNAs were also less abundant in human sperm samples from adult survivors of adversity early in life (Dickson et al., 2018). As with the dietary paradigms above, the small RNA changes reported in the sperm of stressed males differ dramatically from report to report, although here the paradigms differ much more substantially in detail than for the various high fat diet protocols. The diversity of small RNA changes may therefore simply reflect differences between conditions – e.g. social instability as opposed to cortisol injections – in terms of their effects on the recipient.

These and many other findings using various paternal effect models, from low protein diet to pesticide exposure to ethanol consumption, suggest that the sperm RNA payload is surprisingly plastic. Importantly, in the case of sperm RNAs, it has been possible to directly test whether the RNAs in question are sufficient, when microinjected into control zygotes, to program offspring phenotypes. A number of studies have reported successful induction of specific phenotypes – often a limited subset of the phenotypes induced by the relevant paternal environments – following injection of various RNA populations. For example, Gapp et al. reported that injection of total RNA purified from MSUS sperm into control zygotes resulted in glucose intolerance following restraint stress, and depression-like behaviors including increased floating in a forced swim test, in
adult offspring (Gapp et al., 2014). Some of these effects were later reported following injection of gel-purified long (>200 nt) RNAs from MSUS sperm, suggesting that mRNAs or lincRNAs could play some role in programming offspring phenotypes (Gapp et al., 2018). Other studies have used more defined populations of small RNAs – gel-purified tRFs isolated from high fat sperm were able to direct altered glucose metabolism in one study (Chen et al., 2016A). Similarly, tRFs gel-purified from the sperm of males born to high fat mothers were sufficient (when injected into control embryos) to induce increased locomotion after amphetamine exposure, as well as increased preference for sucrose and high fat diet (Sarker et al., 2019). More specifically still, small numbers of synthetic RNAs have been used in some microinjection studies, with nine microRNAs (miR-193, miR-204, miR-29c, miR-30a, miR-30c, miR-32, miR-375, miR-532, and miR-698) causing altered blood brain barrier permeability in injected animals (Rodgers et al., 2015, while microinjection of miR-19b resulted in decreased glucose tolerance in offspring (Grandjean et al., 2015). Taken together, these studies provide strong support for the potential of sperm RNAs to act as the key molecular mediators by which paternal environmental conditions influence metabolic and other phenotypes in F1 offspring.
Given these three classic epigenetic inheritance pathways and a wide range of other factors that could play potential roles in transmission of environmental information from father to child, we next need to address how this environment translates into epigenetic changes in sperm? How is environment (diet, toxin, stress) sensed and where do these changes in sperm occur? Below, we address what we view as the link connecting paternal experiences to epigenetic changes in sperm – the epididymis.

1.4 The epididymis is critical for male fertility

Armed with the knowledge that at least some epigenetic information in germ cells escapes erasure in the next generation, we next set out to explore where this “window of opportunity” lies. Although sperm development primarily occurs during spermatogenesis in the testes, we hypothesize that the translation of environment to germline is occurring during epididymal maturation - rather than interference with spermatogenesis.

It has been clear for over 50 years that sperm exiting the testis are incapable of fertilizing oocytes (Bedford, 2015; Orgebin-Crist, 1967; Bedford, 1967). Morphologically mature sperm leave the testes and travel to the epididymis – the site of sperm maturation. Sperm spend upwards of two weeks traveling through this highly coiled and lengthy organ positioned between the testes and vas deferens. To fully appreciate its intricacy in folding, it is important to note that - when extended - is over six meters long in humans and one meter in mice (Hinton et al., 2011). Post-testicular maturation is essential for fertility, as sperm acquire forward motility and the ability to
interact productively with the oocyte only during epididymal transit, with both these functions are required for fertility (James et al., 2020; Turner, 2008).

The epididymis can be divided into distinct regions (initial segment in rodents, caput, corpus, and cauda) (Figure 1.2) with each region providing a unique secretory microenvironment contributing to serial modifications of sperm. The cell composition throughout the organ consists of principal cells, narrow cells, clear cells, basal cells, and halo cells – each with their own described specialized function and localization. For example, principal cells – the primary cell type of the epididymis – are responsible for secretion of microvesicles (Sullivan and Saez, 2013) and both narrow and clear cells are understood to regulate luminal pH (Cornwall, 2009).
Figure 1.2 The epididymis
A) Illustration of the epididymis as the link between the efferent ducts of the testes on the proximal side and vas deferens on its distal end. The epididymis is anatomically divided in distinct regions - the caput (head), corpus (body), and cauda (tail). B) FVB mouse epididymis.

Additionally, the epididymis serves to concentrate and store sperm, maintain viability for long periods of time (Hinton et al., 1996) as well as providing an antioxidant system to protect sperm against stress, oxidative damage, and infections (Flaherty, 2019; El-Taieb et al., 2009). Given its collective functions, the epididymis plays a critical role in sperm development. Indeed, many gene knockout mice have been generated with the finding that changing the epididymal environment leads to subfertile or infertile mice (Zhou et al., 2018). For example, Gpx5/- male mice have a higher incidence of
miscarriage and developmental defects (Chabory et al., 2009) – we know GPX5 is a specialized antioxidant enzyme functioning to protect sperm from oxidative stress in the epididymis (Aitken, 2009).

1.5 Sperm are extensively remodeled during epididymal transit providing a mechanistic “window of opportunity”

It has long been understood that sperm undergo extensive remodeling during epididymal transit. And given that sperm leaving the testes are transcriptionally and translationally silent (Asano et al., 2010), any changes to protein, RNA content, and otherwise, are presumably the result of external factors acting on sperm during the process of epididymal maturation.

The epididymis secretes proteins, ions, and exosomes and microvesicles – collectively referred to as “epididymosomes” – in a region-specific manner that has been documented through protein and gene expression patterns (Rinaldi et al., 2020; Zhou et al., 2018; Nixon et al., 2015A; Nixon et al., 2015B; Cornwall and Hann, 1995). It is through this regional progression that sperm acquire factors necessary to support sperm function (Zhou et al., 2018; Skerget et al., 2015; Johnston et al., 2005;). For example, it has been documented that sperm modify a number of surface proteins including ADAM7 (Oh et al., 2009), which plays a role in sperm-egg adhesion properties, SPAM1 (Martin-DeLeon, 2006), which encodes a hyaluronidase enabling sperm to penetrate the cumulus cell layer surrounding the oocyte, and CRISP1 (Robert et al., 2008), which is implicated in sperm-egg fusion. These massive changes to surface membranes (Tecle and Gagneux,
2015; Kirchhoff et al., 1997;) include an extensive, dense coating of sugar-rich molecules.

Sperm additionally undergo a considerable shift in small RNA content through epididymal transit – that is, predominantly piRNAs in the testes to tRNA fragments (tRFs), miRNAs, snRNAs, snoRNAs, rRNA fragments in addition to piRNAs. This transition is understood to occur through epididymosomes produced by the epithelial cells of the epididymis fusing with sperm during epididymal transit to deliver a complex payload of proteins and small RNAs to maturing sperm. These epididymosomes are also produced in a region-specific manner with distinct profiles of proteins, and small RNAs – notably miRNAs and tRNA fragments (tRFs) (Stanger et al., 2020; Trigg et al., 2019; Nixon et al., 2019; Sharma et al., 2018; Conine et al., 2018; Sharma et al., 2016). Importantly, epididymosomes provide a mechanism for epididymal epithelial cells to communicate with sperm.

1.6 Overview of this dissertation

In addition to delivering a haploid genome to the zygote, sperm also deliver additional epigenetic information in the form of chromatin structure, RNAs, and DNA methylation (Jenkins and Carrell, 2012). Over the past decade, it has become clear that sperm undergo extensive molecular remodeling in the epididymis. Here, we aim to better understand diverse aspects of epididymal maturation and its impact on the sperm epigenome. In this work, we focus on specific two aspects of this “window of opportunity” – cytosine methylation dynamics and the pursuit of capturing the “true”
small RNA content of mature sperm. Beyond the two major waves of cytosine methylation erasure and re-establishment, there have been hints that cytosine methylation may also exhibit changes during other periods of the life cycle, including during epididymal maturation of sperm – we assess this possibility of a “third wave” of DNA methylation reprogramming Chapter II. Given the enormous body of literature surrounding environment mediating small RNA content of sperm, it will be surprising to learn that the standard small RNA sequencing protocols are inadequate for capturing what we know to be the truth – in Chapter III of this work, we aim to remedy this deficiency.
CHAPTER II.

STABILITY OF THE CYTOSINE METHYLOME
DURING POST-TESTICULAR SPERM MATURATION IN MOUSE
2.1 ABSTRACT

Beyond the haploid genome, mammalian sperm carry a payload of epigenetic information with the potential to modulate offspring phenotypes. Recent studies show that the small RNA repertoire of sperm is remodeled during post-testicular maturation in the epididymis. Epididymal maturation has also been linked to changes in the sperm methylome, suggesting that the epididymis might play a broader role in shaping the sperm epigenome. Here, we characterize the genome-wide methylation landscape in seven germ cell populations from throughout the male reproductive tract. We find very few changes in the cytosine methylation landscape between testicular germ cell populations and cauda epididymal sperm, demonstrating that the sperm methylome is stable throughout post-testicular maturation. Although our sequencing data suggested that caput epididymal sperm exhibit a highly unusual methylome, follow-up studies revealed that this resulted from contamination of caput sperm by extracellular DNA. Extracellular DNA formed web-like structures that ensnared sperm, and was present only in sperm samples obtained from the caput epididymis and vas deferens of virgin males. Curiously, contaminating extracellular DNA was associated with citrullinated histone H3, potentially resulting from a PAD-driven genome decondensation process. Taken together, our data emphasize the stability of cytosine methylation in mammalian sperm, and identify a surprising, albeit transient, period during which sperm are associated with extracellular DNA.
2.2 INTRODUCTION

In addition to contributing a haploid genome to the next generation, germ cells also deliver epigenetic information to progeny that can impact early development and later phenotypes. In mammals, this is best characterized in the context of genomic imprinting, a situation where genes exhibit monoallelic expression from either the maternal or paternal allele. In the classic examples of imprinted gene regulation, the heritable marking of imprinting control regions relies on the covalently-modified cytosine derivative 5-methylcytosine (Hackett and Surani, 2013; Bartolomei and Ferguson-Smith, 2011; Feng et al., 2010; Bestor and Bourc’his, 2004). In mammals, cytosine methylation typically occurs in the context of CpG dinucleotides, and the genomic methylation landscape is copied every S phase by the maintenance methyltransferase DNMT1.

Cytosine methylation patterns in mammals undergo two major reprogramming events; the first during primordial germ cell development, and the second occurring upon fertilization. In the zygote, sperm methylation is rapidly erased shortly after fertilization, apparently via active demethylation, while oocyte methylation patterns are lost more slowly via passive demethylation (replication without maintenance methylation). However, a small number of genomic loci escape this demethylation process, including imprinting control regions and a subset of evolutionarily-young repeat elements (Smith and Meissner, 2013; Law and Jacobsen, 2010). The mechanisms responsible for protecting these loci from demethylation are still being uncovered, but recent studies implicate sequence-specific DNA binding proteins (ZFP57 and ZFP445) in maintenance of methylation levels at a subset of imprinting control regions (Shi et al., 2019; Takahashi
et al., 2019). Although the majority of the sperm methylation landscape is erased upon fertilization, the existence of escaper loci suggests the possibility that environmentally-regulated changes to the sperm methylome could play a role in modulating phenotypes in the next generation. Indeed, several studies have documented changes to sperm cytosine methylation in response to various diets or toxin exposures (Sun et al., 2018; Watkins et al., 2018; Holland et al., 2016; Radford et al., 2014; Wei et al., 2014), raising the question of how the sperm methylome is regulated by environmental conditions.

Intriguingly, an early study on sperm methylation suggested the surprising possibility of a third cycle of methylation reprogramming, occurring during post-testicular sperm maturation in the epididymis (Ariel et al., 1994). Briefly, using methylation-sensitive restriction enzymes, the authors showed that methylation levels at two genomic loci appeared to change as sperm entered the proximal, or caput, epididymis. Given recent findings that another epigenetic information carrier in sperm, the small RNA payload, is extensively remodeled during epididymal transit (Sharma et al., 2018; Sharma et al., 2016; Reilly et al., 2016; Nixon et al., 2015A; Nixon et al., 2015B), we envisioned the exciting possibility that the epididymis may play a broader role in control of the heritable sperm epigenome.

Motivated by the Ariel et al study, we therefore revisited cytosine methylation dynamics during post-testicular maturation, using the gold standard whole genome bisulfite sequencing (WGBS) to characterize the methylation landscape genome-wide, at single-nucleotide resolution, in seven germ cell populations from primary spermatocytes to vas deferens spermatozoa (Figure 1.1A). We found remarkably consistent methylation
profiles in five of the seven germ cell populations, with strong correlations between the various testicular germ cell populations and corpus and cauda epididymal sperm. This general persistence of methylation patterns was interrupted by caput epididymal sperm, which exhibited modest global hypomethylation accompanied by hypermethylation of germline-associated CpG islands.

We ultimately identified cell-free DNA, presumably derived from somatic cells of the epididymis, as the cause of the unusual “caput sperm” methylome–most definitively, we show that treating caput or vas deferens sperm with DNase I prior to sperm lysis restored the aberrant sperm methylation landscape to the same pattern seen in all other sperm populations from testicular through cauda epididymal sperm. Cell-free DNA was associated with the citrullinated histone H3 (citH3) that is characteristic of arginine deimination by peptidylarginine deiminase (PAD) enzymes, and, curiously, was detected only in virgin males. Taken together, our data support a static view of methylation patterns stably persisting throughout post-testicular sperm development, and reveal an intriguing but transient stage of programmed cell-free DNA production in the male reproductive tract.
2.3 RESULTS

**Genome-wide analysis of the sperm cytosine methylation landscape**

We set out to build on prior low-throughput studies documenting differences in cytosine methylation between testicular sperm and sperm obtained from various regions of the epididymis (Ariel et al., 1994). To this end, we collected seven germ cell populations from 10–12 week old FVB males: primary spermatocytes, two populations of round spermatids, and spermatozoa obtained from the caput, corpus, and cauda epididymis, and from the vas deferens (Figure 2.1A). For each population we collected samples from seven different males, isolated and pooled genomic DNA, and prepared libraries for whole genome bisulfite sequencing (WGBS), obtaining an average of ~300 million reads per pool. The resulting data recapitulate well-described features of the sperm methylome (Kubo et al., 2015; Wang et al., 2014; Hammoud et al., 2014; Molaro et al., 2011), such as a high overall level of methylation punctuated by hypomethylated CpG islands (Figure 2.1B and 2.1C) supporting the quality of our dataset.
Figure 2.1 Whole genome cytosine methylation in seven germ cell populations
A) Schematic of seven germ cell populations analyzed in this study: primary spermatocytes, early and late round spermatids from testis, and spermatozoa from the caput, corpus, cauda epididymis and from the vas deferens. Shown in grey are mature testicular spermatozoa, which were not analyzed in our WGBS dataset. B) Typical features of the germline methylome were reproduced in all seven datasets. Top panel shows average methylation profile for metagenes normalized to the same length, along with 2 kb of sequence upstream and downstream of genes. Bottom panel shows methylation data surrounding all annotated CpG islands. C) Box plots show methylation levels for all 200 bp tiles across the genome. As expected, the majority of samples are overwhelmingly methylated, with caput sperm and vas deferens sperm exhibiting modest hypomethylation relative to the other five samples. D) Scatterplots comparing methylation levels for consecutive sperm developmental stages. In each case, scatterplot shows methylation levels averaged for all 200 bp tiles (with at least 10 methylation-informative reads) across the genome. Overall, all samples exhibit robustly correlated methylation landscapes with one another, with the caput and vas deferens samples representing outliers with hundreds of hypo- and hyper-methylated tiles compared to other sperm samples. See also Figure 2.2 for all pairwise comparisons. E) Scatterplot comparing methylation levels in primary spermatocytes and cauda sperm, showing that the genomic loci that exhibit changes in methylation in caput sperm return to their original methylation levels later in the epididymis.
A transient methylation signature in the caput epididymis interrupts otherwise stable methylation throughout sperm maturation

We next turned to the question of how the sperm methylome changes over the course of epididymal transit. Comparing all seven populations, we noted similar overall methylation profiles, with global methylation interrupted by hypomethylated CpG islands in all seven samples (Figure 2.1B and 2.1C). Intriguingly, we found somewhat lower global methylation in caput sperm and, to a lesser extent, in sperm from the vas deferens. Examination of metagenes revealed that these two sperm populations deviated from the characteristic methylation profile observed in the other five samples, with lower methylation levels across coding regions, along with subtly increased methylation at promoters (Figure 2.1B).

To more systematically search for methylation changes between germ cell populations, we averaged methylation levels over 200 bp regions tiled across the genome and compared methylation levels in these tiles between all pairs of samples in our dataset (Figure 2.1D and Figure 2.2) Overall, we found that five germ cell populations exhibited nearly-identical methylation landscapes, with strong correlations between the methylation datasets for the three testicular samples, and for corpus and cauda sperm. Importantly, this means that the differences in methylation between round spermatids and caput sperm are reversed in corpus sperm, rather than these methylation changes being part of an ongoing process of progressive methylation maturation in the epididymis. In other words, the nearly-identical methylation profiles for testicular spermatocytes/spermatids and cauda sperm (Figure 2.1E) indicates that the sperm methylome is essentially unchanged
by the process of epididymal maturation, and strongly argues against a major role for the epididymis in modulating sperm methylation.

Below, we explore the unusual methylome of caput and vas deferens sperm. Given the similar behavior of vas deferens and caput sperm populations, we primarily focus in follow-up experiments on the more dramatic methylation changes observed in caput sperm, but we ultimately show that the aberrant methylation signatures of caput and vas deferens sperm result from similar processes.
Figure 2.2 Correlations between all seven germ cell methylation datasets
Scatterplots are shown as in Figure 2.1D and 2.1E, for all pairwise comparisons in this dataset.
Widespread derangements in CpG island methylation in caput epididymal sperm

In contrast to the nearly-identical methylation profiles obtained from testicular and corpus/cauda sperm, comparisons between caput sperm and any of these samples revealed large-scale changes in methylation across hundreds of 200 bp tiles (Figure 2.1D and Figure 2.2). Overall, as noted above, caput sperm were slightly less methylated than these other sperm samples—examination of hypomethylated loci in caput sperm revealed diffuse hypomethylation over a wide range of both coding and intergenic genomic loci. This signature was also detectable in metagene averages in the highly-methylated regions distant from CpG islands, with caput sperm exhibiting a small global deficit at these loci (Figure 2.1B, insets).

In addition to the widespread hypomethylation of the caput sperm genome, a large group of hypermethylated tiles was readily apparent in these scatterplots. Examination of these tiles revealed that they are largely associated with CpG islands. To visualize this, we calculated the average methylation across all annotated CpG islands; Figure 2.3A shows a scatterplot comparing CpG island methylation in round spermatids and caput sperm—nearly-identical results were obtained in comparisons between caput sperm and other testicular populations, or corpus/cauda epididymal sperm. Together, these analyses reveal extensive methylation changes at regulatory elements in caput sperm, primarily reflecting hypermethylation of CpG islands in this sample (dots above and to the left of x = y) with a relatively small number of hypomethylated islands in caput sperm. This is further illustrated in Figure 2.3B which documents methylation dynamics for differentially-methylated CpG islands across all seven germ cell populations, highlighting
the dramatic changes in CpG island methylation in caput sperm (and, to a lesser extent, in vas deferens sperm). Again, we note that regulatory elements that exhibit aberrant methylation in caput sperm universally return to the spermatocyte/spermatid methylation patterns in the corpus and cauda samples. In other words, whatever methylation changes apparently occur in caput sperm are soon reversed in later sections of the epididymis (with the exception of the vas deferens, discussed later).
Figure 2.3 CpG island hypermethylation in caput and vas deferens sperm
A) Scatterplot showing average methylation levels across all annotated CpG islands, comparing late round spermatids and caput epididymal sperm. Although the majority of CpG islands are still hypomethylated in both samples (lower left corner), caput sperm are characterized by hundreds of abnormally hypermethylated CpG islands (dots above the x = y diagonal). B) Heatmap for CpG islands that are more than 20% hypo- or hypermethylated in caput sperm relative to other samples. Each row represents the methylation values of a single CpG island, normalized relative to the average methylation % across all seven samples. Islands are sorted from hyper-methylated in caput sperm to hypomethylated. As caput-hypomethylated CpG islands are enriched for those located near genes with known reproductive functions (Figure 2.4), notable genes are annotated along the right of the heatmap. We note that the majority of caput-hypomethylated CpG islands (bottom of the heatmap) are located within transcribed regions, and may therefore reflect the modest general hypomethylation observed over transcribed regions (Figure 2.1B and 2.1C)
What biological pathways are affected by epididymal methylation dynamics? To address this question, we sought gene ontology categories enriched in genes associated with caput sperm hypo- and hyper-methylated CpG islands (Figure 2.4). Hypomethylated islands in caput sperm are significantly enriched for a handful of processes related to neuronal function (eg, neuronal cell body). However, these enrichments appear to be primarily driven by the fact that caput-hypomethylated CpG islands are found within transcribed regions—the hypomethylation at these loci presumably reflects the more general hypomethylation characteristic of caput sperm (Figure 2.1B and 2.1C). More intriguingly, hypermethylated islands are enriched for a variety of annotations associated with meiosis- and sperm-specific functions, including piRNA biogenesis, ion signaling, and reproductive process.
For all CpG islands exhibiting >20% methylation differences between caput and cauda sperm, nearest genes were identified and enriched gene ontology categories were identified using Funcassociate (Berriz et al., 2009). Bar plots show p values (expressed as -log10) for selected categories enriched among hyper (red) and hypo (green) methylated CpG islands.
To validate our genome-wide dataset, and to develop a set of cost-effective targets for follow-up mechanistic studies, we analyzed methylation levels at a number of target loci (Figure 2.5) by pyrosequencing of bisulfite-converted DNA. As shown in Figure 2.6, methylation differences between caput and cauda were robustly reproducible in many additional, independent pairs of sperm samples. Moreover, the caput methylome proved to be quite stable over an animal’s lifespan—methylation at two target loci assayed remained abnormal in caput sperm obtained from ten month old animals (Figure 2.7).
Figure 2.5 Genomic context for targeted pyrosequencing followup

WGBS data for genomic loci selected for targeted follow-up. At each locus, red and blue dashes represent individual reads for a given CpG, with red and blue showing methylated or unmethylated reads respectively. Top panels for each region show a wider view of the genomic context, with boxes indicating regions shown in the zoom-in bottom panels. For these five regions, methylation levels obtained from the WGBS dataset are shown for caput and cauda sperm samples, along with methylation levels obtained in follow-up pyrosequencing validation. For all five loci, pyrosequencing qualitatively confirmed the methylation trends observed in WGBS; while there was some quantitative disagreement (eg 83% vs 42% for methylation at Tdrd12 in caput sperm), we note that values inferred for small numbers (~4–5) of CpGs in the WGBS dataset are expected to be somewhat noisy given the relatively low sequencing depth.
Figure 2.6 Pyrosequencing validation of WGBS data
Pyrosequencing data are shown for five loci which exhibited caput-specific methylation levels in our WGBS dataset (Figure 2.5). For each locus, pyrosequencing data are shown for 4–6 samples each of caput, corpus, and cauda epididymal sperm, as well as vas deferens sperm, as indicated. Animals were 10–14 weeks of age at time of collection. Error bars show standard deviation across multiple CpGs assayed at each locus. In all cases the pyrosequencing data recapitulate the differences between sperm populations identified in our WGBS dataset, and differences between caput and cauda methylation levels are highly statistically-significant (t test p values ranging from 1.2e-4 for Tbx2 to 2.0e-6 for Tdrd12) for all five loci.

Figure 2.7 The caput methylation profile is stable for at least ten months in unmated animals
Pyrosequencing data for the two indicated target loci in four unmated males at ten months of age, confirming that the caput methylome is stable throughout a typical male’s lifespan.
These data validate our overall dataset, and we focus on these target loci in the mechanistic follow-up studies described below. Importantly, although we considered the possibility that a signature of hypermethylation of germline regulatory elements might reflect contamination by somatic cells, we ensured that the methylation differences between caput and cauda sperm were robust to several sperm purification protocols including one based on detergent washing of caput epididymis luminal contents, and an alternative Percoll-based isolation of caput sperm in the absence of detergent treatment (see Methods). For all samples we routinely determined, based on the characteristic hook-shaped morphology of the murine sperm head, that our caput sperm preps were >99% free of somatic cell contamination.
The caput sperm methylome is stable in multiple buffer conditions

We next sought to identify the mechanistic basis for the widespread changes in methylation observed in caput sperm, to explore the molecular basis for what would represent a relatively rare case of fully replication-independent cytosine demethylation. One of the signature functions of the mammalian epididymis is to provide a variety of highly distinctive luminal microenvironments that serve a multitude of functions in supporting sperm maturation and preventing premature activation (Gervasi et al., 2017; Breton et al., 2016). It is well known that the ionic composition differs significantly between different luminal compartments, as for example pH and calcium levels vary dramatically between caput and cauda, and the extensive gene expression differences between different segments (Domeniconi et al., 2016; Johnston et al., 2005) imply that many other metabolites will also differ in concentration throughout the epididymis.

We therefore set out to test the hypothesis that sperm methylation dynamics result from “pre-loaded” genome-associated DNA modification enzymes (DNMTs for methylation, TET enzymes for demethylation) whose ongoing activity could be inhibited or activated by changes in either substrate levels (SAM, alpha-ketoglutarate, iron oxidation status, etc.) or buffer conditions (DNMT3a activity is highly pH-dependent (Holz-Schietinger and Reich, 2015)). To test this, we attempted to recapitulate the caput to cauda methylation changes by incubating purified caput sperm in various buffer conditions meant to mimic the cauda epididymal lumenal environment (Figure 2.8). However, none of the buffer conditions tested were able to substantially influence the caput sperm methylome.
Figure 2.8 Caput sperm methylation status is stable under many different buffer conditions

Pyrosequencing data for caput and cauda sperm samples. Caput sperm samples were either processed for genomic DNA shortly after isolation, or were incubated for four hours at 37°C in various buffer conditions, as indicated. Buffer conditions were based on either Donners Basic (DB) or Donners Complete (DC), and were supplemented with various levels of NaHCO3, sodium pyruvate, sodium DL-Lactate, or adjusted to pH 6.5, 7.0, or 7.4 (red triangles). Although not indicated in the figure, cauda sperm samples (right) included samples subject to most of the buffer incubations used for caput sperm samples, none of which affected methylation in these samples.
The caput sperm methylome is lost following mating

During the course of these studies, we found one animal in which the methylation level at our target genes was nearly identical for caput and cauda sperm. Further investigation revealed that a female had accidentally been weaned into an otherwise all-male cage, suggesting that mating might influence the methylation dynamics described above. Indeed, direct testing of this hypothesis confirmed that methylation of our target genes was nearly identical in caput and cauda sperm obtained from animals who had successfully sired offspring (Figure 2.9). This was due to the caput methylation profile shifting to match that of cauda sperm; cauda sperm methylation was totally unaffected by the male’s mating status.

We considered and rejected two hypotheses regarding the change in methylation status of caput sperm in mated animals: 1) that conversion of the testicular sperm methylome to the unusual caput state might require extended incubation of newly-arrived sperm in the caput luminal environment, and 2) that the virgin caput methylation profile reflects sperm originating from the unusual first meiotic wave of spermatogenesis and somehow being captured in the caput epididymis. The first hypothesis was rejected based on the finding that the cauda-like methylation program in caput sperm from mated animals was stable for at least nine weeks following mating (Figure 2.9).
Figure 2.9 The unusual methylation program in caput sperm is lost after mating
Pyrosequencing data shown as in Figure 2.6, for caput and cauda sperm obtained from virgin males or males sacrificed at varying times after confirmed mating.
We next considered the possibility that caput sperm obtained from virgin animals represent an unusual first wave of spermatogenesis (Grive et al., 2019) characterized by an atypical methylation program. This hypothesis would require an unlikely process—in which first wave sperm are slowed or arrested in the proximal epididymis, allowing subsequent waves of sperm to pass into the corpus and cauda epididymis—but we could not rule out the hypothesis a priori. However, enriching first wave sperm by isolation of testicular spermatozoa from 35 day-old animals revealed the same methylation levels at our validation loci as those observed for testicular and cauda sperm from older animals (Figure 2.10) refuting the hypothesis that first wave sperm carry an unusual methylome.

**Figure 2.10 Enriching for the first wave of testicular sperm**
Testicular sperm obtained from prepubertal males (thus enriched with the first wave of spermatogenesis) exhibit similar methylation to cauda, not caput, sperm from mature animals. For each target locus, data are shown for testicular sperm obtained from 35 day or 45 day old males, as well as caput and cauda epididymal sperm obtained from 10–14 week old animals.
Contamination of caput epididymal sperm by cell-free DNA

What then could account for the unique methylation program observed in caput sperm from virginal males? Careful inspection of our genomic DNA preparations revealed that although testicular and cauda sperm samples were characterized by uniformly high molecular weight genomic DNA, there was a dim additional “cloud” of lower molecular weight DNA in the caput sperm. This suggested the possibility that caput sperm might be contaminated by cell-free DNA. Indeed, DAPI staining of caput sperm preparations revealed not only the expected hook-shaped sperm heads, but at higher exposure times we noted the presence of web-like DNA structures, often ensnaring multiple sperm (Figure 2.11).
Figure 2.11 Cell-free DNA in the caput epididymis of virgin males is responsible for the caput methylome

DAPI staining of caput sperm samples. For each sample, left panel shows DIC to highlight sperm locations, middle and right panels show DAPI staining, with far right panel showing longer exposures. Yellow arrows indicate examples of web-like cell-free DNA in caput sperm samples.
These webs were not detected in cauda epididymal sperm preparations. To visualize this extracellular DNA in situ, we stained histological sections of several epididymal regions with DAPI. Intriguingly, in both caput sections and in the vas deferens we find a DAPI-staining rim associated with the apical region of the epithelium (Figure 2.12). DAPI-positive rims were not observed in the cauda epididymis, and were diminished or lost from the caput epididymis and vas deferens following mating (Figure 2.12; bottom panels). We cannot be certain that these DAPI rims are the source of the free DNA that contaminates caput sperm preps—the DAPI rims appear to coincide with the cell bodies of the epididymal epithelium—but it is notable that these rims were only observed in the same samples (caput but not cauda epididymis, virgin animals but not mated animals) that suffer from cell-free DNA contamination.
Figure 2.12 DAPI-staining rim associated with the apical region of the epithelium of caput and vas deferens in virgin males
In situ DAPI staining of the epididymis. Images show DAPI-stained histology sections of the initial segment, caput epididymis, cauda epididymis, and vas deferens of either virgin or mated males, as indicated. Yellow brackets indicate examples of DAPI-positive rim associated with the apical regions of the epithelium of initial segment or caput samples from virgin males. Arrow shows a lumen nearly completely filled with this material.
To definitively test whether extracellular DNA is responsible for the methylation profile of virginal caput sperm, we treated various sperm samples with DNase I to eliminate any extracellular DNA prior to extraction of genomic DNA for pyrosequencing analysis. Remarkably, DNase I treatment completely restored the caput sperm methylome to the methylation levels observed in testicular or cauda epididymal sperm (Figures 2.13, 2.14, and 2.15), demonstrating that contaminating cell-free DNA present in the caput epididymis is responsible for the unusual caput methylome observed in virgins. To ensure that this finding was not idiosyncratic to the FVB/NJ strain background used throughout this study, we also carried out pyrosequencing in epididymal and vas deferens sperm samples obtained from C57Bl6/J males (Figures 2.14 and 2.15) again confirming 1) that caput and vas deferens sperm exhibited aberrant methylation at our target genes in this strain background, and 2) that this aberrant methylation profile was eliminated by DNase treatment of sperm preparations prior to sperm lysis and genomic DNA extraction.
Figure 2.13 Cell-free DNA is responsible for the abnormal caput methylome
Pyrosequencing data are shown for caput and cauda sperm samples, either mock treated or pre-treated with DNase I prior to genomic DNA purification. At both loci, differences between untreated caput and cauda sperm are again highly significant ($p < 1\times 10^{-5}$), consistent with Figure 2.6, while there are no significant differences between either DNase-treated caput sperm and cauda sperm, or between treated and untreated cauda sperm ($p > 0.05$ in all cases).
Figure 2.14 DNase-sensitive aberrant methylation is consistently observed in multiple strain backgrounds
Pyrosequencing data for Irs1 are shown for the indicated sperm samples obtained from three 10–14 week old C57Bl6/J males. Left samples show methylation levels for genomic DNA isolated from washed sperm pellets, while right samples show data for genomic DNA isolated following DNase treatment of sperm pellets. As observed for the FVB strain background used throughout the rest of the manuscript, untreated caput and vas deferens sperm preps from C57 animals also exhibit aberrant methylation which is completely corrected following DNase I treatment. See also Figure 2.15.
Figure 2.15 Resolution of the caput methylome by DNase I treatment
A) Pyrosequencing data show methylation at the three indicated target loci in testicular spermatozoa, caput, corpus, and cauda epididymal sperm, and vas deferens sperm, as indicated. All samples were obtained from 10–14 week old FVB males, and sperm were DNase-treated prior to genomic DNA isolation. In all cases DNase treatment completely eliminated methylation differences between the various samples. B) Data for the indicated samples obtained from 10–14 week old C57 males. Data here are shown for samples either mock-treated (left samples) or DNase-treated (right samples, as indicated) prior to genomic DNA extraction.
Based on its impact on caput sperm methylation, it seems likely that the cell-free DNA originates from somatic cells rather than germ cells. We note that the cell-free DNA observed in our sperm preparations is reminiscent of extracellular DNA released by neutrophils, known as Neutrophil Extracellular Traps, or NETs (Tilvawala et al., 2019; Sollberger et al., 2018; Brinkmann et al., 2004). The process of NETosis is associated with increased peptidyl arginine deiminase (PAD) activity, resulting in arginine deimination on histones and other proteins, leaving behind citrulline and leading to massive chromatin decondensation. Examination of RNA-Seq data from throughout the epididymis (Rinaldi et al., 2020) confirmed expression of Padi2 in this tissue, largely confined to principal cells of the caput epididymis and vas deferens (Figure 2.16A).

Moreover, consistent with the caput cell-free DNA being produced downstream of PAD-dependent deimination, we found robust citH3 staining of our caput sperm preparations, with negligible signal in cauda sperm preps (Figure 2.16B).

Taken together, our data reveal that extracellular DNA, potentially produced via a PAD-associated process, contaminates sperm obtained from the caput epididymis and vas deferens of young males, and that this cell-free DNA is depleted following mating.
Figure 2.16 Cell-free DNA is produced via a NETosis-like process
A) Expression of PAD-encoding genes in the murine epididymis and vas deferens. RNA-Seq data are from (Rinaldi et al., 2020). B) Caput epididymal cell-free DNA is associated with citrullinated histone H3. Caput and cauda sperm samples were stained for DAPI (blue) and for citH3 (green), as indicated.
2.4 DISCUSSION

Here, we characterized the cytosine methylation landscape during late stages of spermatogenesis, and throughout post-testicular sperm maturation in the epididymis. Most importantly, these data reveal that the sperm methylome is quite stable during the process of post-testicular maturation. This is clearest in (Figure 2.1E) which shows the robust correlation between the methylation patterns in primary spermatocytes and in mature cauda epididymal sperm. Our data therefore do not support a model in which sperm methylation is extensively remodeled during epididymal transit. Although we cannot rule out subtle methylation changes occurring in epididymal sperm in response to environmental stressors, the absence of changes in this dataset strongly supports the classic view of germline cytosine methylation patterns as largely static in the absence of replication-coupled remodeling events.

Our study also revealed a surprising feature of the male reproductive tract, in which cell-free DNA is present in the caput epididymis and the vas deferens of virgin males. This DNA clearly derives from somatic cells rather than dying sperm, since the methylation profile of virgin caput sperm departs from the characteristic germline methylome to incorporate aspects more typical of somatic cells. This is apparent in the hypermethylation of key germline regulatory elements, which are unmethylated and active during spermatogenesis but methylated and repressed in most somatic tissues. This is also clear at imprinting control regions, which are either completely methylated or unmethylated in the sperm genome, but are 50% methylated in somatic cells—the contaminated caput methylome exhibits a shift towards 50% methylation at many
imprinting control regions (Figure 2.17). Based on the methylation levels at imprinted regulatory elements, as well as changes in DNA yield following DNase I treatment of caput sperm prior to gDNA isolation, we roughly estimate that in our hands caput sperm genomic DNA is contaminated by ~20% somatic cells. Of course, this level of contamination by intact cells would readily be detectable by inspection of our caput sperm preparations, emphasizing the fact that this contamination arises from cell-free DNA rather than intact somatic cells. That said, it remains plausible that the contaminating DNA comes from cells that lyse during dissection and sperm preparation, although it would be unusual for such DNA to be associated with citrullinated histones.
Figure 2.17 WGBS data for selected imprinted genes
At each locus, red and blue dashes represent individual reads for a given CpG, with red and blue showing methylated or unmethylated reads respectively - for selected imprinted genes. Notably, while testicular germ cells and corpus and cauda sperm exhibit the expected 0% or 100% methylation, our caput sperm (and to a lesser extent vas deferens sperm) data are closer to the 50% methylation expected of somatic cells.
As our study was initially motivated by the findings of cytosine methylation changes occurring during post-testicular sperm maturation as reported by Ariel et al, we show WGBS data for the two loci documented in detail in that study (Figure 2.18). However, the assay used by Ariel et al – digestion using a methylation-sensitive restriction enzyme followed by PCR across the cut site – reports on a single CpG, and our data are not deeply-sequenced enough to confidently assess methylation levels at the two individual CpGs in question. That said, pyrosequencing at one locus—Pgk2—did not recapitulate the findings reported in Ariel et al: we found no difference in methylation of this one CpG in caput vs cauda sperm (31% vs 27%), while Ariel et al reported minimal methylation at this site in caput sperm along with higher methylation in cauda sperm. There are countless potential technical differences that might account for this, including strain differences (CD1 vs FVB and C57), mating status and age (not specified in Ariel et al), sperm preparation (Ariel et al used sonication to deplete somatic cells, we used both detergent and Percoll in different preparations), or genomic DNA extraction (we used 150 mM DTT for sperm lysis, Ariel et al used significantly lower reducing agent—0.001% v/v β-ME), and of course the methylation readout (quantitative pyrosequencing vs qualitative PCR following methylation-sensitive restriction digestion).
Figure 2.18 Loci studied in Ariel et al, 1994

It is unclear which cells are responsible for production of cell-free DNA. Extracellular DNA traps are most famously produced by neutrophils during a process known as NETosis, but they have been reported in multiple inflammatory cell types (Yousefi et al., 2008; Pertiwi et al., 2019). In this regard it is intriguing that sperm have been shown to induce NETosis when coincubated with leukocytes (Zambrano et al., 2016). Nonetheless, we favor the hypothesis that the extracellular DNA in the caput epididymis originates in principal cells of the epididymal epithelium. This is based on 1)
the rarity of neutrophils in the epididymis (Rinaldi et al., 2020); 2) the expression of *Padi2*, rather than *Padi4* which is more typically associated with NETosis, in our RNA-Seq dataset (Figure 2.16A); 3) our inability to detect substantial staining of the neutrophil marker myeloperoxidase in the caput cell-free DNA (not shown); and 4) the expression of *Padi2* specifically in principal cells in our single-cell atlas of the epididymis (Rinaldi et al., 2020). Although in principle the identity of the cell of origin for this DNA could leave a signature in our methylation dataset, examination of methylation at several immune cell marker genes proved inconclusive.

Whatever the cell of origin of the extracellular DNA here, it seems likely that it is produced via a PAD-dependent genome decondensation event analogous to the one that drives the production of Neutrophil Extracellular DNA Traps, or NETs (Sollberger et al., 2018). This hypothesis is motivated by the expression of *Padi2* specifically in the two regions exhibiting cell-free DNA, and by the confirmed presence of abundant citH3 staining in caput sperm preps (Figure 2.16). This raises the question of what purpose extracellular DNA in the epididymal lumen might potentially serve. Extracellular DNA plays a key role in defense against pathogens in several contexts–NETs in the mammalian bloodstream entangle pathogens and aid in their engulfment by antigen-presenting cells (Sollberger et al., 2018) while extracellular DNA in pea (*P. sativum*) roots was shown to protect against fungal infections (Wen et al., 2009).

While a similar role could be imagined for extracellular DNA in the epididymis helping to prevent ascending infections, it is notable that this material is present only in virgins but lost after mating (Figure 2.9). As infections would presumably be more likely
in sexually-active animals than in virgins, we therefore disfavor this hypothesis, although it is plausible that the system serves to ensure the reproductive machinery is protected until a first mating is achieved.

Alternatively, we speculate that the DNA webs observed here could ensnare sperm rather than pathogens, potentially playing a role in removal of early rounds of defective sperm, or masking sperm from inducing an immune response. Of course, the simplest hypothesis would be that the extracellular DNA documented here is a vestige of some earlier developmental process–death of epithelial cells during epididymal morphogenesis, for instance–and serves no biological function. That said, the use of PAD-dependent cell death would be unusual for typical programmed cell death in development.

Although the function, if any, served by extracellular DNA in the epididymis is unclear, the more pressing implication of our findings is that they raise two important technical considerations for reproductive epigenetics studies. First, one important finding here is that at least some aspects of sperm populations are different between virgins and mated males, raising the question of whether other features of the epididymal sperm epigenome–often assayed in virgin males–are stable throughout reproductive life. For example, we and others have previously documented substantial differences between the small RNAs carried by caput and cauda epididymal sperm, with a variety of genomically-clustered microRNAs, and a subset of tRNA fragments such as tRF-ValCAC, being far more abundant in cauda sperm than in caput (Sharma et al., 2018; Sharma et al., 2016; Nixon et al., 2015A; Nixon et al., 2015B). We therefore repeated small RNA-Seq for caput and cauda sperm samples obtained from males three or six weeks after mating.
Importantly, we confirmed the same overall differences between these sperm samples in their small RNA payload (Figure 2.19), demonstrating that at least this observation is not an artifact of some transient early process in virgins.

That said, other aspects of the sperm maturation process may differ between virgins and mated males, emphasizing the importance for reproduction studies to explicitly state the mating status of males being used.

Secondly, our work has important technical implications for studies of the sperm methylome, as future efforts focused on cytosine methylation in sperm must contend with the possibility of contamination by cell-free DNA. Although most such studies do not utilize caput epididymal sperm, it is well known that different labs differ in whether “mature sperm” used for molecular studies or IVF are isolated only from the cauda epididymis, or from a mixture of cauda and vas deferens. Given our findings here with vas deferens sperm, it is clear that dissections that capture variable lengths of the vas deferens will result in variability in sperm methylation due to the contamination of these samples by DNA with a somatic cell methylation program.

In addition, although we found contaminating DNA specifically in the caput and vas of virgins, it is plausible that cell-free DNA could be more persistent or produced in different parts of the epididymis in other strain backgrounds or in males subject to different diets or stressors. A simple solution to this issue in future studies would be to treat all sperm samples with DNase I prior to genomic DNA isolation for methylation analyses.
Several previous studies have documented significant differences between the RNA payload of caput and cauda sperm, including dramatically lower levels of genomically-clustered microRNAs in caput sperm relative to both testicular and cauda sperm. To determine whether this unusual small RNA profile is unique to caput sperm obtained from virgins, we obtained caput and cauda epididymal sperm from virgin males as well as males three and six weeks after successful mating, and characterized small RNAs by deep sequencing. A-B) Scatterplot compares small RNA levels in virgin caput sperm to levels in caput sperm 3 (A) or 6 (B) weeks after mating. Data are shown for all small RNAs with an abundance of at least 10 ppm in virgin caput sperm. C) Scatterplot shows enrichment/depletion in cauda sperm vs. caput sperm (calculated as \( \log_2(\text{Cauda}+1)/(\text{Caput}+1) \)) for virgin males (x axis) compared to males 6 weeks after mating (y axis). Overall enrichments for small RNAs in cauda or caput sperm remain highly correlated after mating. D) Example of cauda-enriched microRNAs from the X-linked miR-465 and miR-880 clusters. Enrichment in cauda sperm, and absence from caput sperm, was unaffected by mating status.
2.5 METHODS

Mice

Unless otherwise specified, tissues were obtained from 10–12 week old male FVB/NJ or C57Bl6/J mice.

Dissection and sperm purification

FVB mice, euthanized at 10–12 weeks of age (unless otherwise noted) according to IACUC protocol, were dissected into four segments that roughly corresponded to caput, corpus, cauda, and vas deferens. Caput, corpus, and cauda epididymis as well as vas deferens were placed into Donners complete media (Hisano et al., 2013) and tissue was cleared of fat and connective tissue before incisions were made using a 26G needle while keeping the bulk tissue intact. Tissue was gently squeezed allowing sperm to escape into solution. After incubation at 37°C for 1 hour, sperm containing media was transferred to a fresh tube and collected by centrifugation at 5000rpm for 5 minutes followed by a 1X PBS wash. To eliminate somatic cell contamination, sperm were subjected to a 1mL 1% Triton X-100 incubation 37°C for 15 mins with 1500 rpm on Thermomixer and collected by centrifugation at 5000rpm for 5 minutes. Somatic cell lysis was followed by a 1x ddH2O wash and 30 second spin 14000 rpm to pellet sperm.

Isolation of caput sperm using a discontinuous Percoll gradient

Sperm collection from the caput epididymis is performed as described above, but without somatic cell lysis. Instead, caput sperm are purified using a Percoll gradient as described
(Krapf et al., 2012) to separate somatic cells away from sperm. Briefly, the caput sperm suspension is carefully layered over a discontinuous Percoll gradient containing 45% Percoll (upper phase) and 90% Percoll (lower phase). After centrifugation for 25 min at room temperature (650xg), the interphase containing caput sperm is washed with 1x PBS and prepped for downstream analysis.

**Testicular spermatocyte and spermatid isolation**

For each isolation, two testes were acquired from one FVB/NJ mouse at 10–12 weeks of age. Cell suspension was prepared by incubating the testes without their tunica albuginea in 5 ml elutriation buffer (100 mM NaCl, 45 mM KCl, 6 mM Na$_2$HPO$_4$, 0.6 mM KH$_2$PO$_4$, 0.23% Sodium DL-Lactate, 0.1% Glucose, 0.1% BSA, 0.011% Sodium Pyruvate, 1.2 mM MgSO$_4$ and 1.2 mM CaCl$_2$) containing 25 μg/ml liberase (Roche Diagnostics GmbH) for 30 min at 37°C with gentle agitation every 5 min. The cell suspension was mixed by pipetting 20 times with a 10-ml plastic pipette. After homogenization by pipetting 10 times through a P1000 pipette, the single cell suspension was filtered twice through a 40-μm cell strainer (Fisher Scientific) on ice and centrifuged at 1500 rpm at 4°C for 10 min, and then the pellet was resuspended with 20 ml elutriation buffer. Separation of testis cell populations was performed by centrifugal elutriation using a JE-5.0 elutriation system and a 4-ml standard elutriation chamber (Beckman Coulter). The assembly of the system followed the manufacturer’s instruction. The precise elutriation conditions are as follows: Fractions 1–3 were run at 3000 rpm and fractions 4–5 were run at 2000 rpm. The flow rate was 14, 18, 31, 23, and 40 ml/min for
fractions 1–5, respectively. During elutriation, the elutriation chamber was maintained at 4°C and the cells were collected into 50-ml conical polypropylene tubes that were packed on ice. Cells in tubes of fractions 3 to 5 were pelleted by centrifugation at 1500 rpm at 4°C for 10 min. All pellets from the same fraction were combined and resuspended in 200 μl of elutriation buffer. Percoll (Sigma-Aldrich) gradient (23–35%) was prepared by using a Gradient Master 108 (Biocomp) following standard program: S1/1, 2:26 (time), 82.0 (angle), 13 (rpm). After loading the cell suspension, centrifugation was carried out with SW40Ti Rotor (Beckman Coulter) at 11,000 rpm at 4°C for 15 min. The cells were then collected in 15-ml conical polypropylene tubes and pelleted by centrifugation at 1500 rpm at 4°C for 15 min. The cell pellets were stored at -80°C and ready for DNA extraction. Based on cell morphology and small RNA data, Fraction 5 was determined to correspond to primary spermatocytes, with fractions 4 and 3 being relatively early and late round spermatids, respectively.

For isolation of the first wave of testicular spermatozoa from postnatal day 35, a cell suspension was prepared by dissecting testes from postnatal day 35 males into 35mm dish containing 150 mM NaCl, removing tunica albuginea, and dissociating tissue using 22G needle and 3 mL syringe with 150 mM NaCl. Cell suspension was then pipetted into a 15 mL conical tube to allow for large tissue to settle. Once settled, top 1mL was loaded onto 52% isotonic Percoll solution. Samples were ultracentrifuged at 15,000 rpm for 10 min at 10 ℃. Following ultracentrifugation, pellet was carefully isolated, washed with 150 mM NaCl to remove residual Percoll, resuspended in PBS, and assessed for purity.
using microscopy (see S1A Fig in Sharma et al., 2018 for typical preparation) before undergoing the genomic DNA isolation protocol.

**Genomic DNA isolation**

750 μL Extraction Buffer (4.24M Guanidine Thiocyanate, 100 mM NaCl, 1% N-Laurylsarcosine, 150 mM freshly prepared DTT, 200 μg/mL Proteinase K) was added to sperm pellets and incubated for 2 hours at 56 °C with shaking. Samples were then allowed to equilibrate to room temperature after which 600 μL isopropanol was added to precipitate DNA followed by a 15 mins spin at 14,000rpm. Supernatant was carefully discarded and pellets were washed 2x with 80% EtOH, dried, and resuspended in 10 mM Tris pH8 (500 μL) with 5 μL RNase A (Qiagen 19101) for 2 hours after which samples were treated with 5 μL Proteinase K (Qiagen 19131) overnight. Following phenol chloroform extraction (UltraPure Phenol:Chloroform:Isoamyl Alcohol ThermoFisher 15593031) using phase lock gel tubes (Quantabio 10847–802), aqueous phase was transferred to a fresh tube containing 1.5 μL glycogen (20 mg/mL), 1.3 μL 5 M NaCl, and 100% EtOH. Genomic DNA integrity was determined using NanoDrop and quantified for downstream applications using Qubit (DNA BR).

**Whole genome bisulfite sequencing**

For each population we collected sperm samples from seven different males, isolated genomic DNA, and prepared libraries for whole genome bisulfite sequencing (WGBS); Bisulfite converted DNA was prepared according using the EZ DNA Methylation-
Lightning Kit (Zymo D5030). Library construction was performed using Accel-NGS Methyl-Seq DNA Library Kit (Swift) according to the manufacturer’s instructions. 30 ng bisulfite converted DNA was used and barcoded following the (Swift) manufacturer’s instructions including 6 rounds of PCR. All final libraries were analyzed by Fragment Analyzer and quantitated using the Qubit as well as the KAPA Library Quantitation qPCR kit.

**Data analysis**

Sequences were trimmed with Trim Galore (v0.4.4; Cutadapt v1.9.1)—Swift libraries were trimmed by 10 bp from their 5’ ends for both R1 and R2. The non-CG methylation levels were consistently very low (~0.2%) indicating good bisulfite conversion rates. The resulting trimmed sequences were mapped to the mouse GRCm38 genome using Bismark (Krueger and Andrews, 2011) (v0.17.0); CpG methylation calls were extracted and analysed using SeqMonk (www.bioinformatics.babraham.ac.uk/projects/seqmonk/). Analyses in Figures 2.1 and 2.2 were carried out using evenly-spaced 200 bp tiles, using only tiles with at least 1 CpG and at least 10 methylation-informative reads in every one of the seven datasets. For analyses in Figures 2.3B and 2.4 each CpG island was assigned to the nearest gene based on proximity to the transcription start site (within 50 kb).

**Pyrosequencing**

Pyrosequencing was performed using the PyroMark Q24 (Qiagen: 9001514) according to the manufacturer’s instructions. Bisulfite converted DNA was prepared using the EZ
DNA Methylation-Lightning Kit (Zymo D5030). PCR was performed with 10–20 ng of bisulfite-converted material for each locus of interest (Table 2.1). Primers (IDT) were designed using the PyroMark Assay Design software (Qiagen).

Table 2.1 Sequences of pyrosequencing primers

<table>
<thead>
<tr>
<th>Region of Interest</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Irs1_F1</td>
<td>TGATGGGAAATGGTAGGAGATGT</td>
</tr>
<tr>
<td>Irs1_R1_BIO</td>
<td>CTACTCCCAATAACCACTACATTTT</td>
</tr>
<tr>
<td>Irs1_Seq1</td>
<td>AAATGGTAGGAGATGTG</td>
</tr>
<tr>
<td>Irs1_F2</td>
<td>GTGTTTTAGGGAATTTGGGGGTGAT</td>
</tr>
<tr>
<td>Irs1_R2_BIO</td>
<td>CATAACTCCACTCAAACATCACTCT</td>
</tr>
<tr>
<td>Irs1_Seq2</td>
<td>GGGAATTGGGGTGATA</td>
</tr>
<tr>
<td>Tdrd6_F</td>
<td>GGATAGGGGTAGGATGTGATTG</td>
</tr>
<tr>
<td>Tdrd6_R_BIO</td>
<td>CTAAACAACCTACTCCAAATCTCC</td>
</tr>
<tr>
<td>Tdrd6_Seq1</td>
<td>GTGTTAGTTATGGAGTT</td>
</tr>
<tr>
<td>Tdrd6_Seq2</td>
<td>TTAGTTTAGGTTTAGTAGTG</td>
</tr>
<tr>
<td>Tdrd12_F</td>
<td>ATGTTTTATTGAGGTTATGGTATATGG</td>
</tr>
<tr>
<td>Tdrd12_R_BIO</td>
<td>CAAAAACTCCCAAACCAATCTTTTA</td>
</tr>
<tr>
<td>Tdrd12_Seq1</td>
<td>GGGAATTGAGATGTT</td>
</tr>
<tr>
<td>Tdrd12_Seq2</td>
<td>AGGAGATTATAGTATAG</td>
</tr>
<tr>
<td>Gpr56_F</td>
<td>GTAAGTGTGTTGAGTATAGGATA</td>
</tr>
<tr>
<td>Gpr56_R_BIO</td>
<td>TCCACCTCCCTACTCCCTCATT</td>
</tr>
<tr>
<td>Gpr56_Seq</td>
<td>TTTTTTTATTTTTGTTGGATT</td>
</tr>
<tr>
<td>Tbx2_F</td>
<td>GGATGTTTTGGAGTAAGAAGTATA</td>
</tr>
<tr>
<td>Tbx2_R_BIO</td>
<td>AACCACAAACTTTAACATCCACTATT</td>
</tr>
<tr>
<td>Tbx2_Seq1</td>
<td>TGGATAAGAAAGTTAAATAT</td>
</tr>
<tr>
<td>Tbx2_Seq2</td>
<td>GGGAAGGTAAAAGGGAT</td>
</tr>
</tbody>
</table>
**Ex vivo sperm incubations**

Prior to genomic DNA extraction, sperm were incubated for 4 hours at 37°C in the following conditions: Donners complete media (Hisano et al., 2013), Donners complete media supplemented with either 200 μM SAM, additional sodium pyruvate (10 mM), additional sodium DL-Lactate (1% vol/vol), NaHCO₃ (200mM) as well as Donners complete buffered to pH 6.5, 7, and 7.4. Additional conditions include: Donners basic (stock solution) (Hisano et al., 2013), Donners basic supplemented with NaHCO₃ (final 25 mM), Donners basic supplemented with BSA (final 20 mg/mL), Donners basic supplemented with Sodium DL-Lactate (final 0.53% vol/vol), and finally Donners basic buffered to pH 6.5, 7, and 7.4.

DNase treatment of caput sperm was performed prior to gDNA isolation as per manufacturer’s instructions (Qiagen79254). After DNase treatment, sperm were spun for 30 sec to (14000 rpm), supernatant aspirated, and sperm samples processed for genomic DNA extraction (see: *Genomic DNA isolation*).

**Visualization of cell-free DNA and immunostaining**

After sperm prep, samples were dried onto VWR Superfrost Plus Micro Slides (48311–703), fixed for 10 minutes in 4% paraformaldehyde solution, washed 3x with PBS, and permeabilized for 10 min with 0.1% triton in PBS. Blocking was performed with 10% BSA in PBS for 1 hour at room temperature in a humidified chamber. Staining with primary antibody, Anti-Histone H3 (citrulline R2 + R8 + R17 (Abcam ab5103), was performed for 2 hours at room temperature in 1% BSA PBS at a concentration of 1:250
in a humidified chamber. Primary antibody was decanted and slides washed three times with PBS for 5 minutes each. Goat anti-Rabbit 488 secondary antibody (Invitrogen A11008) was performed for 1 hour at room temperature at a concentration of 1:500 in a humidified chamber. Secondary antibody was decanted and slides washed three times with PBS for 5 minutes each followed by VECTASHIELD Antifade Mounting Medium with DAPI (Vector Labs H-1200), sealing coverslip, and imaging on the Zeiss Axioskop 2 Plus. All studies were repeated with at least two distinct sperm samples. Exposures for different sperm preps were captured using identical light intensity and exposure times, set to enable visualization of cell-free DNA in DAPI images, and set just below saturation for cit-H3 images. Note that secondary alone controls for cit-H3 were clean (not shown).

**Histology**

Virgin or retired breeders of approximately the same age were anesthetized and perfused with phosphate-buffered saline (PBS) followed by 4% paraformaldehyde (PFA)/PBS. Epididymides were explanted and further incubated in 4%PFA/PBS at 4°C overnight. After washing the excess of PFA with PBS, the sample was incubated at 4°C in 30% sucrose, 0.002% sodium azide in PBS until organ sank to the bottom of the tube. The sucrose solution was replaced and once organ remained at the bottom of the tube the same volume of optimal cutting temperature compound (OCT) was added to the vial and kept ON at 4C under agitation. Samples were then mounted in OCT and frozen at −80°C until sectioning. Sectioning was done at a thickness of 5 μm by the UMASS morphology core. Slides were stored frozen at -20°C.
Slides were placed at a 37°C warm plate for 10 minutes to ensure proper attachment of 
the section to the slide, then washed three times for 5 minutes in PBS 0.02% tween 20 
(PBS-T) to remove OCT, followed by DAPI stain (3 μM 4’,6-diamidino-2-phenylindole 
in PBS) for 5 min. After washing off the excess of DAPI with PBS-T, slides were 
mounted with ProLong gold Antifade (Thermofisher P36930) and imaged the following 
day.

**Small RNA sequencing**

Males from the same litter were split into the following groups: not mated, mated and 
recovered for three weeks, and mated and recovered for six weeks. Mating was 
confirmed by the formation of blastocysts in culture. All males were dissected at the 
same age (14 weeks). Caput and cauda sperm were isolated as described and small RNA 
sequencing was performed. Isolation of 18–40 nts small RNAs was carried out as 
previously described (Sharma et al., 2018). Size selection by purification of RNAs from 
15% polyacrylamide-7M urea denaturing gels and sequencing library preparation using 
Illumina’s TruSeq Small RNA Library Preparation Kit.
CHAPTER III.

SPERM SMALL RNAs THREE WAYS
3.1 ABSTRACT

Small RNAs resulting as cleavage products from mature tRNAs - referred to as tRNA fragments (tRFs) - are an emerging class of small non-coding RNAs (ncRNAs) implicated in many regulatory processes. Our interest lies with tRFs carried in mouse sperm with the potential to be altered in response to paternal environment. However, we know there is a strong bias in capturing tRFs using conventional sequencing methods when compared to Northern blots. Specifically, we have only been successful in capturing 5′ fragments of a handful of tRNA species. This limitation is due in part to structure, RNA termini, and numerous modifications present on tRNAs – all of which interfere with sequencing preparation protocols. To address this discrepancy, we set out to compare the two most widely used conventional small RNA sequencing methods (Illumina and NEB), and a novel protocol developed by Kathleen Collins’ Lab at UC Berkeley - Ordered Two-Template Relay (OTTR). OTTR utilizes a unique reverse transcriptase (RT) capable of adding both adaptors and performing cDNA synthesis in a single step. OTTR confirms abundant 5′ and 3′ fragments of GlyGCC and ValCAC matching previous Northern blot data as compared to only 5′ fragments from Illumina and NEB. These results not only reshape our understanding of the sperm RNA payload of mature sperm as being far more complex than previously thought, but also force us to reassess hypotheses regarding the RNAs payload contributing to future offspring.
3.2 INTRODUCTION

Beyond the primary role of tRNAs in protein synthesis as the physical link between amino acid and mRNA, tRNA fragments are a growing area of research due to their growing documented regulatory functions (Yu et al., 2021; Park et al., 2020; Kim et al., 2020; Su et al., 2020; Keam, 2015; Anderson and Ivanov, 2014). Mature tRNAs are susceptible to targeted cleavage by RNases – for example angiogenin (ANG) - also known as RNase5 - produces 5’ and 3’ tRNA halves (Fu et al., 2009; Yamasaki et al., 2009). ANG is a stress-activated ribonuclease responsive to numerous cellular stress conditions including hypertonic stress and arsenite – an oxidative stressor (Liu et al., 2018; Li and Hu, 2012; Saikia et al., 2012). This cleavage can have downstream implication like translation initiation inhibition (Ivanov et al., 2011) and even function to protect cells from apoptosis during osmotic stress (Saika et al., 2014). Notably, ANG is directly involved in tumor growth – increase of ANG expression and secretion is upregulated in many cancer types (Shimoyama et al., 1996). This, due to angiogenin’s ability to stimulate blood vessel formation (ie ANGiogenesis), - critical for tumors vascularization in a hypoxic microenvironment. (Hartmann et al., 1999). It is then not entirely surprising to learn that abnormal tRF expression have been reported in numerous cancer types (Yu et al., 2020; Zhu et al., 2020). To date, the only known cleavage target of ANG in vivo are tRNAs (Lyons et al., 2017) with ANG overexpression shown to selectively cleave a specific subset of tRNAs (Su et al., 2019).

Our lab is interested in how these tRNA fragments can be modulated by paternal environment. tRFs have also been widely documented to change in response to paternal
diet in the male germline – specifically in mature sperm (Sharma et al., 2016; Chen et al., 2016A). These diet-responsive tRNA-derived fragments in paternal high fat diet include: increased expression of specific 5’ tRFs (tRF-GluCTC and tRF-GluTTC) in rats (de Castro Barbosa et al., 2016) and increased levels in tRF-GlyGCC accompanied by decreased tRF-GluCTC in mice (Cropley et al., 2016) for example. Our lab has focused on the effects of low protein diet and found increased levels of 5’ tRF-GlyGCC in mature sperm which can function to repress MERVL (LTR retroelement) in embryos and mESCs (Sharma et al., 2016). This suppression of MERVL was further found to occur through altered histone gene expression and production (Boskovic et al., 2020).

Given our interest in how paternal environment – namely diet – has the potential to impact offspring phenotype through RNAs delivered in sperm, it is critical to state that we know current strategies for sequencing tRNAs and tRFs are biased. Sperm RNAs are unique in that the majority of sperm RNAs fall below ~200nt in length (Figure 3.1). This requires specialized sequencing protocols as the standard RNA sequencing methods will not work for small RNAs. Currently in the field, the majority of small RNA sequencing is performed with two primary methods: either the TruSeq Small RNA Sample Preparation Kit (Illumina) (Sarker et al., 2019; Wang et al., 2021B; Stanford et al., 2018; Sharma et al., 2016) or the NEBNext Small RNA Library Prep Set (NEB) (Chan et al., 2020; Suvorov 2020; de Castro Barbosa et al., 2016; Cropley et al., 2016).
Figure 3.1 Sperm RNAs have a distinct size distribution as compared to the tissue from which it is generated
A) 15% UREA PAGE GEL with 1ug each: testicular, caput, and cauda tissue as compared to mature cauda sperm B) Representative bioanalyzer of testicular tissue RNA where tissue has distinct rRNA peaks and C) sperm RNA where the majority of RNAs falls below ~200nt.
How do we know current sequencing methods are limited in capturing reality? Our lab and others have largely characterized 5′ fragments of tRF-GlyGCC, tRF-GluCTC, ValCAC, and tRF-HisGTG as the predominant species in mature sperm, but we know from Northern blots that the 3′ tRNA fragments of each of these species are also present - and at nearly equal levels (Sharma 2018 et al., Zhang et al., 2018; Sharma et al., 2016). Of course, due to limitations in starting material (sperm RNA is present in low concentrations and Northern blotting requires micrograms), we cannot perform Northern blotting for every tRNA species.

To address this discrepancy, we set out to compare the two most widely used conventional small RNA sequencing methods (Illumina and NEB) with a novel protocol developed by Kathleen Collins’ Lab at UC Berkeley - Ordered Two-Template Relay (OTTR) (Upton et al., 2021).

**Difficulties in cloning tRFs: modifications**

Characterization of tRNAs and tRFs using standard sequencing methods has been hindered by several technical difficulties. Not only do tRNAs have an incredibly compact structure which can interfere in both adapter ligation and cDNA synthesis, but are also heavily modified. Over 150 RNA modifications have been identified in all domains of life (Boccaletto et al., 2017) and 80% were identified in tRNAs (Suziki et al., 2021 Levi and Arava, 2021 and Frye et al., 2016). And tRNAs, with their cloverleaf-like secondary structure, are heavily modified – between 10-20% of any given tRNA is modified. (Jühling et al., 2009).
The most abundant of the modifications include pseudouridine ($\Psi$), Inosine (I), Queuosine (Q) (Figure 3.2), N6-methyladenine ($m^6A$), 5-methylcytosine ($m^5C$), N1-methylguanosine ($m^1G$), N1-methyladenosine ($m^1A$), and N3-methylcytosine ($m^3C$) and much more complicated and bulky N6-threonylcarbamoyladenosine ($t^6A$) and 5-methoxycarbonylmethyl-2-thiouridine (mcm5s2U) (Boccaletto et al., 2017).
Figure 3.2 Examples of RNA modifications present of tRNAs
MODOMICS-generated figure (Boccaletto et al., 2017) illustrating A) basic tRNA structure with percent modification at given tRNA position and B) Queuosine (Q), Pseudouridine (Ψ), Inosine (I) modifications and their respective structures.
Functionally, these modifications play critical roles in everything from gene regulation (Roundtree et al., 2017) to directing translational control. In the case of pseudouridylation (Ψ) (Guzzi et al., 2018) to provide stability and protection against RNA cleavage (Lyons, et al., 2018). With respect to protection against tRNA cleavage, specific modifications have been shown as protective – for example, queuosine (Q) (Wang et al., 2018). Even site-specific modifications like methylation by Dnmt2 at position 38 (m^5C38) (Schaefer et al., 2010) and methylation of position 58 (m^1A) have been found as critical for tRNA stability. (Zhang and Jia, 2018).

Notably, queuosine (Q) (usually occurring in tRNAs at position 34) is a nutritionally-modulated tRNA modification. That is, mammals cannot synthesize queuosine or any of its precursor forms – it must be obtained from diet and the microbiome (Fergus et al., 2015). Remarkably, germ-free mice fed a Q-free diet lost queuosine in tRNAs (Reyniers et al, 1981). Even more remarkable, Q-tRNA levels can impact other modifications - germ-free mice with reduced levels of dietary Q-tRNA had lower levels of m^5C38-tRNA methylation - dietary supplementation with Q was shown to reverse these effects (Tuorto et al., 2018). This is of particular interest to us as it provides a direct link between diet (microbiome) and the potential for tRNA cleavage in sperm through modulation of modification levels.

Modifications can interfere with standard cloning protocols and commonly used reverse transcriptases (RTs) in two ways: 1) “hard-stop” modifications and 2) RT fidelity through misincorporation. “Hard-stop” modifications like N1-methyladenosine (m^1A), N1-methylguanosine (m^1G), and N2,N2-dimethylguanosine (m^2G) can all block RT
extension altogether (Wilusz, 2015). And specific modifications – namely 5-hydroxymethyluridine and N6-methyladenine (m^6A) – have been shown to increase RT error rate leaving different substitution signatures (Potapov et al., 201; Ebhardt et al., 2009). These modifications present a serious hindrance to contend with as any given tRNA has an average of ~13 modifications (Pan, 2018).

**RNA termini: The motivation for T4 Polynucleotide Kinase (T4 PNK) treatment**

Beyond being heavily modified, small RNAs can have a variety of prime (5′) and three prime (3′) termini (ends). These come in a variety of forms: 5′phosphate (5′ P), 5′-triphosphate (5′ PPP), 5′ hydroxyl (5′ OH), 3′-phosphate (3′ P), 3′ hydroxyl (3′ OH), and 2′,3′-cyclic phosphate (3′-cP) to name a few. Successful adapter ligation for small RNAs is dependent on a 5′ P and a 3′ OH - other ends decrease ligation and downstream cloning efficiency. Effectively, RNAs containing a 2′,3′ cyclic phosphate (3′-cP) escape 3′ adapter ligation and remain “uncaptured” (Wang et al., 2021A; Honda et al., 2016). This is especially problematic for tRNAs as for example, ANG cleavage leaves behind a 5′ hydroxyl group (5′ OH) and a 2′,3′ cyclic phosphate (3′-cP) – in fact many ribonucleases have been reported to generate a 3′-cP 3′ ends (Shigematsu et al., 2018).

To eliminate diversity of RNA termini and increase adapter ligation efficiency, we employed T4 polynucleotide kinase (T4 PNK) treatment (alongside a control – “no treatment” group). T4 PNK functions to phosphorylate 5′ hydroxyl (5′ OH) ends to generate 5′ P as well as to remove of phosphoryl groups (3′ P) and 2′,3′-cyclic phosphate (3′-cP) termini to generate 3′ OH ends. (Figure 3.3) This treatment is intended to increase
our yield of RNA containing the “ideal” substrate requirements (5’P and 3’OH) for successful adapter ligation (Das and Shuman 2013; Amitsur et al., 1987; Cameron and Uhlenbeck, 1977; Richardson, 1965)

Figure 3.3 T4 Polynucleotide Kinase (T4 PNK)
T4 Polynucleotide Kinase (T4 PNK) functions to phosphorylate 5’ hydroxyl (5’ OH), remove phosphoryl groups (3’ P), and 2’,3’-cyclic phosphate (3’-cP) termini to generate 3’ OH ends for subsequent ligation.
Why OTTR? Limitations of commercially available kits

Illumina’s TruSeq Small RNA Library Preparation Kit and NEB’s NEBNext Small RNA kit both follow the two-adaptor ligation approach – first, direct ligation of a preadenylated adaptor to the 3’ end of the RNA succeeded by ligation of an adaptor to the 5’ end. This introduces our first bias – successful adapter ligation is dependent on 5’ P and 3’ OH which leads to enrichment of transcripts with – not surprisingly - 5’ P and 3’ OH ends. This bias creates selection and preference for specific small RNA classes – largely miRNAs and siRNAs (Carthew and Sontheimer, 2010). Additional bias can be introduced in the ligation steps due to the inherently complex tRNA secondary structure and inaccessible ends – simply put, RNA ligases prefer accessible ends (Fuchs et al., 2015; Zhuang et al., 2012).

NEB’s NEBNext Small RNA kit has improved on Illumina’s protocol by supplementing polyethylene glycol (PEG) into the ligation reaction to increase ligation efficiency as well as reduce bias. PEG functions as a macromolecular crowding agent effectively increasing the concentration of the ligase and the RNA substrate (Heyer et al., 2015; Viollet et al., 2011; Harrison and Zimmerman, 1984).

Lastly, RNA modifications can interfere with the reverse transcriptase (RT) used for cDNA synthesis – Illumina uses SuperScript II and NEB *ProtoScript* II, respectively. As mentioned earlier, specific modifications can lead to RT misincorporation or, more likely, cause the RT to fall off altogether when it reaches a “hard-stop” modification.
Enter: OTTR

Ordered Two-Template Relay (OTTR), a novel protocol developed by Kathleen Collins’ Lab at UC Berkeley relies on a novel, modified, reverse transcriptase (RT) from *Bombyx mori*. This non-long terminal repeat (non-LTR) retroelement-encoding reverse transcriptase (RT) has unique properties allowing for end-to-end RNA sequence capture. By taking advantage of both the terminal transferase activity and the processive template jumping of the RT, the Collins Lab developed a remarkable protocol that produces one cDNA from discontinuous templates copied end-to-end in a single tube. Instead of sequential ligation of adaptors followed by cDNA synthesis (Illumina and NEB), OTTR uses cDNA synthesis to add both adaptors and produce the following: the 5′ adaptor primer, cDNA to the input RNA strand of interest, and the 3′ adaptor primer. (Upton et al., 2021).
3.3 RESULTS

As the OTTR protocol is completely unlike both the Illumina and NEB small RNA protocols it is critical to minimize bias between library builds to allow for the best possible comparison. To do this, after size selection, RNA was pooled and split into a “no treatment” control group and a “T4 PNK + ATP” group before further splitting into OTTR, Illumina, and NEB builds. Essentially, the libraries were built from the same starting pools to minimize mouse to mouse variation or minor technical differences in size selection. Additionally, in previous experiments (Sharma et al., 2016, Chen et al., 2016A) we (and others) have gel purified RNAs 18-40 nt in length for library preparation. In keeping with recommendations from the Collins lab, we used column-based size selection (mirVana™). This method of size selection and purification captured everything below ~75nt in length (used for all library builds).

We sequenced technical replicates (5-10 million reads per condition), removed adaptors, performed quality filtering, and sequential mapping alignment using our standard sequencing pipeline to known rRNAs, miRNAs, tRNAs, piRNAs, and snRNAs. Any unmapped reads are then mapped to the standard, annotated transcriptome (Yukselen et al., 2020; Sharma et al., 2018; Sharma et al., 2016).

**OTTR captures 3’ tRNA fragments**

The first (and most pressing) question we set out to address: Is OTTR capable of capturing 3’ tRNA fragments that have evaded capture thus far? To start, we focus only on tRFs GlyGCC and ValCAC – these are the tRFs we have Northern blot validation for.
showing equal distributions of the 5’ and 3’ tRFs (Sharma et al., 2016; Sharma 2018 et al., Zhang et al., 2018). Comparing tRF-GlyGCC (Figure 3.4) from all three library builds shows us equal distributions of the 5’ and 3’ tRNA fragments with OTTR (Figure 3.4A) and only the 5’ ends with both the Illumina (Figure 3.4B) and NEB (Figure 3.4C) library builds (as expected and previously published). This holds true for ValCAC (Figure 3.5) with nearly equal fractions 5’ and 3’ tRNA fragments with OTTR (Figure 3.5A). This alone convinced us that OTTR is superior as it is the first method that has reflected Northern blot results.

**OTTR is unaffected by T4 PNK treatment**

What about after T4 PNK treatment? Do we resolve additional fragments?

In all library builds – for tRFs GlyGCC (Figure 3.6) and ValCAC (Figure 3.7) – T4 PNK treatment does little to resolve 3’ tRNA fragments. There is a very subtle increase in 3’-tRF-GlyGCC with Illumina + T4 PNK treatment (Figure 3.6B), but may even be detrimental to 3’ tRF-GlyGCC with OTTR + T4 PNK treatment (Figure 3.6A). We suspect this is likely due to 3’-tRF RNA ends not being capable of conversion to a 5’ P and 3’ OH with T4 PNK alone and therefore cannot be ligated. In the case of tRF-ValCAC, we do see one noticeable difference after T4 PNK treatment with the NEB build (Figure 3.7C) – a large increase of reads ending at position 23 that is virtually absent without T4 PNK treatment. This likely indicates that a ValCAC cleavage product is generated, but only accessible after transitioning RNA termini to 5’ P and 3’ OH.
Figure 3.4 Reads mapping to tRF-GlyGCC
With A) OTTR B) IT C) NEB library builds
Figure 3.5 Reads mapping to tRF-ValCAC
With A) OTTR B) IT C) NEB library builds
Figure 3.6 Reads mapping to tRF-GlyGCC after T4 PNK Treatment
With A) OTTR B) IT C) NEB library builds
Figure 3.7 Reads mapping to tRF-ValCAC after T4 PNK Treatment
With A) OTTR B) IT C) NEB library builds
OTTR provides overwhelming diversity in tRF species captured

Now that we know OTTR is capable of effectively capturing 3’ ends providing us with both 5’ and 3’ tRFs GlyGCC and ValCAC, which we know from Northern blot analysis to be “the truth”, we set out to look at overall tRF abundance. The most abundant tRFs in sperm with Illumina have been: GlyGCC, GluCTC, HisGTG, and ValCAC. But are these truly the most abundant or are these just the tRF species that are most clonable with Illumina’s inherent bias? When we look at all tRNA species, OTTR reveals an unprecedented diversity (Figure 3.8A) when compared to either Illumina (Figure 3.8B) or NEB (Figure 3.8C)

Does OTTR successfully capture 5’ and 3’ sides of all the tRFs?

Furthermore, when we assess the coverage across any given tRNA, OTTR provides us with 5’ and 3’ tRFs for all tRNAs as well as some internal tRNA fragments (Figure 3.9A) while Illumina (Figure 3.9B) or NEB (Figure 3.9C) clearly pile up on the left side of the graph signifying almost all reads mapping to the 5’ end. Generating tRNA coverage across all tRNAs – not just GlyGCC and ValCAC - further illustrates how OTTR is superior in capturing of both tRNA ends.
Figure 3.8 Barplots illustrating diversity of tRNA-mapped reads by species
With A) OTTR B) IT C) NEB library builds
Figure 3.9 tRNA Coverage plots illustrating alignment position of tRNA by species. With A) OTTR B) IT C) NEB library builds where each line is one tRNA mapped read.
T4 PNK treatment slightly increases diversity in tRF species captured with Illumina and NEB, but not OTTR

T4 PNK – the enzymatic treatment required to convert to 5’ P and 3’ OH – does not appear to impact OTTR in terms of the diversity of tRNAs (Figure 3.10– or read coverage (Figure 3.13). This finding suggests that OTTR does not necessarily have preference for a 5’ P and 3’ OH ends. Although this needs to be further explored with spike in control oligos with distinct ends, if proven, would not only eliminate the need for enzymatic treatment altogether, but also any unintended bias that may arise from enzymatic treatment. T4 PNK treatment does appear to improve Illumina and NEB slightly in terms of 1) diversity of tRNAs captured Illumina (Figure 3.11 and Figure 3.12) as well as 2) picking up slightly more 3’ tRFs (Figure 3.14 and Figure 3.15).
Figure 3.10 Barplot illustrating diversity of tRNA-mapped reads by species for OTTR
Where A) no enzymatic treatment and B) T4 PNK + ATP treatment of RNA prior to library preparation.
Figure 3.11 Barplot illustrating diversity of tRNA-mapped reads by species for Illumina TruSeq
Where A) no enzymatic treatment and B) T4 PNK + ATP treatment of RNA prior to library preparation.
Figure 3.12 Barplot illustrating diversity of tRNA-mapped reads by species for NEBNext
Where A) no enzymatic treatment and B) T4 PNK + ATP treatment of RNA prior to library preparation.
Figure 3.13 Coverage plot for OTTR protocol by tRNA
Comparing OTTR tRNA coverage with A) no enzymatic treatment and B) T4 PNK + ATP treatment of RNA prior to library preparation.
Figure 3.14 Coverage plot for Illumina Truseq by tRNA
Comparing Illumina Truseq tRNA coverage with A) no enzymatic treatment and B) T4 PNK + ATP treatment of RNA prior to library preparation.
Figure 3.15 Coverage plot for NEBNext by tRNA
Comparing NEBNext tRNA coverage with A) no enzymatic treatment and B) T4 PNK + ATP treatment of RNA prior to library preparation.
Each tRF species captured needs to be assessed individually

Beyond the tRFs we have Northern blot validation for, we now take an exploratory look at tRF-LeuCAG (Figure 3.16) where three library builds yield three entirely different results. In typical OTTR fashion, both 5’ and 3’ reads can be found spanning LeuCAG – with and without T4 PNK treatment. Upon careful inspection, OTTR also has a small subset of reads ending in the anticodon loop (positions 34 and 35) as well as reads starting around position ~36 (Figures 3.16A and 3.16D). Illumina is very capable of capturing 5’ ends tRF-LeuCAG (Figures 3.16B and 3.16E) with some reads ending in the anticodon loop (positions 34 and 35) as well, but the majority of reads ending at position 26 – a site known to be RNA modified to N2,N2-dimethylguanosine (m²G26) (Edqvist et al., 1995). And while NEB captures very few 5’ ends of tRF-LeuCAG, the majority of the reads fall between positions ~50-65- consistent with the T loop of tRNAs (Figures 3.16C and 3.16F) (see Figure 3.2 for relative T loop location). This difference between Illumina and NEB is somewhat unexcepted given that they are effectively the same protocol - with the addition of PEG in the NEB build.
Figure 3.16 Reads mapping to LeuCAG - with and without T4 PNK Treatment
Comparing LeuCAG mapped reads across three builds A,D) OTTR, B,E) Illumina Truseq, and C,F) NEBNext as well as without T4 PNK Treatment (A,B,C) and with T4 PNK Treatment (D,E,F).
In another example, in the case of tRF-AspGTC, OTTR again has both 5’ and 3’ reads as we have come to expect, but interestingly OTTR also has a subset of reads starting in the tRNA D loops and stem around position ~20 (Figures 3.17A and 3.17D), which is found in larger quantities in both Illumina and NEB, only after T4 PNK treatment (Figures 3.17E and 3.17F). These reads starting around position 20 are entirely absent from the Illumina and NEB controls (Figures 3.17B and 3.17C) likely revealing an example of where endonuclease cleavage (ANG or other) resulted in a 5’ OH which T4 PNK treatment successfully converted to a 5’ P.
Figure 3.17 Reads mapping to AspGTC - with and without T4 PNK Treatment
Comparing AspGTC mapped reads across three builds A,D) OTTR, B,E) Illumina Truseq, and C,F) NEBNext as well as without T4 PNK Treatment (A,B,C) and with T4 PNK Treatment (D,E,F).
Distribution of read length mapping to tRNAs varies with OTTR

Another notable difference in the OTTR-mapped tRNA reads as compared to both Illumina and Truseq is the distribution of read lengths. Until now, we have previously only gel extracted RNAs 18-40 nt in length and thus expected a tight distribution in this range. This is the first time we are looking at all tRFs below ~75nt.

The majority of tRNA-mapped reads with OTTR (Figure 3.18A) are below 40nt - this in contrast to both Illumina (Figures 3.18C) and NEB ((Figures 3.18E) which have narrow windows of read length ranging from ~29nt to 34 nt. Based on length alone, if we only had the Illumina (Figures 3.18C) and NEB (Figures 3.18E) data, we would imagine all tRNAs to be cleaved along the anticodon loop generating tRNA halves. Based on the OTTR data, another interpretation is that yes, we do indeed have tRNA halves, but also internal tRNA fragments that could have been generated from cleavage all along the tRNA – not just limited cleavage at the anticodon loop. We see a hint of this with Illumina and NEB after T4 PNK treatment (Figures 3.18D and 3.18E) with a notable, but not dramatic, increase in smaller fragments.
Figure 3.18 Read length of insert mapped to tRNAs
Overall distribution of read length varies with OTTR

Until now, we have focused heavily on tRFs even though we know there are many other small RNAs and RNA fragments found in sperm. Now, we step back and turn our attention to look at all RNAs cloned. As mentioned, this is the first time we are looking at all RNAs below ~75nt. In looking at read length distribution (Figure 3.19) we will see substantial bias between the library builds. OTTR is quite capable at producing reads below ~40nt in length with the majority falling below ~26nt in length. As has become apparent, OTTR is indifferent to T4 PNK treatment and thus with or without, the read distribution looks virtually identical (Figures 3.19A and B). This in contrast to both Illumina (Figures 3.19C) and NEB (Figures 3.19E) where without T4 PNK treatment the bulk of reads fall into a narrow window ranging from ~29 nt to 34 nt reads that we know map to tRNAs. After T4 PNK treatment, we see shift in the distribution of reads with Illumina (Figure 3.19D) which ultimately correlates to a large increase of reads mapping to rRNA (Table 3.1). This is also the case for NEB (Figures 3.19E) after T4 PNK treatment (Table 3.1).
Figure 3.19 Read length of insert after adaptor removal, but prior to mapping
T4 PNK treatment functions to increase 18S and 28S rRNA reads captured with Illumina and NEB

With respect to this marked increase in rRNA mapping reads after T4 PNK treatment with Illumina and NEB – as well as a slight increase in OTTR (Table 3.1), we set out to resolve which rRNAs are responsive to T4 PNK treatment. Of total reads mapped to rRNA, about half mapped to 45S rRNA – the precursor for 18S, 5.8S and 28S rRNA (Henras, 2014). After T4 PNK treatment in Illumina and NEB libraries, 18S and 28S rRNA are increased (Figure 3.20) suggesting the 5’ and 3’ ends of the rRNAs are generated in a mechanism that allows them to be susceptible to end treatment. This finding holds true with recently published work addressing the impact of T4 PNK treatment on small RNA sequencing in HEK293T cells with both Illumina and NEB (Wang et al., 2021A).
Figure 3.20 Relative rRNA ratios mapping in each condition and library build
Table 3.1 Sequential mapping to determine percent of reads mapping

<table>
<thead>
<tr>
<th>Sample Treatment</th>
<th>Percent of reads mapping to:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>rRNA</td>
<td>miRNA</td>
</tr>
<tr>
<td>OTTR No Treatment</td>
<td>25.17%</td>
<td>0.87%</td>
</tr>
<tr>
<td>OTTR T4 PNK + ATP</td>
<td>40.30%</td>
<td>0.61%</td>
</tr>
<tr>
<td>Illumina Truseq No Treatment</td>
<td>29.88%</td>
<td>2.80%</td>
</tr>
<tr>
<td>Illumina Truseq T4 PNK + ATP</td>
<td>74.71%</td>
<td>0.13%</td>
</tr>
<tr>
<td>NEBNext No Treatment</td>
<td>22.46%</td>
<td>1.35%</td>
</tr>
<tr>
<td>NEBNext T4 PNK + ATP</td>
<td>70.05%</td>
<td>0.13%</td>
</tr>
</tbody>
</table>

Sequential mapping to rRNA, miRNA, tRNA, piRNA, snRNA, repetitive elements.
The majority of OTTR reads captured remain “unaligned” to the genome

To assess mapping rates for small RNAs other than rRNAs, we removed rRNA reads from our analysis and focus only on the following mapping rates: miRNA, tRNA, piRNA, snRNA, repetitive elements, transcriptomic reads, and remaining unaligned reads (Table 3.2). Most alarmingly, OTTR has >70% of reads remaining as “unaligned”. We suspect – with maximal mismatch allowance – that the majority of these reads will actually map to tRNAs, but we have yet to perform these analyses systematically. If, on average, any given tRNA can have 13 modifications, we can assume allowing 1, 2, or even 3 mismatches for mapping will not be sufficient to capture all potential misincorporations thus leaving the majority of tRNA fragments classified as “unaligned”. We also see a sharp increase in unaligned reads with Illumina and NEB after T4 PNK treatment suggesting that perhaps ligation of these cleaved tRNA fragments is successful, but downstream misincorporations yield unalignable reads. This is a computational problem that certainly needs further investigation.
<table>
<thead>
<tr>
<th>Sample Treatment</th>
<th>miRNA</th>
<th>tRNA</th>
<th>piRNA</th>
<th>snRNA</th>
<th>Repetitive elements (rmsk)</th>
<th>Reads Remaining After Sequential Mapping</th>
<th>Transcriptome Aligned Reads</th>
<th>Remaining Unaligned Reads</th>
</tr>
</thead>
<tbody>
<tr>
<td>OTTR</td>
<td>1.16%</td>
<td>4.77%</td>
<td>8.71%</td>
<td>0.66%</td>
<td>2.14%</td>
<td>82.56%</td>
<td>9.21%</td>
<td>73.34%</td>
</tr>
<tr>
<td>OTTR T4 PNK + ATP</td>
<td>1.03%</td>
<td>3.81%</td>
<td>8.02%</td>
<td>0.84%</td>
<td>1.55%</td>
<td>84.76%</td>
<td>12.46%</td>
<td>72.30%</td>
</tr>
<tr>
<td>Illumina Truseq</td>
<td>3.99%</td>
<td>48.58%</td>
<td>4.27%</td>
<td>0.68%</td>
<td>0.11%</td>
<td>42.38%</td>
<td>11.52%</td>
<td>30.86%</td>
</tr>
<tr>
<td>Illumina Truseq T4 PNK + ATP</td>
<td>0.53%</td>
<td>4.06%</td>
<td>4.54%</td>
<td>1.96%</td>
<td>0.36%</td>
<td>88.55%</td>
<td>30.84%</td>
<td>57.71%</td>
</tr>
<tr>
<td>NEBNext</td>
<td>1.74%</td>
<td>55.25%</td>
<td>1.86%</td>
<td>0.88%</td>
<td>0.05%</td>
<td>40.22%</td>
<td>14.19%</td>
<td>26.03%</td>
</tr>
<tr>
<td>NEBNext T4 PNK + ATP</td>
<td>0.43%</td>
<td>5.67%</td>
<td>3.74%</td>
<td>10.88%</td>
<td>0.27%</td>
<td>79.01%</td>
<td>25.58%</td>
<td>53.43%</td>
</tr>
</tbody>
</table>

Table 3.2 Sequential mapping to determine percent of reads mapping after rRNA removal
Sequential mapping to miRNA, tRNA, piRNA, snRNA, repetitive elements.
3.4 DISCUSSION

Here, we set out to compare Illumina and NEB small RNA protocols to the novel OTTR protocol as a means to 1) establish the “ground truth” of tRF content in sperm 2) assess OTTR in overcoming technical difficulties and limitations due to tRNA structure, tRF termini, and numerous modifications present on tRNAs. We compare a “no treatment” control group and a “T4 PNK enzymatic treatment group (to minimize variation in RNA termini) for all three builds: OTTR, Illumina, and NEB.

First, can OTTR establish the “ground truth” of tRF content in sperm? Is OTTR superior at capturing 3′ tRFs? For the two tRF species we have Northern blot validation for, the answer is a resounding “YES!” That said, moving forward it would be wise to validate OTTR using at least several additional probes – beyond GlyGCC and ValCAC – using the gold standard: Northern blotting.

Additionally, we can gather relative abundance from Northern blots, but not precise counts like we can from sequencing data. One current limitation of the OTTR protocol is that we do not know if we can trust the precise abundance of these tRNA fragments. To address this, it will be critical to combine the OTTR protocol with unique molecular identifiers (UMIs) – something the lab is already investing in. This will be incredibly valuable and useful for looking at sperm generated from dietary paradigms as well as other paternal effects. And if our interest lies specifically in tRFs moving forward, rRNA depletion of our sperm samples would enable us to gain more information as we currently lose ~50% of reads to rRNA – even more after T4 PNK treatment.
Second, how does OTTR handle the technical difficulties of cloning small RNAs with various termini where other library preparation methods have failed? Given that the OTTR “no treatment” control group and a “T4 PNK + ATP” group looked identical in nearly all cases investigated, we can be fairly confident that OTTR is unaffected by RNA termini. This alone simplifies the protocol as it eliminates the need for enzymatic treatment. In some cases – for example AspGTC (Figure 3.17) OTTR shows a less biased version of what you would get if you averaged together the control group and T4 PNK + ATP treatment groups of either Illumina or NEB.

Although we suspect OTTR to be unaffected by enzymatic treatment targeting RNA ends, we can only say this with confidence for T4 PNK. T4 PNK treatment will address 5’ hydroxyl (5’ OH), 3’ phosphate groups (3’ P), and 2’,3’-cyclic phosphate (3’-cP), but RNA termini can be much more varied. To address this, further experiments using additional enzymatic treatments are necessary. One example worth trying: Tobacco Acid Pyrophosphatase (TAP) and RNA 5’ pyrophosphohydrolase (RppH) can be used to convert 5’ triphosphate groups (5’PPP) – another group know to decrease cloning efficiency for small RNA sequencing – to 5’ phosphate groups (5’ P) (Almeida et al., 2019).

Of course, the “real truth” is: everything has a bias. With both Illumina and NEB, we are confident in knowing where our bias is occurring. With OTTR, we are entering an unknown territory of bias - we still don’t know the full limitations of this protocol. One item that does stand out is OTTR’s preference to clone shorter fragments – those under ~40 nt and most predominantly, fragments under 26nt – in comparison to the
average insert length of either Illumina and NEB - this agrees with what the Collins lab described (Upton et al., 2021). Another standout circumstance of the OTTR protocol is the incredibly high rate of reads that are “unaligned”. Although we suspect these to be misincorporations needing a higher allowance of mismatches in mapping, this needs to be addressed. Moving forward, it will be critical to test this hypothesis computationally by allowing 1 to N mismatches and assessing mapping efficiency and specificity. Additionally, although costly, it could be useful to generate synthetically modified RNAs (with modified phosphoramidites) at specific sites to gain insight over how OTTR handles modifications.

Overall, we are thrilled that OTTR captures 3’ tRNA fragments that are invisible to all other protocols examined. Taken together, our data provide an updated view of the mouse sperm small RNA payload and highlight the utility of OTTR for analysis of tRNA fragments by deep sequencing.
3.5 METHODS

Mice

Unless otherwise specified, tissues were obtained from 12 week old male FVB/NJ

Dissection and cauda sperm purification

Ten FVB males, euthanized at 12 weeks of age according to IACUC protocol, were dissected for cauda epididymal sperm. Briefly, cauda epididymis was placed into Donners complete media (Hisano et al., 2013) and tissue was cleared of fat and connective tissue before incisions were made using a 26G needle while keeping the bulk tissue intact. Tissue was gently squeezed allowing sperm to escape into solution. After incubation at 37°C for 1 hour, sperm containing media was transferred to a fresh tube and collected by centrifugation at 5000rpm for 5 minutes followed by a 1X PBS wash. To eliminate somatic cell contamination, sperm were subjected to a 1mL 1% Triton X-100 incubation 37°C for 15 mins with 1500 rpm on Thermomixer and collected by centrifugation at 5000rpm for 5 minutes. Somatic cell lysis was followed by a 1x ddH2O wash and 30 second spin 14000 rpm to pellet sperm.

Sperm RNA purification and size selection

Immediately following cauda sperm purification, sperm RNAs were isolated using the mirVana™ miRNA Isolation Kit following the enrichment procedure for small RNAs as per manual. Protocol was modified with one half volume of 100% ethanol added to the aqueous phase recovered from organic extraction (recommended volume is one third).
**T4 Polynucleotide Kinase (PNK) treatment**

All column purified small RNAs were pooled and split into groups: T4 PNK treatment with ATP and a control no treatment group. T4 PNK treatment with ATP was incubated at 37°C for 30min in T4 PNK reaction buffer, 10mM ATP, and 50U T4 PNK (NEB M0201S). Control no treatment group was treated the same – minus the 50U T4 PNK. All sperm small RNAs samples were then cleaned and concentrated using RNA Clean & Concentrator™-5 (Zymo) prior to library preparation.

**Small RNA sequencing**

Small RNA sequencing library preparation was performed using three protocols: commercially available TruSeq Small RNA Library Preparation Kit (Illumina) and NEBNext® Small RNA Library Prep Set for Illumina® (NEB) as well as Collins Lab Library Preparation Kit (Upton et al., 2021). All three protocols were performed as per manual.

**Data analysis**

Small RNA sequencing was mapped to mm10 using Bowtie2. Abundance was determined by normalizing to parts per million based on total number of reads mapping to rRNA, microRNAs, tRNAs, piRNAs (repeatmasker and unique piRNAs), and mRNA (Refseq).(Yukselen et al., 2020). R was used to generate all tRNA coverage figures.
CHAPTER IV.

DISCUSSION
Overview

This thesis work set out to better understand diverse aspects of epididymal maturation and its impact on the sperm epigenome. The experiments in Chapter II started as exploration into a possible “third wave” of DNA methylation reprogramming during post-testicular sperm maturation. This work resulted in refuting the hypothesis of a “third wave” and definitively showing the sperm methylome is stable through post-testicular sperm maturation. In addition, the work in Chapter III sets out to remedy a known deficiency in the standard small RNA sequencing methods. Indeed, we expose these deficiencies with a newly developed sequencing method (OTTR) allowing us to capture specific tRNA fragments that evaded detection with sequencing previously. This work brings to light two specific aspects of post-testicular sperm maturation and broadens our overall understanding of sperm epigenomics.
In setting out to characterize and validate sperm cytosine methylation dynamics during epididymal maturation from seven germ cell populations using WGBS, we revealed consistent and stable methylation profiles throughout with exceptions only in caput and vas deferens sperm. In the process of validating what appeared to be a cytosine methylation-remodeling event we found, by chance, that the unique caput sperm methylome is lost following mating. That is, after mating and recovery, caput sperm never reverted back to the unmated methylation state. Following up on this observation led to the discovery of a surprising feature of the male reproductive tract: extracellular DNA is present in virgin males in the caput epididymis and the vas deferens, but not in mated males and contaminating extracellular DNA was associated with citrullinated histone H3. Most importantly, the DNA methylation status of virgin caput sperm can be resolved with DNase I treatment prior to sperm genomic DNA isolation providing a straightforward solution for future sperm methylation assays.

The unexpected finding of contaminating extracellular DNA in caput virgin males raised additional questions; was this unique to FVB mice? Was this a transient age-related phenomenon? To test whether this finding unique to FVB mice, we pyrosequenced epididymal sperm from another strain background (C57BL/6) in the presence and absence of DNase I and found similar results confirming that this unusual caput methylome is consistent in other strain backgrounds. To explore if the extracellular DNA in the caput of virgin males is age-related, we repeated our pyrosequencing analysis in virgin males from 10 weeks to 10 months of ages and found consistent methylation across all ages assayed.
Given that the cell-free DNA was associated with the citrullinated histone H3 (citH3) in the caput of virgin males and eager to follow up on the possibility of a neutrophil-mediated process in the epididymis, we stained with a standard NET marker: anti-myeloperoxidase (MPO). Unfortunately, this antibody did not work in my hands and proved inconclusive. Additionally, NET-associated citrullination is documented as a PAD4-mediated process and although we cannot rule this out, it is Padi2 – not Padi4 – that is expressed in the caput epididymis.

The finding that extracellular DNA was present in sperm samples obtained from the caput epididymis of virgin males, but not mated males, raised additional questions of whether there were other detectable differences between virgin and mated male mice. For example, does mating change the small RNA profile of sperm in caput epididymis? We set out to repeat small RNA sequencing from both caput and cauda sperm from virgin males - as well as mated males - to find that mating status does not impact the small RNA content of sperm. That said, this small RNA sequencing was performed with Illumina’s Truseq method – which as we know from work in Chapter III, provides highly biased results.

In Chapter III, we set out to compare Illumina and NEB small RNA protocols to the novel OTTR protocol. We find that OTTR superior at capturing 3’ tRFs for the two tRF species we have Northern blot validation for. We find the OTTR “no treatment” control group and a “T4 PNK + ATP” group looked identical in nearly all cases of tRFs assessed giving us confident that OTTR is unaffected by RNA termini. This simplifies
our current protocol as it eliminates the need for enzymatic treatment prior to library preparation. Most shockingly, OTTR uncovered a rainbow of tRNA species that were previously undetectable. With both Illumina and NEB we saw only GlyGCC, ValCAC, GluCTC, and HisGTG. With OTTR we see all tRNA species represented to a certain extent – and almost always with equal distribution of 5′ and 3′ tRFs.

Of course, we have to consider that everything has a bias. And with both Illumina and NEB, we know where the bias lies - with OTTR, we are entering an unknown territory. We still don’t know the exact limitations of OTTR. What we cannot ignore is that OTTR captures 3′ tRNA fragments that are invisible to all other protocols examined. This alone will provide an updated view of the mouse sperm small RNA payload, and highlights the utility of OTTR for analysis of tRNA fragments by deep sequencing.

**Implications for the field**

The first major implications of the work in this thesis for the field are technical. In the case of the sperm methylome (Chapter II), it is imperative to treat sperm with DNase prior to any methylation analysis regardless of sperm origin (epididymal or vas deferens). Not only will DNase treatment of sperm prior genomic DNA extraction resolve any contamination from extracellular DNA for downstream methylome analysis, but quite frankly, this is an easy step to incorporate into any protocol. The second technical implication is that this extracellular DNA likely contaminates many published epigenetic measurements of sperm. For example, HiC measurements that claim that sperm resemble fibroblasts are most likely measurements of extracellular DNA combined with improper
and incomplete “opening” of sperm DNA. Additionally, many paternal effects papers cite changes in sperm DNA methylation (5-10% changes) due to paternal environment. These results can easily be explained by our extracellular DNA finding and even mature sperm can benefit from DNase treatment. In the case of small RNA cloning protocols (Chapter III), it is not ideal to return to the standard small RNA sequencing methods given what we know about broad variety of tRFs OTTR is able to capture – we cannot move the field forward with knowingly incomplete information.

The biological implications of the work in Chapter II remain untested. Our best hypotheses are that: A) the extracellular DNA in caput is a remnant of organ formation or B) it serves a protective function to prevent ascending epididymal infections. These hypotheses are highly speculative as we have tested neither – we are not set up to look at the immunological consequences of ascending reproductive infections nor studying organogenesis. These hypotheses also rely on the knowledge that this extracellular DNA is cleared upon mating which further complicates the meaning behind its biological function. Whatever the case, the proximal epididymis would be a strange place for a NETosis-like mechanism and requires further exploration. The biological implications for Chapter III are more open-ended as we are forced to re-evaluate what small RNAs sperm are delivering to an oocyte upon fertilization. We now face the reality that sperm contain not only 5′ and 3′ tRFs, but also some amount of all tRNA species – none of which have been well characterized for function in the developing oocyte.
Future directions

With respect to the sperm methylome (Chapter II), it is still unclear which cells are responsible for the production of extracellular DNA in the caput of virgin males, and several experimental questions would be valuable to address. First, testing additional antibodies (myeloperoxidase (MPO), neutrophil elastase (NE)) would help us to rule out a neutrophil-mediated process in the caput epididymis of unmated males.

Ideally, analyzing PAD KO mice - starting with PAD2 KO as Padi2 is the most highly expressed in the caput epididymis – will allow us to definitively determine if this is a PAD-mediated process and what cell type is responsible.

Furthermore, given the prevalence of studies documenting changes to sperm cytosine methylation in response to environmental exposure to toxins (for example BPA, dioxin) and more recently, studies documenting the similarities in three-dimensional genomic organization of sperm compared to somatic cells, it is necessary to report whether extracellular DNA is a source of contamination in other epigenomic data. For example, assaying sperm cytosine methylation with and without DNase treatment after toxin exposure at published, targeted loci to determine if extracellular DNA is the cause of the change in methylation - particularly at imprinted control regions (ICRs).

Additionally, perform Hi-C, a chromosome conformation capture sequencing technique, on mature sperm both with and without DNase treatment prior to crosslinking to determine the impact of extracellular DNA on 3C methods.

With respect to OTTR (Chapter III), it is critical that we expand OTTR to our paternal paradigms – starting with diet. We have established that males fed a low protein
diet have increased 5′ tRF-GlyGCC, but until now, we have only been able to see the 5′ side. With OTTR we will be able to discern if this increase in 5′ tRF-GlyGC, correlates with an increase with 3′ tRF-GlyGCC. We know the impact of 5′ tRF-GlyGCC on the developing offspring – what about 3′? As we have not tested this, OTTR opens the door to knowing the true small RNA content of sperm under different paternal environments and guiding how we – and others - design future experiments with respect to paternal effects paradigms.

One of the fundamental questions we want to address is how the small RNA payload of sperm has the possibility to impact offspring development. Before we invest in OTTR for paternal paradigms, would be wise to validate OTTR using a least several additional probes with Northern blotting and optimize the OTTR protocol with unique molecular identifiers (UMIs). It would also beneficial to test OTTR using additional enzymatic treatments – alongside T4 PNK treatment – using Tobacco Acid Pyrophosphatase (TAP) or pyrophosphohydrolase (RppH) to convert 5′ triphosphate groups (5′PPP) to 5′ phosphate groups (5′ P).

For this work, additional computational attention is strongly needed. As mentioned, with the OTTR protocol we have an incredibly high rate of reads that are “unaligned” (>50%). We need to address if these are truly mapping to tRFs – and just misincorporations providing insight into how OTTR handles modifications – or if these are reads mapping to another class of heavily modified small RNA. We should test this computationally, but also practically; one way to try to get at this would be to generate synthetically modified RNAs (with modified phosphoramidites) at specific sites to gain
insight into how OTTR handles these modifications – and what site-specific misincorporations it makes. With this, we can provide significant and much needed insight into the mechanisms and factors modulating sperm epigenomics during post-testicular sperm maturation.
APPENDIX I.

IMPACT OF AGING ON THE SPERM EPIGENOME
A.1 ABSTRACT

Advanced paternal age is associated with adverse changes in reproductive hormones, sperm count, motility, and morphology as well as conception rates, miscarriage rates, and even long-term consequences to offspring health. Here, we set out to investigate if advanced paternal age can change the gene expression profile of the epididymis and consequently, the composition of the sperm epigenome. We first set out to explore the effects of aging on gene expression in the testes and epididymis by generating RNA-sequencing datasets from males of three ages: 3 weeks, 6 weeks, and 40 weeks. We find striking gene expression changes in both caput and cauda epididymis – this includes a decrease in gene expression in known fertility and maturation markers as well as an increase in immune response – consistent with “inflammaging”. A pilot study exploring the effects of age on the sperm epigenome reveal changes in the small RNA content of sperm in 40-week-old males as compared to the younger, control group. These results increase our curiosity in age-related changes to the sperm epigenome and the potential impact in altering offspring phenotype. These preliminary experiments are foundational in allowing us to address the impact of aging on the understudied male reproductive accessory organs – the epididymis. Additionally, the results presented here provide new insight into the aging process and its impact on the epigenetic information carried in sperm.
A.2 INTRODUCTION

Delayed parenthood is a visible trend in developed countries although not without reproductive consequences. The reproductive risks associated with advanced maternal age are well-studied, but less is known about the impact of paternal age or its potential impact on offspring phenotype. This is likely due to decline in male fertility being more subtle than age-associated decline in female fertility, which is often depicted as a “fertility cliff”. Nevertheless, advanced paternal age is associated with adverse effects to changes in reproductive hormones and consequently negative impacts to sperm count, motility, and morphology as well as conception rates, miscarriage rates, and even long-term consequences to offspring health (Jenkins et al., 2018).

Interestingly, advanced paternal age is associated with increased risk of psychiatric disorders like schizophrenia, bipolar disorder, and obsessive compulsive disorder (Vervoort et al., 2021; Rieske et al., 2020; Buizer-Voskamp et al., 2011; Miller et al., 2011; Petersen et al., 2011; Dalman, 2009; Sipos et al., 2004) – even in mouse studies, offspring of older males had social deficits (Zhao et al., 2020; Smith et al., 2009). Of course, these studies are only correlative, and the molecular impact of advanced paternal age on offspring development and increased psychiatric disease risk remains ambiguous.

Many hypothesize that advanced paternal age contributes to de novo mutations in sperm and thus, logically, is assumed to be the cause and underlying mechanism (Graten et al., 2016). However, we find it intriguing that (presumably random) de novo mutations would lead to the same phenotype in offspring (ie psychiatric disorders). As it is now well-established that epigenetic information in the germline can respond to environmental
conditions – we wonder if age is just another “environmental condition” – like diet or exposure to toxins.

As we have covered the impact of the paternal effect paradigms in mouse on altered offspring phenotype - and the hypothesis that the sperm epigenome is contributing to programming offspring phenotype in the Introduction of this work (Chapter I), we focus here on age as a paternal effect. Advanced paternal age been shown to impact sperm DNA methylation levels. In comparing mature sperm from old and young mice, many have documented loss of methylation in sperm derived from older males. While the genomic location varies across studies, the data trends towards hypo-methylated regions in the genome in the sperm of aged males (Yoshizaki et al., 2021; Kobayashi et al., 2016; Milekic et al., 2015). Although DNA methylation is a plausible mechanism, we have been deceived by “changes” in DNA methylation before (see Chapter II) and would need to convince ourselves further given the technical necessity of treating sperm with DNase prior to genomic DNA extraction. Of course, these documented changes in DNA methylation in aged sperm could be real, but another hypothesis is that age is correlated to decreased sperm count skewing the methylation results towards any contaminating extracellular DNA from somatic cells. In these documented cases, it is impossible to rule out somatic cell contribution and therefore we are wary of the final interpretations - Kobayashi et al. (2016) even acknowledge that although they confirm the absence of somatic cells under a microscope, it is not possible to completely exclude somatic cell contamination.
In another notable example, advanced paternal age in mice significantly impacted several factors resulting in smaller embryonic placental weight, fetal weight, and length (Denomme et al., 2020). Examining DNA methylation and gene expression levels at imprinting control regions in the placenta (for example, Kcnq1ot1 ICR) found increased DNA methylation levels and decreases in gene expression for paternally-expressed imprinted genes. This is a striking finding as paternally-expressed imprinted genes are important for growth and development of the placenta – loss of paternally-expressed Mest causes placental growth restriction (Coan et al., 2005). Additionally, in our other paradigms, we are very much interested in the “black box” between paternal environment and offspring phenotype - alterations in placentation is our strongest best candidate for a potential mechanism.

Collectively, what we know about the mechanisms surrounding aging is complex. The progressive functional decline over time is linked to accumulating molecular damage, oxidative stress, and decline in mitochondrial quality and activity. The aging process is also accompanied by the chronic burden of background inflammation over time – termed “inflammaging” (Franceschi et al., 2018).

As we know sperm are heavily remodeled as they transit the epididymis, it is entirely feasible that age-induced changes to gene expression as well as protein and lipid production will negatively impact sperm outcome. Briefly, we recap what is known about the epididymis and aging – all studies in the brown Norway rat. We know hormone production is altered in the epididymis - Northern blot analysis was used to detect a decrease in 5α-reductase type 1 and type 2 mRNA in the caput and corpus of the brown
Norway rat (Viger and Robaire, 1995). This is significant as 5α-reductase (5AR) is the enzyme that converts testosterone to dihydrotestosterone – the predominant androgen-mediated hormone necessary for sperm maturation in the epididymis (Robaire and Hamzeh, 2011). Additionally, no significant decrease of 5α-reductase was found in the cauda suggesting that even with age – gene expression in the epididymis is region-specific. Morphologically, the epididymal lumen diameter has been documented to significantly reduce with age as well as an increase in the thickness of the basement membrane in all segments of the epididymis (Serre and Robaire, 1998). This decrease in diameter throughout the organ tells us that overall volume of the organ decreases – and likely contributes to lower sperm count – or the other way around. Either way, less space equals less sperm. Aging in the brown Norway rat was also accompanied by an increase in the number of halo cells in the corpus epididymis (Levy and Robaire, 1999). The halo cells – broadly known to be immune cells – were then further characterized using immunostaining of immune cell markers and found to be monocytes, helper T lymphocytes, and cytotoxic T lymphocytes – all increasing with age (Serre and Robaire, 1999). Most recently, cDNA microarray analysis focused on oxidative stress-related genes and a handful of other pathways to find a decrease in gene expression throughout the epididymis. Of the 254 genes assayed, they found no genes increasing in expression with age (Jervis and Robaire, 2002). This result, we believe to be a limitation of the genes assayed and microarray technology – something we aim to resolve with NGS.
Here, we set out to investigate if advanced paternal age can change the composition of the sperm epigenome – potentially resulting in altered offspring phenotype. To test this hypothesis we aim to perform the preliminary experiments described here. First, how does the gene expression profile of the epididymis change with age? Second, can we identify changes to the small RNA content of mature sperm aging changes the sperm epigenome starting (Figure A.1). To the best of our knowledge, no work has been done to address the small RNA content of aged sperm as compared to younger sperm in mice – all the focus has been on DNA methylation as the potential driver of phenotype. These preliminary experiments are foundational in allowing us to pursue single cell RNA-sequencing (scRNA-seq) of young and aged epididymis – as well as expanding further to include male accessory organs – seminal vesicles and prostate – in the future. Together, these studies will provide new insight into the aging process of understudied reproductive accessory organs in males and its impact on the epigenetic information carried in sperm. With this, we can test the hypothesis that age-related changes in epididymal gene expression can impact the sperm epigenome – and offspring phenotype.
A.3 RESULTS

Effects of aging on gene expression in the testes and epididymis

We first set out to explore the effects of aging on gene expression in the testes and epididymis by generating RNA-sequencing datasets from two to three males of three ages: 3 weeks, 6 weeks, and 40 weeks (Figure A.1). To convince ourselves of the quality of the dataset, we first set out to confirm tissue-specific gene expression for testes, caput, and cauda. For example, we confirmed expression of meiotic gene expression of the testes, and well-known genes of the proximal, caput epididymis (Rnase10 and Lcn8), and distal, cauda epididymis (Defb2 and Spink14) (Rinaldi et al., 2020).

We also know the first wave of spermatogenesis in mice takes ~35 days to complete the overall differentiation process. With this, we can confirm that the genes expressed in the later stages of spermatogenesis are upregulated in the testes at 6 weeks and 40 weeks (when compared to 3 weeks) (Ernst et al., 2019; Grive et al., 2019). These genes (Tnp1, Tnp2, Prm1, Prm2) are expressed in round and elongating spermatids – around days 25-30 – and therefore absent from the 3 week testes (Figure A.2).
Figure A.1 Schematic of RNA-seq and small RNA experimental samples
Preliminary experiments include A) RNA-seq of testes, caput and cauda tissue from males aged 3 weeks, 6 weeks, and 40 weeks as well as B) small RNA-seq of mature, cauda sperm from males aged 6 weeks and 40 weeks.
Figure A.2 Confirmation of Expression of *Prm1, Prm2, Tnp1, and Tnp2* in 3 week, 6 week, and 40-week-old testes samples

*Prm1*, *Prm2, Tnp1, and Tnp2* – protamine genes and transition protein genes are expressed in the later stages of spermatogenesis and thus absent from the 3 week testes samples. Calculated average TPM (Transcripts Per Million) from at least $n=2$ samples where error bars indicate standard deviation. P-values calculated for $<0.05$
With these proof of principle experiments, we are convinced by the quality of our RNA-seq data set and move on to make more interesting comparisons. In the testes, we see dramatic changes in gene expression between 3 weeks and 6 weeks (Figures A.3A and A.4) as well as 3 weeks and 40 weeks (Figures A.3B and A.4). As expected, enriched GO terms of the most significantly upregulated genes included spermatogenesis, sperm chromatin organization, and spermatid development. There were only modest - and not statistically significant - changes between 6 weeks and 40 weeks (Figure A.3C) in the testes. This suggested that aging in the testes likely subtle -this is in stark contrast to changes we see in the epididymis. Although we do see interesting changes between 3 weeks (the epididymis has yet to see sperm) and 6 weeks in both caput and cauda epididymis, here we focus our efforts on changes between 6 weeks and 40 weeks in the interest of gene expression changes with advancing age.
Figure A.3 Scatterplot analysis of normalized gene expression of testes
Normalized gene expression values comparing A) 6-week-old (x-axis) to 3-week-old (y-axis) B) 40 week (x-axis) to 3-week-old (y-axis) and C) 40 week (x-axis) to 6-week-old (y-axis) where significantly upregulated genes are highlighted (red) as well as downregulated genes (blue).
Figure A.4 Heat map of top ten most variably expressed genes in testes
Genes significantly increasing in expression by 6 and 40 weeks include Tnp1, Tnp2, Prm1, Prm2, Akap12, Odf2, Hk1, and Atp8B3. Showing only top ten most variable genes with Z score cut off 1.96.
In comparing changes between 6 weeks and 40 weeks in the caput epididymis, we see widespread changes providing a detailed view of the regulatory alterations accompanying aging. We see downregulated genes as well as the majority of statistically significantly genes being upregulated with age (Figure A.5) – something microarray experiments were not able to capture. Analysis of enriched downregulated pathways are primarily transport pathways (Table A.1). – specifically, transmembrane transport, ion transport, and carbohydrate transport. We know the epididymis ships necessary factors to sperm for function and maturation, so it is notable to see such significant changes at 40 weeks. Many of these downregulated genes - Clu, Gpx5, Adam7, Slc7a5, Slc38a5, RNase 9, RNase10, and Tmem258 (Figures A.6 and A.7) - we know these to be paramount to proper epididymal function. For example, GPX5 is a specialized antioxidant enzyme functioning to protect sperm from oxidative stress (Aitken, 2009) and Gpx5-/- male mice have a higher incidence of miscarriage and developmental defects (Chabory et al., 2009). Another, ADAM7 (a disintegrin and metalloprotease 7) – is a membrane-anchored protein transferred to sperm membranes – and found in high abundance in caput epididymosomes (Nixon et al., 2019). Adam7-/- mice are abnormal in sperm function and morphology with decreased mobility and deformed tails suggesting Adam7 is critical for proper sperm maturation (Choi et al., 2015). This is because ADAM7 forms complexes with a number of proteins of the sperm membrane – specifically calnexin (CANX), heat shock protein 5 (HSPA5), and integral membrane protein 2B (ITM2B) (Han et al., 2011). Although functionally unclear, gene expression of Tmem258 has been documented to decrease with age (Glass et al., 2013).
Figure A.5 Scatterplot and heatmap analysis of normalized gene expression of caput
Normalized gene expression values comparing A) 40 week (x-axis) to 6-week-old (y-axis) males where significantly upregulated genes are highlighted (red) as well as downregulated genes (blue). B) Heatmap illustrating post differentially expressed genes in DEseq2 with a FDR cutoff of 0.05 and minimum fold change of 2
<table>
<thead>
<tr>
<th>Direction</th>
<th>adj.Pval</th>
<th>nGenes</th>
<th>Pathways</th>
</tr>
</thead>
<tbody>
<tr>
<td>Down regulated</td>
<td>1.40E-06</td>
<td>199</td>
<td>Establishment of localization</td>
</tr>
<tr>
<td></td>
<td>1.40E-06</td>
<td>89</td>
<td>Transmembrane transport</td>
</tr>
<tr>
<td></td>
<td>1.00E-05</td>
<td>33</td>
<td>Glycoprotein metabolic process</td>
</tr>
<tr>
<td></td>
<td>1.20E-05</td>
<td>189</td>
<td>Transport</td>
</tr>
<tr>
<td></td>
<td>1.20E-05</td>
<td>43</td>
<td>Lipid biosynthetic process</td>
</tr>
<tr>
<td></td>
<td>1.70E-05</td>
<td>88</td>
<td>Ion transport</td>
</tr>
<tr>
<td></td>
<td>1.70E-05</td>
<td>29</td>
<td>Glycoprotein biosynthetic process</td>
</tr>
<tr>
<td></td>
<td>2.90E-05</td>
<td>25</td>
<td>Glycosylation</td>
</tr>
<tr>
<td></td>
<td>3.00E-05</td>
<td>24</td>
<td>Protein glycosylation</td>
</tr>
<tr>
<td></td>
<td>4.40E-05</td>
<td>72</td>
<td>Lipid metabolic process</td>
</tr>
<tr>
<td></td>
<td>2.10E-04</td>
<td>41</td>
<td>Anion transport</td>
</tr>
<tr>
<td></td>
<td>2.10E-04</td>
<td>117</td>
<td>Organic substance transport</td>
</tr>
<tr>
<td></td>
<td>5.30E-04</td>
<td>17</td>
<td>Carbohydrate transport</td>
</tr>
<tr>
<td></td>
<td>7.90E-04</td>
<td>15</td>
<td>Membrane lipid biosynthetic process</td>
</tr>
<tr>
<td>Up regulated</td>
<td>5.60E-58</td>
<td>407</td>
<td>Immune system process</td>
</tr>
<tr>
<td></td>
<td>5.80E-57</td>
<td>312</td>
<td>Cell motility</td>
</tr>
<tr>
<td></td>
<td>1.50E-56</td>
<td>360</td>
<td>Movement of cell or subcellular component</td>
</tr>
<tr>
<td></td>
<td>8.20E-53</td>
<td>329</td>
<td>Locomotion</td>
</tr>
<tr>
<td></td>
<td>3.50E-48</td>
<td>277</td>
<td>Cell migration</td>
</tr>
<tr>
<td></td>
<td>4.90E-44</td>
<td>258</td>
<td>Regulation of immune system process</td>
</tr>
<tr>
<td></td>
<td>3.90E-42</td>
<td>507</td>
<td>Regulation of response to stimulus</td>
</tr>
<tr>
<td></td>
<td>4.80E-39</td>
<td>198</td>
<td>Positive regulation of immune system process</td>
</tr>
<tr>
<td></td>
<td>1.00E-38</td>
<td>199</td>
<td>Regulation of cell motility</td>
</tr>
<tr>
<td></td>
<td>1.60E-38</td>
<td>192</td>
<td>Regulation of cell migration</td>
</tr>
<tr>
<td></td>
<td>2.80E-38</td>
<td>365</td>
<td>Cell surface receptor signaling pathway</td>
</tr>
<tr>
<td></td>
<td>8.90E-38</td>
<td>151</td>
<td>Blood vessel morphogenesis</td>
</tr>
<tr>
<td></td>
<td>9.80E-38</td>
<td>241</td>
<td>Biological adhesion</td>
</tr>
<tr>
<td></td>
<td>1.30E-37</td>
<td>208</td>
<td>Regulation of cellular component movement</td>
</tr>
</tbody>
</table>

Table A.1 Enriched pathways in DEGs for caput epididymis
Figure A.6 Heat map of top ten most variably expressed genes in caput
Genes significantly decreasing in expression by 40 weeks include Clu, Adam7, and Gpx5 and genes significantly increasing in expression by 40 weeks include mitochondrial (MT) genes and Malat1. Showing only top ten most variable genes with Z score cut off 1.96.
Figure A.7 Decreasing expression of Clu, Gpx5, Adam7, Slc7a5, Slc38a5 and Tmem258 in 40-week-old caput epididymis as compared to 6-week-old. Calculated average TPM (Transcripts Per Million) from at least n=2 samples where error bars indicate standard deviation. P-values calculated for <0.05.
Upregulated pathways are highly enriched for immune system processes (Table A.1) – among them *Cd3g, Il1b, Ccl8, Ccl6, Pecam* and *Ccl21a* - a CCR7 ligand that controls the migration of immune cells (Koozai et al., 2017) (Figure A.8) - strongly suggesting immune infiltration into the aging caput epididymis. This is consistent with the process of “inflammaging” as a hallmark factor characterizes many aging tissues (Franceschi et al., 2018). When we zoom in to the top ten most variably expressed genes in caput between 6 weeks and 40 weeks (Figure A.6), we see a number of genes upregulated to be involved in mitochondrial respiration – this is consistent with the mitochondrial dysfunction seen in older animals (Haas, 2019). Among our strongly upregulated genes we see *Apod* (Apolipoprotein D) in aging caput epididymis - a known marker of aging tissues (Rassart et al., 2020; Dassati et al., 2014) - giving us confidence in the validity of this dataset (Figure A.8). Furthermore, this data is convincing in the widespread effects of aging the epididymis and its likely impact on sperm maturation.
Figure A.8 Increasing expression of *mt-Co1, Ccl21a, Apod, Pecam, Aqp1* and *Ly6a* in 40-week-old caput epididymis as compared to 6-week-old
Calculated average TPM (Transcripts Per Million) from at least n=2 samples where error bars indicate standard deviation. P-values calculated for <0.05
In comparing changes between 6 weeks and 40 weeks in the cauda epididymis, we again see widespread changes in gene regulation, but more genes significantly downregulated than upregulated (Figure A.9). Unlike caput where downregulated genes pathways are enriched primarily in transport, in cauda we see downregulation pathway terms like: sperm motility, male gamete generation and reproductive process (Table A.2). Some notable examples required for male fertility include Drc7, Iqcg, Catsper1, Spem1, and Rnase9. Focusing on a subset of these downregulated genes by 40 weeks, (Figures A.10 and A.11) we find Crisp1, Itm2b, and Cd52, RNase9, RNase12, and Adam7, and focus on their roles and importance in male fertility.

As mentioned, ADAM7 forms complexes with a number of proteins of the sperm membrane - including (ITM2B) thus playing a critical role in a functional sperm membrane. (Han et al., 2011). Crisp1 is expressed in the epididymis, found in the lumen and proteins are found on the surface of sperm – it is known to play a role in sperm-egg adhesion – specifically in the first step of sperm binding to the zona pellucida (Cohen et al., 2008). Cd52 is expressed on lymphocytes - and epididymal cells. This antigen binds to the sperm membrane and functions as a known maturation marker for sperm (Yamaguchi et al., 2008).

Of note, the RNase A family is thought to be the only enzyme family that is vertebrate-specific – with a familiar family member - RNase 5 (angiogenin). Interestingly, RNases 9-13 are expressed specifically in the male reproductive tract and are considered the “non-canonical” family members due to their lack of documented ribonuclease activity. Furthermore, RNases 9 and 10 exclusively expressed in the
epididymis (Cho et al., 2005). $RNase\ 9$ is expressed throughout the epididymis (and decreases with age in caput and cauda in our dataset), is androgen dependent (Zhu et al., 2007), and sperm motility is impaired in $RNase9$-null sperm informing us that its loss impaired sperm maturation overall (Westmuckett et al., 2014).
Figure A.9 Scatterplot and heatmap analysis of normalized gene expression of cauda
Normalized gene expression values comparing A) 40 week (x-axis) to 6-week-old (y-axis) where significantly upregulated genes are highlighted (red) as well as downregulated genes (blue). B) Heatmap illustrating post differentially expressed genes in DEseq2 with a FDR cutoff of 0.05 and minimum fold change of 2
<table>
<thead>
<tr>
<th>Direction</th>
<th>adj.Pval</th>
<th>nGenes</th>
<th>Pathways</th>
</tr>
</thead>
<tbody>
<tr>
<td>Down regulated</td>
<td>1.20E-77</td>
<td>252</td>
<td>Male gamete generation</td>
</tr>
<tr>
<td></td>
<td>2.00E-71</td>
<td>239</td>
<td>Spermatogenesis</td>
</tr>
<tr>
<td></td>
<td>3.20E-67</td>
<td>306</td>
<td>Sexual reproduction</td>
</tr>
<tr>
<td></td>
<td>7.70E-62</td>
<td>269</td>
<td>Gamete generation</td>
</tr>
<tr>
<td></td>
<td>1.30E-55</td>
<td>320</td>
<td>Multi-organism reproductive process</td>
</tr>
<tr>
<td></td>
<td>1.00E-54</td>
<td>290</td>
<td>Multicellular organism reproduction</td>
</tr>
<tr>
<td></td>
<td>1.60E-52</td>
<td>284</td>
<td>Multicellular organismal reproductive process</td>
</tr>
<tr>
<td></td>
<td>5.90E-49</td>
<td>377</td>
<td>Reproductive process</td>
</tr>
<tr>
<td></td>
<td>9.60E-44</td>
<td>167</td>
<td>Cellular process involved in reproduction in multicellular organism</td>
</tr>
<tr>
<td></td>
<td>5.10E-35</td>
<td>55</td>
<td>Cilium movement</td>
</tr>
<tr>
<td></td>
<td>1.80E-33</td>
<td>84</td>
<td>Meiotic nuclear division</td>
</tr>
<tr>
<td></td>
<td>1.10E-31</td>
<td>86</td>
<td>Meiotic cell cycle process</td>
</tr>
<tr>
<td></td>
<td>2.40E-30</td>
<td>124</td>
<td>Germ cell development</td>
</tr>
<tr>
<td></td>
<td>6.10E-29</td>
<td>61</td>
<td>Sperm motility</td>
</tr>
<tr>
<td>Up regulated</td>
<td>4.10E-100</td>
<td>650</td>
<td>Regulation of multicellular organismal process</td>
</tr>
<tr>
<td></td>
<td>2.70E-94</td>
<td>540</td>
<td>Cell surface receptor signaling pathway</td>
</tr>
<tr>
<td></td>
<td>3.20E-92</td>
<td>529</td>
<td>Immune system process</td>
</tr>
<tr>
<td></td>
<td>1.60E-91</td>
<td>705</td>
<td>Regulation of response to stimulus</td>
</tr>
<tr>
<td></td>
<td>5.60E-86</td>
<td>804</td>
<td>System development</td>
</tr>
<tr>
<td></td>
<td>7.90E-80</td>
<td>350</td>
<td>Biological adhesion</td>
</tr>
<tr>
<td></td>
<td>7.90E-80</td>
<td>551</td>
<td>Regulation of developmental process</td>
</tr>
<tr>
<td></td>
<td>7.30E-78</td>
<td>345</td>
<td>Cell adhesion</td>
</tr>
<tr>
<td></td>
<td>1.40E-76</td>
<td>463</td>
<td>Regulation of multicellular organismal development</td>
</tr>
<tr>
<td></td>
<td>2.10E-73</td>
<td>423</td>
<td>Positive regulation of multicellular organismal process</td>
</tr>
<tr>
<td></td>
<td>2.20E-73</td>
<td>528</td>
<td>Anatomical structure morphogenesis</td>
</tr>
<tr>
<td></td>
<td>2.50E-72</td>
<td>356</td>
<td>Cell migration</td>
</tr>
<tr>
<td></td>
<td>6.80E-69</td>
<td>348</td>
<td>Immune response</td>
</tr>
<tr>
<td></td>
<td>1.40E-67</td>
<td>334</td>
<td>Regulation of immune system process</td>
</tr>
<tr>
<td></td>
<td>7.40E-67</td>
<td>456</td>
<td>Positive regulation of response to stimulus</td>
</tr>
</tbody>
</table>

Table A.2 Enriched pathways in DEGs for cauda epididymis
Figure A.10 Heat map of top ten most variably expressed genes in cauda

Genes significantly decreasing in expression by 40 weeks include \textit{Crisp1}, \textit{Itm2b}, and \textit{Cd52} and genes significantly increasing in expression by 40 weeks include mitochondrial (MT) genes and \textit{Malat1}. Showing only top ten most variable genes with Z score cut off 1.96.
Figure A.11 Decreasing expression of *Crisp1, Cd52, RNase9, RNase12, Adam7* and *Itm2b* in 40-week-old cauda epididymis as compared to 6-week-old
Calculated average TPM (Transcripts Per Million) from at least n=2 samples where error bars indicate standard deviation. P-values calculated for <0.05.
What about upregulated pathways in the cauda with age? Like caput, these are highly enriched for immune system processes (Table A.2) and include \textit{Bcl6b, Pecam, Aire, Icoslg, Ccl3, Ccl6, Ccl8, Ccl9, Ccl21a, Il1b, Tlr2, Tlr4, and Tlr9} amongst others. And like caput, this immune pathway enrichment strongly suggests systemic immune infiltration throughout the epididymis and is consistent with the process of “inflammaging” (Franceschi et al., 2018). We also see a number of genes upregulated as involved in mitochondrial respiration - consistent with caput by 40 weeks. (Figures A.10 and A.12) And, we see \textit{Apod} (Apolipoprotein D) in aging (40 week) cauda epididymis - a known marker of aging tissues (Rassart et al., 2020; Dassati et al., 2014) again confirming that this is a systemic process – not limited to any specific region of the epididymis.
Figure A.12 Increasing expression of *mt-Cytb, Cxcl12, Apod, Pecam, Ccl21a*, and Ccl8 in 40-week-old cauda epididymis as compared to 6-week-old
Calculated average TPM (Transcripts Per Million) from at least n=2 samples where error bars indicate standard deviation. P-values calculated for <0.05.
Given the drastic changes we see in the caput and cauda epididymis with age, we next set out to ask, does age impact the small RNA levels in sperm? As 3-week-old males do not have mature sperm, we performed small RNA sequencing (Illumina Truseq) with cauda sperm isolated from 6 week and 40-week-old males (Figure A.1B) to test this hypothesis. We find that although the samples all exhibited the expected ratio we found previously - with the majority of small RNAs being 5′ tRNA fragments, we do see specific tRFs as changing with age. For example, we see a strong drop in 5′ tRF-ValCAC levels in sperm from 40-week-old males compared to 6-week-old males (Figure A.13A) - 5′ tRF-ValCAC is usually the third most abundant tRF in mature sperm. We also see an incredible increase in snoRNA cleavage in the sperm from 40-week-old males, which are absent in the younger samples (Figure A.13B). These results increase our curiosity in age-related changes to the sperm epigenome, but it should be noted that these experiments were performed prior to the knowledge of the OTTR protocol – this is work we will almost certainly transition to the OTTR protocol moving forward to assure ourselves that it holds true (covered in Chapter III).
**Figure A.13 Small RNAseq from 6 week and 40-week-old sperm**

In A) comparing levels of 5′ tRF-ValCAC and B) snord70 from 6week and 40-week-old males where each bar is equal to 1 mouse (6 weeks n=4, 40 weeks n=2) and y-xis is after normalizing to TPM.
A.4 DISCUSSION

The work presented here is intended as the foundation of a project with a bigger scope. Overall, we find striking gene expression changes in both caput and cauda epididymis – this is in stark contrast to the testes where modest, but not statistically significant changes occur. This informs us that the process of spermatogenesis is seemingly, largely unaffected with age. That said, it is entirely possible we have not looked at old enough males – generally aging studies use mice upwards of two years old. Even so, by 40 weeks, the process of post-testicular sperm maturation is sensitive and responsive to age, which we document through gene expression changes with decrease to known fertility and maturation markers as well as increase in immune response signifying low level immune infiltration. It is not surprising that we do not see any changes to immune regulation in the testes – that is because the testes are immune privileged in mammals. That is, the immune system doesn’t recognize sperm as “self” and to prevent the immune system from triggering an immune response to sperm, the testes can tolerate the introduction of “antigens” (ie sperm) without stimulating an immune response (Qu et al., 2019; Fijak and Meinhardt, 2006).

The bigger scope of this work will set out to analyze age-related changes in the epididymis on a single cell level. This will allow us to characterize the effects of aging on cell composition which we cannot detect at the same resolution using bulk RNA-seq. Additionally, we will include more – and older timepoints to gain a better understanding of the trajectory of aging. We would expand this to include data about the sperm
epigenome – specifically, histone retention, small RNA-seq, and DNA methylation in aged sperm samples.

We are also interested to know if age has an effect on preimplantation development. To address this, we can mate young and old males, isolate 2-cell, 4-cell, morula, and blastocyst stage embryos, and perform RNA-seq. With this, we can determine if age of sperm contributes to gene expression changes in the embryo. Next, we will select interventions (likely starting with caloric restriction) to see if we can slow – or even reverse these age-associated changes in gene expression in the epididymis, the epigenome, and even in gene expression changes in the preimplantation embryo if any are noted with age.

Beyond this, we want to expand past the epididymis and into additional male reproductive tissues – namely, seminal vesicles and prostate. These organs provide the fluid through which sperm is transported yet we understand very little about how their function changes with age. By comprehensively assaying the effects of paternal age on male reproductive tissues, we aim to better understand the mechanism underlying age as a paternal environment and how this can contribute to offspring phenotype.
A.5 METHODS

Mice

Testes, caput, and cauda tissues were obtained from 3, 6, and 40-week-old FVB/NJ males.

Dissection and tissue preparation

FVB mice, euthanized according to IACUC protocol, were dissected for testes, caput, and cauda. Testes and caput tissue were immediately snap frozen. Cauda epididymis was placed into Donners complete media (Hisano et al., 2013) and tissue was cleared of fat and connective tissue before incisions were made using a 26G needle while keeping the bulk tissue intact. Tissue was gently squeezed allowing sperm to escape into solution. After incubation at 37°C for 1 hour, cauda tissue was washed 3x in PBS and snap frozen while sperm containing media was transferred to a fresh tube and collected by centrifugation at 5000rpm for 5 minutes followed by a 1X PBS wash. To eliminate somatic cell contamination, sperm were subjected to a 1mL 1% Triton X-100 incubation 37°C for 15 mins with 1500 rpm on Thermomixer and collected by centrifugation at 5000rpm for 5 minutes. Somatic cell lysis was followed by a 1x ddH2O wash and 30 second spin 14000 rpm to pellet sperm.

Tissue RNA sequencing

1ug of total RNA from testes, caput and cauda were enriched for poly(A) RNA using NEBNext® Poly(A) mRNA Magnetic Isolation Module. Sequencing libraries were e
generating using NEBNext® Ultra™ RNA Library Prep Kit for Illumina® as per manufacturers instructions for paired-end sequencing.

**Sperm small RNA sequencing**

Isolation of sperm RNA was carried out as previously described (Sharma et al., 2018) followed by size selection of 18-40 nt RNAs from 15% polyacrylamide-7M urea denaturing gels. Sequencing library preparation was performed using Illumina’s TruSeq Small RNA Library Preparation Kit.

**Data analysis**

Small RNA sequencing was mapped to mm10 using Bowtie2. Abundance was determined by normalizing to parts per million based on total number of reads mapping to rRNA, microRNAs, tRNAs, piRNAs (repeatmasker and unique piRNAs), and mRNA (Refseq). (Yukselen et al., 2020). Tissue RNA-sequencing was mapped to GRCm38 (mm10) with STAR followed by read quantification using featureCounts and differential gene and differential gene expression analysis using DESeq2.
REFERENCES


Amitsur, M., Levitz, R., & Kaufmann, G. (1987). Bacteriophage T4 anticodon nuclease, polynucleotide kinase and RNA ligase reprocess the host lysine tRNA. The EMBO journal, 6(8), 2499–2503.


Brinkmann, V., Reichard, U., Goosmann, C., Fauler, B., Uhlemann, Y., Weiss, D. S.,


Buizer-Voskamp, J. E., Laan, W., Staal, W. G., Hennekam, E. A., Aukes, M. F.,


Carone, B. R., Fauquier, L., Habib, N., Shea, J. M., Hart, C. E., Li, R., Bock, C., Li, C.,
Gu, H., Zamore, P. D., Meissner, A., Weng, Z., Hofmann, H. A., Friedman, N., & Rando,


