2019-07-08

Mechanistic Basis for Control of Early Embryonic Development by a 5' tRNA Fragment

Xin Y. Bing
University of Massachusetts Medical School

Mechanistic Basis for Control of Early Embryonic Development by a 5' tRNA Fragment

A Dissertation Presented

By

XINYANG BING

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

July 8th 2019

Interdisciplinary Graduate Program

MECHANISTIC BASIS FOR CONTROL OF EARLY EMBRYONIC DEVELOPMENT BY A 5' tRNA FRAGMENT

A Dissertation Presented

By

XINYANG BING

This work was undertaken in the Graduate School of Biomedical Sciences

Interdisciplinary Graduate Program

Under the mentorship of

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

Paul Kaufman, Ph.D., Member of Committee

Sean Ryder, Ph.D., Member of Committee

Pavel Ivanov, Ph.D., External Member of Committee

Thomas Fazzio, Ph.D., Chair of Committee

Mary Ellen Lane, PhD.,

Dean of the Graduate School of Biomedical Sciences

July 8th 2019

This work is dedicated to

Mom and dad

Silvia

And

Helena,

Baba will always love you.

ACKNOWLEDGEMENTS

“A teacher for one day is a father for life” – Chinese idiom. With that in mind, I would like to thank my thesis mentor Oliver Rando. Ollie has taught me so much about not only how to rigorously conduct scientific inquiry, but how to approach science, mentorship, and life in general. His keen insight, breadth of knowledge, patience, generosity, and work ethic was a constant source of inspiration. The wealth of knowledge and experience that he has so kindly shared with me, I think sometimes maybe even unwittingly, will surely guide me throughout the rest of my career.

To my colleagues, thank you. It takes a village to raise a child, and without the members of the Rando lab and the UMass scientific community in general, I would certainly not have been able to complete my dissertation. From my first day in the lab, Upasna Sharma has been a patient and wonderful mentor to me. Since my project is also based on her findings, I owe much to her trailblazing. I specially thank Fengyun Sun for his guidance, expertise, and friendship; Io Long Chan, Shweta Kukreja, Nils Kriestenstein, Caroline Galan, Ebru Kaymak, Vera Rinaldi, and Colin Conine for help and discussions; past lab members Stanley Hsieh Tsung-Han, Caitlin Connelly, Ben Carone, Amanda Hughes, Jeremy Shea, Markus Vallaster for friendship and discussions. Thank you Lina Song for helping provide timely advice and support for the lab and I. Special thank you to Thomas Fazzio and his lab members Sarah Hainer, Kurtis McCannel, and Benson Chen for their help and reagent support. I am indebted to my thesis committee members Tom, Paul Kaufman and Sean Ryder for their support and discussions.

Last but not definitely not least, thank you so much Ana Bošković for being a great collaborator, superb mentor and friend.

I would also like to thank our collaborators Joanna Jachowicz and Maria-Elena Torres-Padilla for giving me the opportunity to better my computational skills. Thank you to all who have provided me with guidance, advice, and know-how: Emiliano Ricci, Wu Pei-Hsuan, Michal Rabani, Ozlem Yildirim, Dante Lepore, Tania Silva, Nick Rhind lab, Marian Walhout lab, Anastasia Khvorova lab, and Phil Zamore lab.

Lastly, thank you to my friends who have supported me throughout my Ph.D. and inspired me when I was lost, Ammeret Rossouw, Juliano Sangalli, Teresa Rzezniczak, and Michael Packer. And thank you to my old mentors Thomas Merritt and Brent Sinclair for giving me the opportunity to train and learn with you, and all the people I worked with in their labs.

ABSTRACT

Ancestral environmental conditions can instruct offspring development, although the mechanism(s) underlying such transgenerational epigenetic inheritance is unclear. In murine models focused on paternal dietary effects, we and others have identified tRNA fragments (tRFs) in mature sperm as potential carriers of epigenetic information. In our search for molecular targets of specific tRFs, we observed that altering the level of 5'-tRF Glycine-GCC (tRF-GG) in mouse embryonic stem cells (mESCs) and preimplantation embryos modulates the expression of the endogenous retrovirus MERV-L and genes regulated by MERV-L. Intriguingly, transient derepression of MERV-L is associated with totipotency of two-cell stage embryos and a subset of two-cell-like mESCs.

Here, I reveal the mechanistic basis for tRF-GG regulation of MERV-L. I show that tRF-GG supports the production of numerous small nuclear RNAs associated with the Cajal body, in mouse and human embryonic stem cells. In particular, tRF-GG modulates the levels of U7 snRNA to ensure an adequate supply of histone proteins. This in turn safeguards heterochromatin-mediated transcriptional repression of MERV-L elements. Importantly, tRF-GG effects on histone mRNA levels, activity of a histone 3'UTR reporter, and expression of MERV-L associated transcripts can all be suppressed by appropriate manipulation of U7 RNA levels. I also show that hnRNPF and H bind directly to tRF-GG, and display a stark overlap of *in vivo* functions to tRF-GG. Together, this data uncovers a conserved mechanism for a 5' tRNA fragment in the fine-

tuning of a regulatory cascade to modulate global chromatin organization during pre-implantation development.

TABLE OF CONTENTS

List of Figures

List of Copyrighted Materials Produced by the Author

Chapter II is reprinted and adapted from the co-authored work with permission from AAAS.

Sharma, U., Conine, C.C., Shea, J.M., Bošković, A., Derr, A.G., **Bing, X.Y.**, Belleannee, C., Kucukural, A., Serra, R.W., Sun, F.Y., Song, L., Carone, B.R., Ricci, E.P., Fauquier, L., Moore, M.J., Sullivan, R., Mello, C.C., Garber, M. and Rando, O.J. (2016) Biogenesis and function of tRNA fragments during sperm maturation and fertilization in mammals. *Science* 351(6271): 391-396.

List of Third Party Copyrighted Material

Figure 1.1	AAAS	Open
Figure 1.2	Elsevier	4626471164300
Figure 1.3	Springer Nature	4626471327028
Figure 1.4	Springer Nature	Open
Figure 1.5	Springer Nature	Open
Figure 1.6	Elsevier	4626480342926
Figure 1.7	Elsevier	4626480544349
Figure 1.8	Elsevier	4626480808540

Preface

Work in Chapter II was a collaborative project between the Rando lab, Mello lab, Moore lab, Garber lab, and Sullivan lab. Experiments were designed by US, CCC, JMS, AB, **XYB** and BRC. Experiments were conducted by US, CCC, JMS, AB, **XYB**, CB, RWS, FYS, LS, BRC, EPR, XZL, and LF. Data was analyzed by OJR, US, CCC, JMS, AB, AGD, XYB, CB, AK, MG. The manuscript was prepared by OJR and US. XZL, LF, MJM, RS, CCC, and MG provided infrastructure and funding support for specific experiments. This work is published:

> Sharma, U., Conine, C.C., Shea, J.M., Bošković, A., Derr, A.G., **Bing, X.Y.**, Belleannee, C., Kucukural, A., Serra, R.W., Sun, F.Y., Song, L., Carone, B.R., Ricci, E.P., Fauquier, L., Moore, M.J., Sullivan, R., Mello, C.C., Garber, M. and Rando, O.J. (2016) Biogenesis and function of tRNA fragments during sperm maturation and fertilization in mammals. *Science* 351(6271): 391-396.

Work in Chapter III was conducted entirely within the Rando lab. Experiments were designed by OJR, AB, and **XYB**. Experiments were conducted by AB, **XYB**, and EK. Data was analyzed by **XYB**, OJR, AB, and EK. The manuscript was prepared by OJR, AB, and **XYB**. This work has a preprint at:

> Boskovic, A., **Bing, X.Y**., Kaymak, E., and Rando, O.J. (2018) Control of noncoding RNA production and histone levels by a 5' tRNA fragment. Preprint. BioRxiv.

Work in Appendix I and II was conducted within the Rando lab. Experiments were designed and conducted by XYB, with input from OJR. Data was analyzed by XYB. The work is unpublished.

CHAPTER I

INTRODUCTION

Overview

Crops in cultivation often throw “rogues” that do not grow true to the variety sown. However, in 1915, William Bateson and Caroline Pellew described something peculiar about the “rogue” culinary peas that emerged sporadically from genetically identical self-fertilized seeds: their genetics did not agree with the prevailing chromosome theory of heredity, commonly referred to as Mendelian inheritance (Bateson and Pellew, 1915). The “rogue” phenotype was not only dominant in initial crosses, resulting heterozygous offspring continued to dominate crosses as if homozygous. This was in disagreement with how genes that reside on chromosomes are expected to segregate according to Mendel’s laws, and Bateson and Pellew used this example as a counterargument to the chromosome theory of heredity. Fate would have it that this work was largely ignored given that Thomas Hunt Morgan published definitive evidence in support of Mendelian heredity that same year (Morgan, 1915). Curiously, Alexander Brink described something similar to Bateson and Pellew in *Zea mays* at the *red1* (*r1*) locus decades later, which he termed paramutation (Brink, 1956, 1958). As in the “rogue” peas, these paramutagenic alleles that controlled the aleuronic pigmentation of maize emerged spontaneously, and converted one allele form to another in a heritable manner, violating Mendel’s Law of Segregation.

Exceptions, even if rare, to Mendelian inheritance often point to gaps in our understanding of eukaryotic genetics. Given hindsight, Alexander Brink and

colleagues did not attempt to describe paramutation as a violation of Mendelian inheritance, but sought alternative explanations that built upon the chromosome theory of heredity, such as chromosomal communication via homolog pairing, or cytoplasmic plasmids (Brink, 1956, 1958). We now know that the manner by which paramutable alleles are inherited immediately suggest the phenomenon as epigenetic in nature – the phenotypes are at times variable and reversible. In fact, these descriptions of paramutation were the earliest verifiable descriptions of transgenerational epigenetic inheritance (Arteaga-Vazquez and Chandler, 2010; Heard and RA Martienssen, 2014) – the inheritance of epigenetic information across generations.

Strictly defined, epigenetics refers to the perpetuation of phenotypes across cell divisions without changes to the DNA sequence (Ptashne, 2007; Berger *et al.*, 2009). Molecular and genetic studies have identified numerous pathways associated with epigenetic gene regulation, including covalent DNA modifications (Jones and Takai, 2001), transcription factors (Takahashi and Yamanaka, 2006; Ptashne, 2007), chromatin architecture (Alabert and Groth, 2012), small RNAs (Ghildiyal and Zamore, 2009; Borges and Martienssen, 2015), and prions (Harvey, Chen and Jarosz, 2018). The environment, both intrinsic and external, plays a crucial role in shaping the cellular epigenome (Feil and Fraga, 2012). Well studied examples range from intercellular communication such as the canonical Wnt signaling pathway (Komiya and Habas, 2008), to the phenotypic plasticity of whole organisms such as vernalization in plants (Kim *et al.*, 2009). In recent decades however, epigenetic mechanisms have been

implicated in inheritance of ancestral environmental cues in various organisms, harkening back to “Lamarckian” inheritance or the inheritance of acquired traits, and which led to the emergence of the field of transgenerational epigenetic inheritance (Heard and RA Martienssen, 2014; Miska and Ferguson-Smith, 2016).

As expected, transmission of epigenetic information across generations requires stable modifications to the germline epigenome. However, this process faces two theoretical barriers – the Weissman principle of the germplasm (1893), and epigenetic reprogramming (Morgan *et al.*, 2005; Heard and RA Martienssen, 2014). An abundance of studies in various model organisms, including plants, worms, flies, and rodents have revealed how ancestral environmental cues can bypass these barriers to affect epigenetic information in the germline, and how this information can be propagated through many generations (Daxinger and Whitelaw, 2012; Heard and RA Martienssen, 2014; Bošković and Rando, 2018). In addition, interest in the field continues to expand, given implications of public health and disease, as epidemiological evidence suggests that risk of metabolic diseases can be epigenetically inherited (Hales and Barker, 1992; Bateson, 2001; Lumey *et al.*, 2007). Importantly, the extent to which, and the mechanisms by which, this process occurs in mammals remain elusive and controversial (Pembrey *et al.*, 2014; Rando and Simmons, 2015; Horsthemke, 2018).

I describe herein a novel mechanism by which epigenetic information may be passed on through the male germline in mammals. First, I briefly review the topics that form the basis of the biology described (Chapter I), and then present

data suggesting a novel carrier of epigenetic information (tRNA fragment) capable of regulating embryonic gene expression (Chapter II). I next describe our efforts to uncover the molecular function of this tRNA fragment in the mouse embryo (Chapter III), and finally discuss the implications and potential experiments that could lead to further molecular insight and discovery (Chapter IV). In addition, I discuss experiments that could potentially contribute to our understanding of small fragment diversity (Appendix I) and tRNA fragment binding partners (Appendix II), which have not been published.

Inter- and transgenerational epigenetic inheritance

Mendelian and Non-Mendelian Inherited Epialleles

Phenotypic deviation is frequently observed in genetically identical organisms, termed epivariation due to the presumably epigenetic nature of this variation. In certain rare cases however, these variations are heritable through many generations, sometimes in a non-Mendelian manner, as in paramutation. Importantly, these epialleles invariably demonstrate metastable phenotypes, or a probabilistic nature of expression, and eventually, phenotypic reversibility.

Paramutation describes a process that results in heritable epigenetic changes of gene regulation and *trans*-homologue interactions (Chandler, 2007; Hollick, 2017). As one of the first examples of transgenerational epigenetic gene silencing, paramutation has been extensively studied in various organisms, including various species of plants, *Drosophila*, *C. elegans*, and mammals (Chandler, 2007). The most extensively studied cases of paramutation are of

maize pigmentation loci such as *r1* and *booster* (b1), which encode transcriptional factors that regulate production of flavonoid pigments and therefore have easily detectable phenotypes (Chandler, Eggleston and Dorweiler, 2000). However, detectability does not equate paramutability, as various loci of different pathways have been shown be involved in or amenable to paramutation (Chandler and Stam, 2004; Regulski *et al.*, 2013).

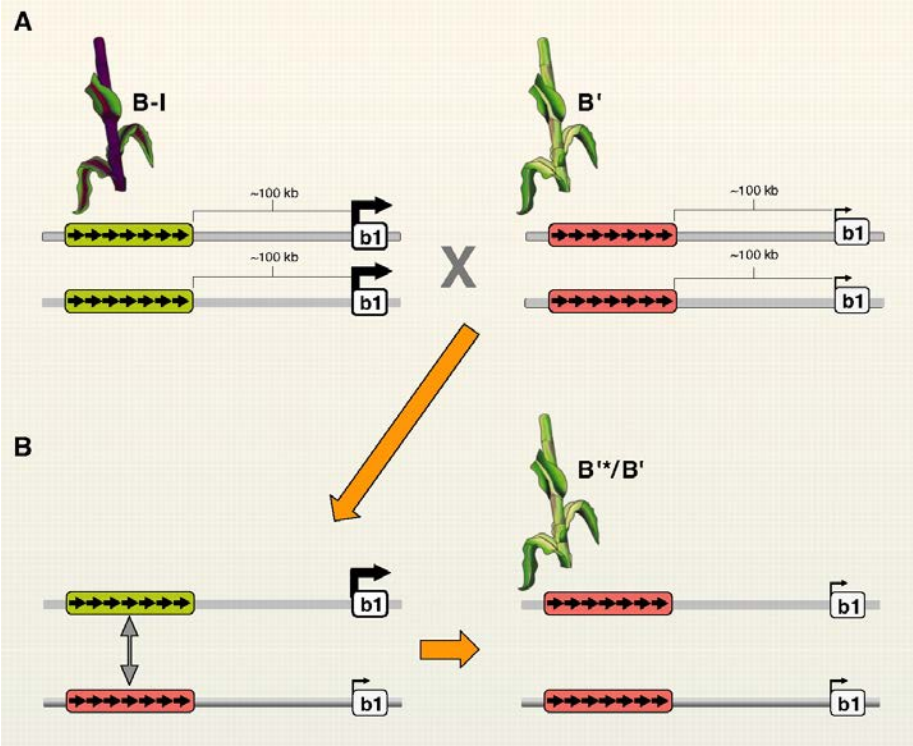


Figure. 1.1 Properties of *b1* paramutation. (**A**) *B-I* and *B′* phenotypes (dark and light colored plants, respectively) and diagrams of the *b1* locus and associated regulatory regions; because maize is diploid, the two diagrams for each plant represent the two alleles. The *b1* locus (white box labeled *b1*) encodes a transcription factor that activates the anthocyanin biosynthetic pathway, which produces purple coloration. When *b1* is highly transcribed (*B-I*, thick arrow above white box), a dark purple plant is observed. When transcription is low (*B′*, thin arrow above white box), a lightly pigmented plant is observed. *B-I* and *B′* have identical DNA sequences, including seven tandem copies of an 853–base pair (bp) repeat unit, located ~100 kb upstream of the *b1* coding region [indicated by seven black arrows within green (*B-I*) or red (*B′*) blocks]. The green and red blocks symbolize the distinct chromatin structures within the repeats in *B-I* and *B′*, respectively. Extensive data

demonstrate that the tandem repeats are required for *b1* paramutation and the high expression in *B-I*. The repeats have not been found elsewhere in the maize genome and are transcribed noncoding sequences that produce 24-nt siRNAs in both alleles. (**B**) The result of crossing *B-I* and *B′* is that *B-I* is always changed into *B′* by unknown mechanisms. The diagram portrays a two-step process (orange arrows), such that before establishment of paramutation both *B-I* and *B′* epigenetic states exist (left). Paramutation is established between early embryogenesis and the formation of 10 leaf primordia through unknown mechanisms mediated by the repeats (symbolized by the double-headed gray arrow), resulting in *B-I* being changed into *B′* (right). The new *B′* allele (*B-I* in the previous generation) is denoted *B′**, is mitotically and meiotically stably silenced, and is as capable as *B′* at changing *B-I* into *B′* in subsequent generations (not diagrammed). *From (Chandler, 2010). Reprinted with permission from AAAS.*

Importantly, the critical factor for paramutagenic alleles appears to be the number of repetitive sequences (Kermicle, Eggleston and Alleman, 1995; Hövel, Pearson and Stam, 2015; Hollick, 2017). In the famous example of the *b1* locus (**Fig 1.1**), the paramutagenic *B'* and paramutable *B-I* alleles both contain seven tandem repeat expansions of an 853-bp sequence – otherwise unique in the genome – in an enhancer region 100-kb upstream of the transcription start site that is required for paramutation to occur (Stam *et al.*, 2002; Belele *et al.*, 2013). Other examples of paramutation result from silencing mechanisms targeting transposable elements (TEs), which are inherently repetitive (Hollick, 2017). Therefore, repetitive sequences, whether exogenous or endogenous, appear to be the signal that recruits ancient host genomic surveillance mechanisms that target these genes for heritable silencing.

The silencing of these repetitive sequences that mediate paramutation behavior rests upon a major pillar of epigenetic information transfer – RNA interference (RNAi) (Hollick, 2017). Forward genetic screens have revealed the requirement of RNA-dependent RNA polymerase (RdRP) components in maize

(Dorweiler *et al.*, 2000; Hollick and Chandler, 2001), which are integral to small-RNA directed DNA methylation (RdDM) mechanisms that primarily target repetitive elements such as transposons for *de novo* cytosine methylation (Matzke, Kanno and Matzke, 2015). Since RdDM in *Arabidopsis* also requires histone 3 lysine 9 (H3K9) methylation marks to recruit RdRP to generate 24nt small-RNAs (Nobuta *et al.*, 2008; Erhard *et al.*, 2009; Hale *et al.*, 2009), all three major epigenetic information carriers appear to be involved in the transgenerational gene silencing of paramutable repeat elements in plants (Heard and RA Martienssen, 2014; Hollick, 2017). Paramutation in *Drosophila* and *C. elegans* similarly require the production of species-specific classes of small-RNAs, and their reinforcement of gene silencing also require H3K9me3 marks (Ashe *et al.*, 2012; Shirayama *et al.*, 2012; Luteijn and Ketting, 2013; Le Thomas *et al.*, 2014; Zhang *et al.*, 2014), underscoring the importance of the epigenetic crosstalk in the targeting and reinforcement of heritable gene silencing in the germline in animals and plants.

At about the same time as Brink and Marcus Rhoades were describing paramutation, Barbara McClintock recognized that the *Activator* and *Suppressor Mutant* transposons she discovered “cycled” between active and silent phases that are heritable through multiple generations (McClintock, 1951, 1961). These “cycling” transposons spontaneously generated epialleles of genes residing nearby, resulting in transgenerational leaf and seed color changes (Slotkin and Martienssen, 2007a; Lisch, 2012). In contrast to the unusual non-Mendelian inheritance patterns of paramutation, most cases of heritable epivariation

segregate in a Mendelian manner during meiosis, albeit in a metastable and reversible way. Examples of epialleles are abundant in plants, including in *Arabidopsis*, where powerful genetic screens laid bare the mechanistic basis of epivariation, which show a striking similarity to that of paramutation. With regards to the epialleles that resulted from McClintock's "cycling" transposons, the activity of transposons correlates with their level of DNA methylation (Chandler and Walbot, 1986) and heterochromatic marks (Li, Freeling and Lisch, 2010). Specifically, the expression of a hairpin transcript that generates *trans*-acting small-interfering-RNAs (siRNAs) induces transcriptional gene silencing through increased levels of DNA methylation surrounding the transposons (Slotkin, Freeling and Lisch, 2005). These siRNAs in turn influence the local chromatin context of TEs, as mutants in maintenance DNA methyltransferases, histone deacetylases, and chromatin remodelers lose transposon silencing potential (Law and Jacobsen, 2010). As in paramutation, epigenetic crosstalk is critical, as histone modifications and RNAi can rescue methylation defects to some extent (Mathieu *et al.*, 2007; Mirouze *et al.*, 2009; Zemach *et al.*, 2013; Creasey *et al.*, 2014).

Perhaps the most famous and well-studied case of epivariation and transgenerational epigenetic inheritance in mammals involves the agouti viable yellow (A^{vy}) allele in the mouse. Here as in plants, the A^{vy} allele arose from a spontaneous insertion of a transposon – an intracisternal A particle (IAP) retrotransposon 100kb upstream of the agouti gene. The long terminal repeat (LTR) of the IAP element acts as an alternative promoter that causes ectopic

expression of agouti protein, resulting in pleiotropic phenotypes of yellow fur, diabetes, and increased susceptibility to tumors (Daxinger and Whitelaw, 2012). Just like the cycling transposons in maize, A^{vy} mice display phenotypic mosaicism for IAP activity, due to variation in DNA methylation, resulting in deviation of coat color phenotype. Importantly, this epivariation is heritable through the maternal germ line, as embryo transfer experiments show that coat color is transmitted via oocytes, instead of potential intrauterine environmental effects (Morgan *et al.*, 1999). Mechanistically, the heritability of A^{vy} in mouse relies on similar pathways to those in *Arabidopsis*, including DNA methylation, histone deacetylases and chromatin remodelers (Daxinger and Whitelaw, 2012).

Together, epialleles that are meiotically inherited in plants and animals present a converging picture: they invariably arise from novel transposon insertions, and the genomic surveillance mechanism that leads to silencing of these transposons, while ancient, can be leaky, leading to metastable and reversible phenotypes. Perhaps most importantly, certain TEs consistently evade surveillance pathways in unexpected ways. For example, IAP elements escape complete demethylation in primordial germ cells and preimplantation embryos (Lane *et al.*, 2003; Smith *et al.*, 2012), but this process appears to differ in the male and female germline, resulting in maternal specific inheritance of DNA mosaicism, as in the A^{vy} example (Daxinger and Whitelaw, 2012). This dichotomy reveals two key insights into the processes of inter- and transgenerational epigenetic inheritance: male and female germ-lines have distinct reprogramming pathways, and TEs continue to conflict with eukaryotic genomes by “exploiting”

these windows of opportunity in a contemporary evolutionary arms race. Excitingly, the interactions of the host surveillance pathway and the transposon may generate new avenues of evolution that drive organismal innovation and genetic conflict, such as the probable case of genomic imprinting.

Genomic Imprinting

Recognized by mule breeders over 3000 years ago, but not formally described until 1991 at the *Igf2r* locus (Morison and Reeve, 1998), the process of genomic imprinting in mammals and flowering plants (Rodrigues and Zilberman, 2015) presents a case for epigenetic information which can be said to be "programmed": imprinted loci demonstrate highly heritable monoallelic expression (Kinoshita, Ikeda and Ishikawa, 2008; Bartolomei and Ferguson-Smith, 2011). The accurate parent of origin epigenetic marking, or in other words, the "escape" of epigenetic reprogramming, of these loci have over time become essential for normal mammalian development: elegant pronuclear transplantation experiments in mice demonstrated that appropriate imprinting is required for early mammalian development (McGrath and Solter, 1984; Surani, Barton and Norris, 1984). In humans, phenotypically distinct diseases also result from deletion of either the maternal or paternal imprinted region of chromosome 15q11-q13 (Angelman's syndrome or Prader-Willi disease, respectively, Knoll et al. 1989). It could be hypothesized therefore, that imprinting represents an evolved epigenetic inheritance, fixed into the population after coincidentally acquiring some essential developmental functions.

Our current understanding of the mechanism for imprinting in mammals supports this hypothesis. Maintenance of imprinted marks is provided by Krüppel-associated box (KRAB)-containing zinc finger proteins (KZFPs) ZFP57 (Li *et al.*, 2008) and ZNF445 (Takahashi *et al.*, 2019a). KZFPs are one of the fastest evolving gene families in mammals, and generally function to recruit repressive epigenetic modifiers that deposit H3K9me2/3 marks to TEs in a species-specific manner (Imbeault, Helleboid and Trono, 2017). At least in the case of ZFP57, its methylation-dependent binding to TGCmCGC present in imprinting control regions (ICRs) protects methylation erasure in preimplantation embryos, when global waves of DNA-demethylation take place (discussed below). Binding also recruits KAP1 (or TRIM28) and histone methyltransferases for H3K9 to further reinforce the memory of epigenetic silencing. Genomic hallmarks of ICRs include repeat invasion and expansion, including at the *PEG10* locus, and intronless pseudogenes, such as those in the *SNRPN* locus implicated in Prader-Willi-Angelman syndrome (Renfree, Timothy A Hore, *et al.*, 2009). Together, these facts point to novel retrotransposition events of TEs, which naturally evade epigenetic reprogramming, as key drivers of ICRs.

Taken together, given the numerous examples of paramutation, epivariation, and genome imprinting in plants and animals, heritable epigenetic variation exists in nature. These events mostly result from random integrations of TEs that lead to heritable epigenetic changes in expression of gene(s) located near the transposition event (Slotkin and Martienssen, 2007b). Various epigenetic mechanisms function cohesively to maintain the transgenerational

inheritance of these variable alleles, depending on the arsenal available to that organism, and which epigenetic marks are completely or incompletely erased during germline and embryonic reprogramming. In certain cases, such as imprinting in plants (Vaughn *et al.*, 2007; Schmitz *et al.*, 2011) and mammals (Renfree, Timothy A. Hore, *et al.*, 2009; Barlow and Bartolomei, 2014), these random epialleles ultimately resulted in “programmed” escape alleles (possibly from an epigenetic selective sweep) that is functionally integral to the proper development of the organism. However, these variations of epialleles do not support an adaptive epigenetic mechanism that could contribute to macroevolution. Support for “soft inheritance” or Lamarckian inheritance would require examples of adaptive epigenetic variation induced by the environment that is heritable across many generations, in animals and plants. Evidence supporting this idea remains elusive.

Transgenerational effects of environmental perturbations

Various environmental stressors have been shown to induce specific adaptive (or nonfunctional, or even maladaptive) phenotypes in future generations that are associated with changes in epigenetic states. In plants, maternal exposure to a variety of abiotic stresses, such as temperature (Brink, Styles and Axtell, 1968; Lacey, 1996), light (Schmitt, Niles and Wulff, 1992; Galloway, 2005), hyperosmotic stress (Wibowo *et al.*, 2016), and nutrient availability (Miao, Bazzaz and Primack, 1991; Schmitt, Niles and Wulff, 1992; Secco *et al.*, 2015), can induce adaptive phenotypes in the offspring. Maternal

effects are also apparent for biotic stresses such as predation (Agrawal, Laforsch and Tollrian, 1999) and pathogenic bacteria (Slaughter *et al.*, 2012). These adaptive phenotypes have in many cases been shown to have an epigenetic basis (Secco *et al.*, 2015; Wibowo *et al.*, 2016), with the best studied example being vernalization (Andrés and Coupland, 2012). Following prolonged exposure to cold, the ability to flower is re-acquired through epigenetic silencing of a floral repressor FLOWERING LOCUS C, via H3K27 methylation and antisense transcription silencing of the locus (Song *et al.*, 2012). However, this process is robustly reset in the germline and early embryo (Sheldon *et al.*, 2008), and cannot be inherited as an acquired trait (infamously attempted and failed by the Russian agronomist Trofim Lysenko). Attempts to demonstrate adaptive epigenetic variation in stress tolerance have also had limited success in plants, and maternal effects generally disappear within a couple of generations (Pecinka and Mittelsten Scheid, 2012).

Interestingly, temperature, especially heat stress, appears to be a consistent modulator of epigenetic inheritance in plants and animals, particularly in terms of the expression of TEs (Horváth, Merenciano and González, 2017). This effect was noted by both McClintock (McClintock, 1983) and Brink (Brink, Styles and Axtell, 1968) for maize TEs. Interestingly, this derepression of TEs in response to heat stress can be propagated for several generations (Ito *et al.*, 2011; Suter and Widmer, 2013; Migicovsky, Yao and Kovalchuk, 2014). Transgenerational effects of heat stress on TE expression have also been noted in *C. elegans* in terms of transgene silencing (Klosin *et al.*, 2017), and *Drosophila*

in terms of heterochromatin associated position-effect variegation (Gowen and Gay, 1933; Hartmann-Goldstein, 1967). However, the epigenetic mechanisms underlying TE derepression upon ancestral heat stress are usually species specific (Horváth, Merenciano and González, 2017), and it is also unclear how loosening of TE repression could impart adaptive fitness in the offspring.

Onset of diseases later in life have long been associated with adverse intrauterine environments in humans, in particular the nutritional status of the mother (Hales and Barker, 1992, 2001; Lane, Robker and Robertson, 2014). For example, cohort studies from the Dutch Hunger Winter clearly demonstrates that prenatal exposure to famine is associated with obesity, diabetes and cardiovascular diseases later in life (Lumey *et al.*, 2007). These studies helped lay the foundation for the developmental origins of health and disease, a formal variation of the “thrifty phenotype” hypothesis (Hales and Barker, 2001). Given the critical dependence of the fetus on intrauterine factors, it is expected that perturbations to the maternal environment would affect fetal development, *i.e.* fetal programming (Lane, Robker and Robertson, 2014; Rando and Simmons, 2015). Along these lines, *in utero* exposure to endocrine disruptors such as diethylstilbestrol (fetal poisoning) leads to birth defects (Titus-Ernstoff *et al.*, 2010), infertility (Palmer *et al.*, 2001), and cancer (Herbst, Ulfelder and Poskanzer, 1971) later in life. However, these effects were gone by the second generation (Titus-Ernstoff *et al.*, 2010), reminiscent of intergenerational maternal effects in plants (**Fig. 1.2**) (Heard and RA Martienssen, 2014).

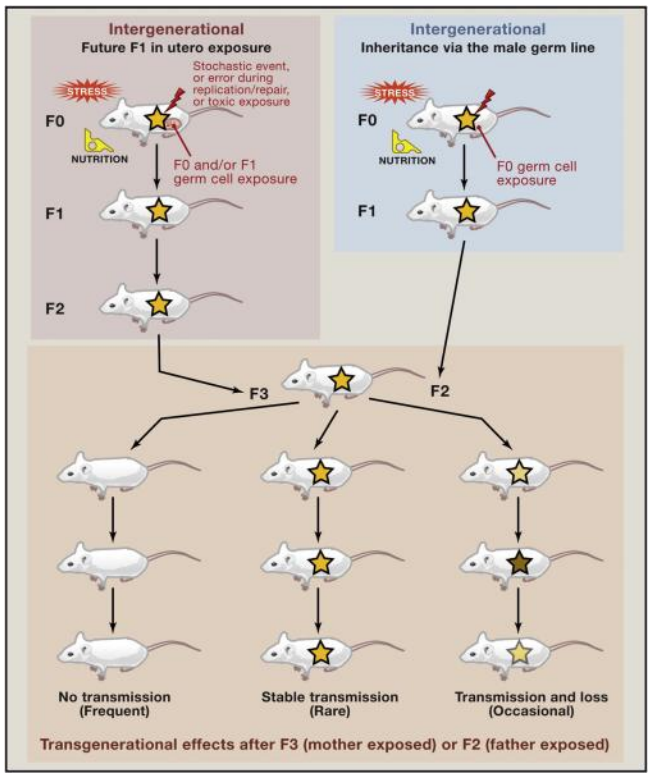


Figure 1.2. Transgenerational and Intergenerattional Epigenetic Effects. Epigenetic changes in mammals can arise sporadically or can be induced by the environment (toxins, nutrition, and stress). In the case of an exposed female mouse, if she is pregnant, the fetus can be affected in utero (F1), as can the germline of the fetus (the future F2). These are considered to be parental effects, leading to intergenerational epigenetic inheritance. Only F3 individuals can be considered as true transgenerational inheritance in the absence of exposure. In the case of males in which an epigenetic change is induced, the individual (F0) and his germline (future F1) are exposed; the F1 is thus considered as intergenerational. Only F2 and subsequent generations can be considered for evidence of transgenerational inheritance. *From (Heard and RA Martienssen, 2014). Reprinted with permission from Elsevier.*

Mechanistically, early embryonic development is highly sensitive to oviductal fluid composition (Leese *et al.*, 2008), in large part determined by maternal nutrition (Fleming *et al.*, 2012), which can influence blastocyst cell delineation (Kwong *et al.*, 2000; Sun *et al.*, 2014) and placentation (Fowden *et al.*, 2008). Importantly, pre-gestational and gestational influences on development are believed to occur through environment-induced modification of

the embryo's epigenome (Hochberg *et al.*, 2011). One of the best known examples involve studies of the A^{vy} locus in mice, where specific manipulations of methyl-donor availability during pregnancy can alter the heritability of the epiallele (Waterland and Jirtle, 2003). These direct maternal nutritional effects on the embryonic epigenome were later confirmed in *axin fused* mice (Waterland *et al.*, 2006), and have been demonstrated to modulate epiallele DNA methylation in human populations (Dominguez-Salas *et al.*, 2014). Other prominent maternal dietary manipulations in mouse models include high-fat diet (Sasson *et al.*, 2015; Huypens *et al.*, 2016) and zinc deficiency (Beach, Gershwin and Hurley, 1982). However, there is scant evidence that these effects can persist for more than one generation through the maternal lineage in animal models (Benyshek *et al.*, 2008) and epidemiological studies (Pembrey *et al.*, 2006). Instead, the epigenetic signal appears more readily passed on in the paternal lineage following maternal exposure *in utero* (Aiken and Ozanne, 2014; Rando and Simmons, 2015).

Transgenerational effects of undernutrition in the male lineage have been documented to occur in human populations (Kaati, Bygren and Edvinsson, 2002; Pembrey *et al.*, 2006). In studies of the Overkalix cohort, food availability in paternal grandparents is linked to obesity and cardiovascular disease in the grandchildren in a sex specific manner. Critically, the time of exposure was important to the observed effect: poor food availability in early adulthood was linked to decreased mortality risk, while the opposite is true for exposure in early adolescence (Pembrey *et al.*, 2006). This perhaps reflects the differential

sensitivity of male gametes to environmental perturbations during different processes of sexual development.

Despite strong evidence from epidemiological and animal model studies suggesting that paternal effects can influence up to two generations of offspring development, very few studies have reported phenotypes that persist past three generations in mammals (Dunn and Bale, 2011; Rando, 2012; Soubry, 2015). It is therefore unlikely that transgenerational effects of the environment can broadly influence macroevolution in humans, or any other organisms for that matter (Heard and RA Martienssen, 2014; Horsthemke, 2018). Crucially, it is clear that early nutritional cues can adversely affect offspring health later in life. Given the implications for public health, numerous rodent and other animal models have been developed with the explicit intent of understanding the extent and mechanism of inter- and transgenerational epigenetic inheritance in mammals.

A wide range of environmental perturbations can trigger paternal lineage offspring programming (Rando, 2012; Soubry, 2015; Boškovi´c and Rando, 2018). Rodent models have focused on three primary categories of environmental triggers: dietary interventions, stress exposures, and toxin exposures (Lane, Robker and Robertson, 2014). To identify pathways affecting metabolic parameters such as glucose control and lipid metabolism in the offspring, dietary intervention paradigms include high-fat diets, low-protein diets, and caloric restriction (Anderson *et al.*, 2006; Jimenez-Chillaron *et al.*, 2009; Carone *et al.*, 2010; Ng *et al.*, 2010a). Other studies attempt to explore parental effects on altered cortisol release, metabolism, and blood-brain barrier function

(Bale, 2015), stress exposures include social defeat (Dietz *et al.*, 2011), maternal separation (Gapp *et al.*, 2014a), and chronic unpredictable stress (Rodgers *et al.*, 2015), or simply odorant conditioning (Dias and Ressler, 2014). Finally to directly study ancestral toxic exposure on offspring development, a range of toxins and bioactive compounds have been utilized, from endocrine disruptors (vinclozolin, BPA etc.) (Anway *et al.*, 2005; Manikkam *et al.*, 2013) to drugs of abuse such as alcohol (Lam *et al.*, 2000), nicotine (Vallaster *et al.*, 2017), cocaine (Vassoler *et al.*, 2013) and morphine (Slamberová, Riley and Vathy, 2005). Many studies attempt to model late childhood and early adulthood stresses as in the Overkalix cohort, and involve perturbations applied from weaning to sexual maturity (Carone *et al.*, 2010; Ng *et al.*, 2010b; Fullston *et al.*, 2012). Most however are more interested in fetal development as in the Dutch Hunger Winter, and therefore subject pregnant females to the chosen treatment (Jimenez-Chillaron *et al.*, 2009; Pentinat *et al.*, 2010; Drake *et al.*, 2011; Dunn and Bale, 2011; Frantz *et al.*, 2011; Ding *et al.*, 2012; Martínez *et al.*, 2014; Radford *et al.*, 2014). Intriguingly, as observed in the Overkalix cohort in humans, the majority of studies observe a sex-specific effect, perhaps reflecting sex-specific germline reprogramming (Barlow and Bartolomei, 2014), or simply sex-specific embryonic developmental differences (Lane, Robker and Robertson, 2014).

What is the molecular mechanism for paternal transfer of epigenetic information? The jury is still out on this question (**Fig. 1.3**). First, non-germline mechanisms including seminal fluid, cryptic maternal effects, and transfer of microbiome can contribute to male-line effects (Curley, Mashoodh and

Champagne, 2011; Rando, 2012). Only a handful of studies have intentionally ruled these possibilities out by demonstrating gametic epigenetic information transfer (Dias and Ressler, 2014; Chen, M Yan, *et al.*, 2016; Huypens *et al.*, 2016; Sharma *et al.*, 2016). Second, most studies never rule out DNA sequence mutation as a possible mechanism, which becomes particularly problematic for treatments that could specifically influence DNA repair pathways (Skinner, Guerrero-Bosagna and Haque, 2015). For example, while outbred rats exposed to vinclozolin in utero exhibit diminished male fertility for up to four generations (Anway *et al.*, 2005), inbred rats subjected to the same treatment demonstrated no such effect (Schneider *et al.*, 2008), raising the possibility that genetic variability actually underlies the observed phenotype (Shea *et al.*, 2015). Third, reported epigenetic alterations to sperm and offspring in response to perturbations have been diverse and seemingly conflicting. These include sperm chromatin (Vassoler *et al.*, 2013; Siklenka *et al.*, 2015a; Ben Maamar *et al.*, 2018a), cytosine methylation (Carone *et al.*, 2010; Ng *et al.*, 2010b; Fullston *et al.*, 2012; Dias and Ressler, 2014), and small RNAs (Gapp *et al.*, 2014a; Rodgers *et al.*, 2015; Chen, M Yan, *et al.*, 2016; Grandjean *et al.*, 2016; Sharma *et al.*, 2016). Could all of these pathways mediate paternal epigenetic inheritance in mammals? If so, do they collaborate or exclusively signal one or a combination of environmental signals? Finally, what is the signaling pathway that causes specific epigenetic changes in the mature sperm, and how does this influence development of the fertilized embryo? Answering these burning questions will be

central to our understanding of inter- and transgenerational paternal epigenetic inheritance.

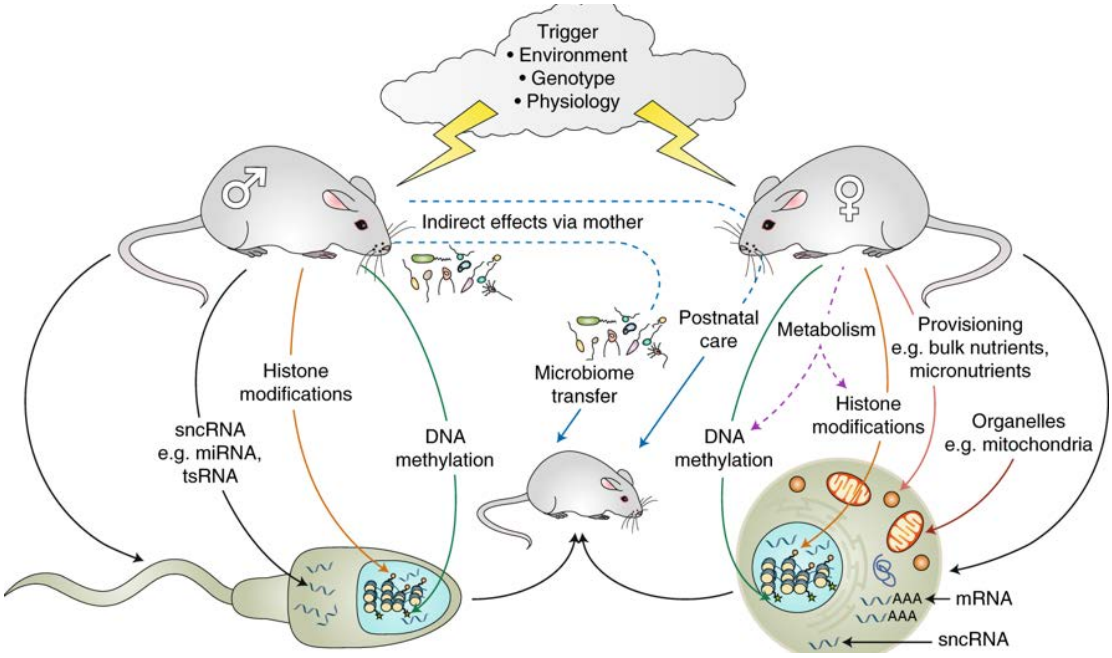


Figure 1.3. Mechanisms of transfer of information about ancestral environment or physiology over generations. Many mechanisms of transmission of information about environmental experience or physiological state can underlie inheritance over a single generation, from parents to progeny, both genome-associated (for example, covalent modifications of histones, sncRNAs, including tsRNAs (tRFs) and miRNAs, and DNA methylation, among others) and genome-independent (for example, microbiome transfer). Paternal effects are not always mediated by gametes but may act via the mother indirectly. *From* (Perez and Lehner, 2019)*. Reprinted with permission from Springer Nature.*

Escape of epigenetic reprogramming

In order to understand how certain genomic regions or regulatory factors can escape germline and early embryonic reprogramming, a mechanistic understanding of the processes themselves is necessary. Germ cells and embryos are unique in their epigenomic features at the chromatin, DNA methylation and small RNA composition levels. I will briefly describe these

unique aspects of both male, female germ lineages and developing embryos in animals, but a detailed description of mammalian male germline epigenetic reprogramming will be discussed in the next subchapter.

Chromatin in germ cells is highly unique in terms of both packaging and distribution of materials (Kimmins and Sassone-Corsi, 2005). In most animals, including mammalian sperm, histones are replaced by protamines to allow compaction of DNA into the sperm head (Sassone-Corsi, 2002). This essential process involves temporal replacement of nucleosomes with histone variants (Kimmins and Sassone-Corsi, 2005; Montellier *et al.*, 2013), then transition proteins, and finally protamines. Mammalian sperm genomes are hypermethylated at 90% of CpGs, particularly at repeats, guided by piRNAs (see below) (Popp *et al.*, 2010). However, a subset of promoter CpGs independent of CpG islands are hypomethylated (Messerschmidt, Knowles and Solter, 2014; Molaro *et al.*, 2014), and have been shown to be associated with the remaining 2-10% of nucleosomes (Hammoud, David A. Nix, *et al.*, 2009; Brykczynska *et al.*, 2010; Erkek *et al.*, 2013), although this topic remains highly controversial (discussed below) (Yamaguchi *et al.*, 2018). It is yet unclear how these regions escape epigenetic reprogramming. There is evidence that histone retention is developmentally programmed during spermatogenesis (Siklenka *et al.*, 2015a), and at least in zebrafish, sperm nucleosomes can act as placeholders to exclude DNA methylation and facilitate gene activation at embryonic genome activation, or EGA (*i.e.* act as epigenetic memory) (Liu *et al.*, 2018; Murphy *et al.*, 2018). Lastly, small-RNAs in sperm also carry a distinct cargo of mostly tRNA fragments

but also miRNAs, gained apparently via exosome fusion (discussed below) (Chen, M Yan, *et al.*, 2016; Sharma *et al.*, 2016, 2018). These small-RNAs that are gained during transit through the epididymis represent another source of epigenetic information that can potentially contribute to paternal epigenetic inheritance (Champroux *et al.*, 2018).

The epigenomic features of female gametes are quite different from their male counterparts, although also distinct from somatic cells. Oocytes likewise have germline specific histone variants, such as H1oo in mammals (Tanaka *et al.*, 2001), and display distinct post-translational modification profiles genome-wide, including H3K4me3 which spreads across domains of promoters and distant loci (Dahl *et al.*, 2016; Liu *et al.*, 2016; Zhang *et al.*, 2016). Probably due to epigenetic crosstalk, these intergenic H3K4me3 domains also correspond to hypomethylation of DNA (Zhang *et al.*, 2016). These broad domains cover about 40% of the maternal genome (Mayer *et al.*, 2000; Wang *et al.*, 2014), in stark contrast to the heavily methylated sperm genome. Surprisingly, gene bodies that are transcribed during oogenesis become hypermethylated in the mature oocyte (Tomizawa, Nowacka-Woszuk and Kelsey, 2012). Mechanistically, Dppa3/PGC7/Stella protects DNA from *de novo* and maintenance methylation in oocytes (Li *et al.*, 2018; Han *et al.*, 2019). This protective effect is at least partially mediated by preventing the nuclear accumulation of DNA methylation regulator UHRF1 (Bostick *et al.*, 2007). Finally, while the biogenesis of the maternal pool of small-RNAs in mammalian oocytes is understudied, the available data paint a peculiar picture: a reliance on endo-siRNAs, and

dispensable piRNAs and miRNAs (Tam *et al.*, 2008; Watanabe *et al.*, 2008; Yang *et al.*, 2016). Although abundant piRNAs exist in oocytes targeting specific TEs such as LINE-1 and IAP elements (Aravin *et al.*, 2008; Ohnishi *et al.*, 2010; Kabayama *et al.*, 2017), depletion of piRNA biogenesis pathway genes show no apparent phenotype (Deng and Lin, 2002; Fazio *et al.*, 2011; Reuter *et al.*, 2011). In addition, miRNA activity is suppressed in oocytes, possibly to prevent degradation of maternally deposited mRNAs essential to early embryonic development (Ma *et al.*, 2010; Suh *et al.*, 2010). These unusual features of mammalian oocyte small RNAs are at least partially mediated by expression of specific isoforms involved in small RNA activity, such as oocyte specific *Ago2* (Freimer *et al.*, 2018) and *Dicer* (Flemr *et al.*, 2013).

Fertilization triggers extensive epigenetic reprogramming in both parental nuclei. Sperm-specific nucleosomes are globally replaced by the replication-independent histone variant H3.3 in most organisms (Bonnefoy *et al.*, 2007; Santenard *et al.*, 2010; Autran *et al.*, 2011). Recent allelic characterization of H3K4me3 and H3K27me3 suggests that the H3.3 that replaces sperm histones does not retain immediate epigenetic “copying” (Zhang *et al.*, 2016; Zheng *et al.*, 2016). Nevertheless, there is evidence that heterochromatic and polycomb associated proteins are present in the male pronucleus shortly after sperm decondensation (Santos *et al.*, 2005), which awaits genome-wide analysis.

On the contrary, the maternal genome retains its maternally deposited chromatin for at least a few cell cycles, notably H3K4me3 (Zhang *et al.*, 2016) and H3K9me2/3 (Arney *et al.*, 2002), which only start to be displaced following

EGA. In yet another demonstration of epigenetic crosstalk, paternal DNA cytosine methylation, like its chromatin content, are rapidly turned over (actively demethylated) (Mayer *et al.*, 2000; Oswald *et al.*, 2000). While there is active demethylation in the maternal genome as well, this represents a small part of the passive demethylation that occurs globally due to nuclear exclusion of DNMT1 (Howell *et al.*, 2001; Ratnam *et al.*, 2002; Nakamura *et al.*, 2007). Mechanistically, the maternally deposited, H3K9me2/3 binding, STELLA blocks entry of TET proteins and DNMT1 globally (Nakamura *et al.*, 2007; Li *et al.*, 2018). Interestingly, while paternally deposited STELLA protects demethylation at paternal ICRs by binding to H3K9me2 (Nakamura *et al.*, 2007), maternally deposited STELLA plays no role in protecting maternal ICRs from demethylation (Li *et al.*, 2018). Instead, an oocyte specific and later somatic isoform of DNMT1 are able to evade exclusion and maintain methylation at these ICRs (Cirio *et al.*, 2008; Hirasawa *et al.*, 2008; Kurihara *et al.*, 2008). This evasion is achieved via recruitment by KAP1/TRIM28 (Messerschmidt *et al.*, 2012), which is bound to KZFPs ZFP57 (Li *et al.*, 2008; Mackay *et al.*, 2008) and ZFP445 (Takahashi *et al.*, 2019b) at ICRs.

Importantly, young retrotransposons such as IAPs in mice (Smith *et al.*, 2012), Alu's and LINE-1s in humans (Guo *et al.*, 2014), consistently escape proper demethylation, and may contribute to epigenetic reprogramming escape by A^{vy} epialleles (Daxinger and Whitelaw, 2012). It is probably not a coincidence that young retrotransposons are marked by H3K9me2/3 in embryonic tissue (discussed below) (Walter *et al.*, 2016), and this could be the precise mechanism

by which these loci escape reprogramming. In other words, perhaps the same mechanisms that protect ICRs from DNA demethylation in PGCs and during embryogenesis may protect young retrotransposons from being reprogrammed as well. More strikingly, the logic works in reverse as well: perhaps young retrotransposons escaping genetic reprogramming accidentally acquired important regulatory roles in development, and genomic imprinting arose from such novel transposition events (Rodriguez-Terrones and Torres-Padilla, 2018).

With the advent of ultra low-input chromatin profiling and small RNA technologies, our understanding of the unusual features of mammalian oocyte and sperm epigenetic landscape has greatly improved. However, key questions remain to be addressed that are critical to the mechanistic understanding of inter- and transgenerational epigenetic inheritance. First, the genomic locations of H3K9me2/3 nucleosomes in the sperm and oocyte genome pre-fertilization, and the mechanism by which they are copied, needs to be elucidated, since these nucleosomes clearly demarcate locations that are capable of escaping both chromatin and DNA methylation erasure. Second, is there crosstalk between small RNAs carried in both sperm and oocytes, and the epigenetic reprogramming taking place in the nucleus? All signs point towards the affirmative. Plants, worms and flies utilize small-RNAs to communicate with nuclear epigenetic machinery in the germline (Volpe and Martienssen, 2011; Heard and RA Martienssen, 2014; Miska and Ferguson-Smith, 2016), and piRNAs guide epigenetic reprogramming of the male mammalian germline (Aravin *et al.*, 2008; Kuramochi-Miyagawa *et al.*, 2008). This crosstalk has not

been identified in oocytes however, where siRNAs and piRNAs appear to exclusively target post-transcriptional gene silencing of endogenous retroviruses, or ERVs (Tam *et al.*, 2008; Watanabe *et al.*, 2008).

If crosstalk in the oocyte does not exist, can sperm-borne small-RNAs generate signaling or communication pathways to influence the epigenetic landscape of the developing embryo? A recent study suggests that sperm-borne small RNAs minimally impact the maternally deposited pool of small RNAs quantitatively (Yang *et al.*, 2016). However, microinjections of sperm-borne small RNAs have been repeatedly shown to mediate paternally induced phenotype in the offspring (Rassoulzadegan *et al.*, 2006; Gapp *et al.*, 2014b; Rodgers *et al.*, 2015; Chen, M Yan, *et al.*, 2016; Sharma *et al.*, 2016; Conine *et al.*, 2018). Perhaps sperm-borne small RNAs are functionally privileged by being bound to effector proteins or contain the proper modifications? This leads to the question of the mechanistic basis by which sperm-borne small RNAs can regulate epigenetic reprogramming in PGCs and early development in mammals, which I will introduce herein.

The Sperm Epigenome

Epigenesis

In mice, epigenetic reprogramming of the male germline begins at embryonic day 7.5 (E7.5), in the global erasure of epigenetic marks laid just a few days ago (**Fig. 1.4**) (Seki *et al.*, 2007). Here, newly emerged primordial germ cells (PGCs) experience global erasure of H3K9 methylation, increase in

H3K27me3 and various histone variants, and global loss of DNA methylation as they migrate to the genital ridge at E10.5-11.5 (Seki *et al.*, 2007; Popp *et al.*, 2010). Once PGCs populate the gonads, sexual dimorphism occurs and male PGCs continue to proliferate until an arrest at E14 (Tam and Snow, 1981), where they remain arrested until post-natal day 2. However, *de novo* DNA methylation of repeat elements and establishment of paternal imprints continue to occur in these quiescent cells (Kato *et al.*, 2007; Watanabe *et al.*, 2011), mediated by DNMT3A/B and DNMT3L proteins (Bourc'his and Bestor, 2004), and this process is essentially complete by birth.

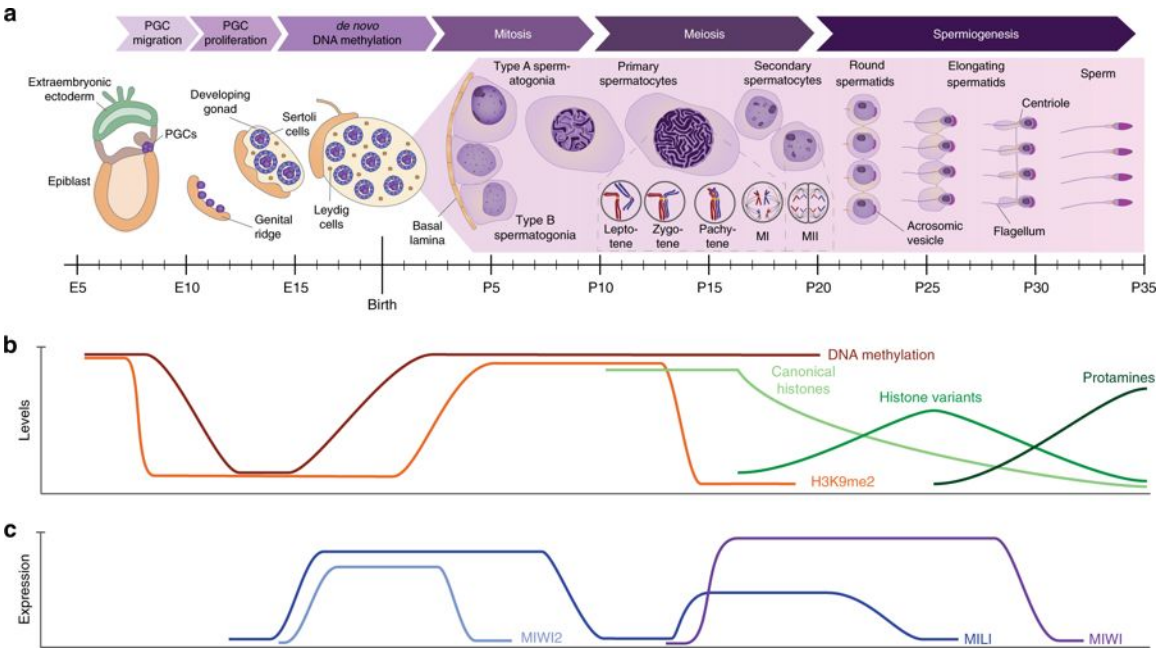


Figure 1.4. Male germ cell nomenclature and developmental dynamics of mouse spermatogenesis. A) Gametogenesis starts during embryonic development when primordial germ cells (PGCs) are defined and migrate to the genital ridge to form the gonads. Spermatogenesis initiates shortly after birth in synchronized waves. At 10 days post birth (P10), spermatogonial stem cells differentiate into primary spermatocytes that are committed to undergo meiosis. Two consecutive cell divisions (meiosis I and II (MI and MII)) without an intermediate S-phase result in the production of haploid gametes that are called round spermatids. These cells can be found as early as P20 and then undergo spermiogenesis during which the cells elongate and develop sperm-

specific structures such as the acrosome and the flagellum to form mature sperm cells. **B)** The process of gametogenesis is associated with extensive epigenetic reprogramming accompanied by drastic changes in DNA methylation and histone modifications such as H3K9me2. During later stages of spermatogenesis, global changes in histone composition and finally a histone-to-protamine exchange result in chromatin compaction. **C)** The three PIWI proteins encoded in the mouse genome show very specific expression profiles throughout spermatogenesis and reflect functionally distinct aspects of the piRNA pathway at different stages of spermatogenesis. *From* (Ernst, Odom and Kutter, 2017)*. Reprinted with permission from Springer Nature.*

This *de novo* DNA methylation is thought to be guided by piRNAs, which begin to be expressed at this stage – bound by mouse PIWI homologs MILI and MIWI2 (Kuramochi-Miyagawa *et al.*, 2008; Di Giacomo *et al.*, 2013). The biogenesis of these so called pre-pachytene piRNAs begins with single-stranded Pol II transcripts from defined genomic locations called piRNA clusters (Aravin *et al.*, 2006; Lau *et al.*, 2006). However, retrotransposon transcripts also make up a large proportion of piRNA precursors at this stage of development (Aravin *et al.*, 2007, 2008). Once exported into the cytoplasm with the help of Maelstrom (Castañeda *et al.*, 2014), the RNA helicase MOV10L1 unwinds and shuttles piRNA precursors to be cleaved by the endonuclease PLD6 or Zucchini (Zheng *et al.*, 2010; Ipsaro *et al.*, 2012; Han *et al.*, 2015; Vourekas *et al.*, 2015). MILI then binds with a strong bias towards 5’ uridines (Cora *et al.*, 2014), and 3’ exonucleolytic shortening of the 3’ end proceeds until the protected footprint of PIWI proteins has been reached – 26nt for MILI and 28nt for MIWI2, the characteristic lengths for these respective piRNA species (Girard *et al.*, 2006; Aravin *et al.*, 2007; Kawaoka *et al.*, 2011; Saxe *et al.*, 2013). Finally, piRNAs are 2’-O methylated by RNA methyltransferase HENMT1 for stability and proper

binding to PIWI proteins (Lim *et al.*, 2015). These MILI-bound primary piRNAs have strong homology to retrotransposons such as LINE-1 and IAPs, and MILI sliced transcripts anti-sense to primary piRNAs generate secondary piRNAs, which can be bound by either MILI or MIWI2, initiating an amplification loop called the ping-pong cycle (Aravin *et al.*, 2008). In this way, piRNA populations can be pruned to enrich for actively expressed TEs. MIWI2-bound piRNAs then translocate into the nucleus, where it detects nascent TE transcripts and recruits DNMT3/B/L for two waves of *de novo* DNA methylation over older followed by younger TEs (Molaro *et al.*, 2014; Itou *et al.*, 2015; Kojima-Kita *et al.*, 2016). MIWI2 is also able to recruit histone methyltransferases to establish repressive chromatin marks over TEs (Pezic *et al.*, 2014). However, this pathway is far from comprehensive, as there appears to be both MIWI2-independent nuclear silencing of TEs (Manakov *et al.*, 2015; Nagamori *et al.*, 2015), and MILI-independent pre-pachytene piRNA biogenesis pathways in fetal testis (Vasiliauskaitė *et al.*, 2017). Importantly, the function of pre-pachytene piRNAs does not appear to be in the repression of novel retrotransposition events *per se,* but instead in maintaining the proper chromatin landscape amenable to recombination during meiosis following birth (Zamudio *et al.*, 2015).

At post-natal day 2, male PGCs resume cell division, giving rise to self-renewing type-A spermatogonia, which in turn give rise to meiotically active type-B spermatogonia (Bellve *et al.*, 1977). Meiosis synchronously commences at P10, as spermatogonia move into the seminiferous tubule as pre-leptotene/leptotene, zygotene, pachytene, and finally diplotene spermatocytes at

P20, following the completion of meiosis I. During this period, H3K9me2 is lost as part of the developmental program driving H3K4me3 gain to allow for engagement of recombination machinery such as SPO11 (Zamudio *et al.*, 2015). Pre-pachytene piRNA-guided *de novo* DNA methylation over H3K9me2 marked TEs appears to bookmark and prevent H3K4me3 encroachment over these genomic regions, thereby preventing aberrant double-stranded breaks that can lead to meiotic arrest.

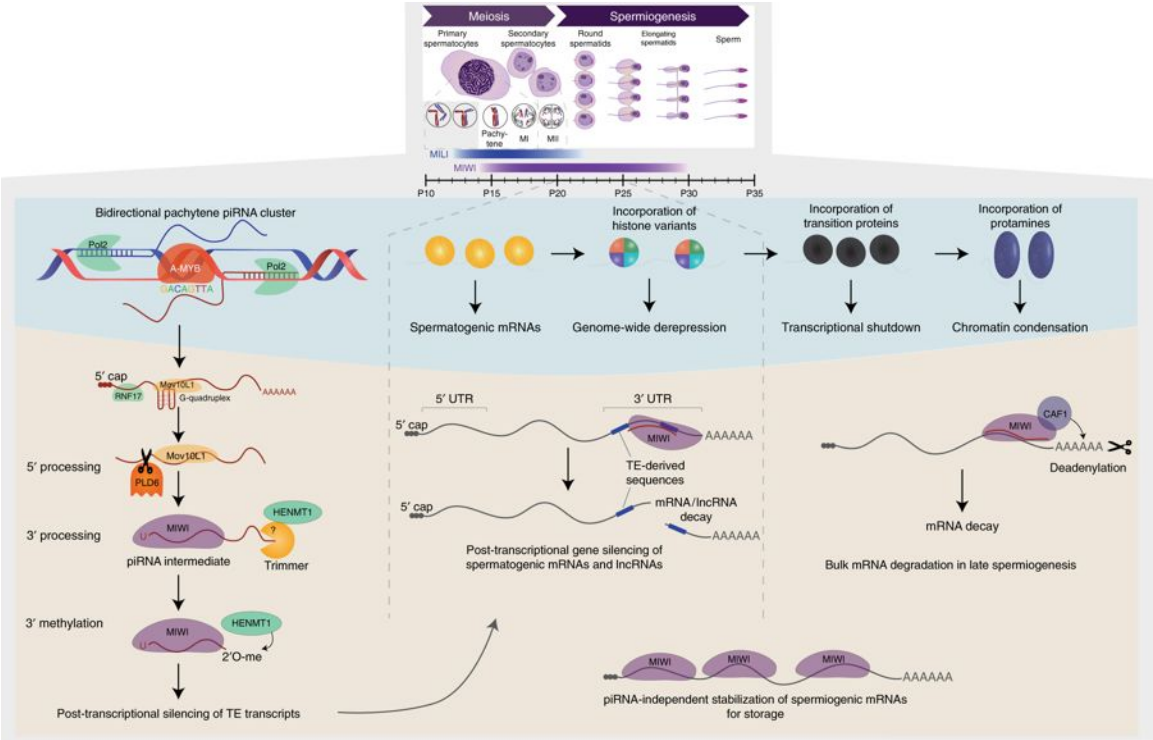


Figure 1.5. Multiple roles of action for pachytene piRNAs during mouse spermatogenesis. Pachytene piRNA expression starts during prophase of the first meiotic division when spermatocytes reach the pachytene stage. This expression is orchestrated by the transcription factor A-MYB, which drives expression of piRNA clusters and other piRNA pathway-related genes. Pachytene piRNAs are transcribed from distinct clusters in the genome often carrying bidirectional promoters. piRNA precursors are then processed in the same fashion as pre-pachytene piRNAs and mature pachytene piRNAs associate with either MILI or MIWI. Pachytene piRNAs engage in a myriad of functions throughout spermatogenesis, including the post-transcriptional silencing of TE transcripts but also non-TE-related functions. After meiosis,

round spermatids undergo extensive epigenetic remodeling, resulting in a genome-wide derepression due to the incorporation of histone variants, followed by transition proteins that lead to a transcriptional shutdown, and finally replacement of histones with protamines leading to chromatin condensation. During this differentiation process, pachytene piRNAs regulate spermatogenic mRNAs and lncRNAs that become transcribed due to the genome-wide derepression. Furthermore, MIWI is involved in the piRNA-independent stabilization of spermiogenic mRNAs to allow storage and translation after the transcriptional shutdown. Toward later stages, pachytene piRNAs direct global mRNA degradation in association with MIWI, which recruits CAF1 to induce deadenylation resulting in mRNA decay. *From (Ernst, Odom and Kutter, 2017). Reprinted with permission from Spinger Nature.*

After rapid completion of meiosis II, a process called spermiogenesis begins in round spermatids at P20 (**Fig. 1.5**) (O'Donnell, 2014). During spermiogenesis, sperm-specific structures such as the flagellum form, but another crucial epigenetic reprogramming process also occurs – nuclear compaction. Histones are replaced first by testis-specific histone variants, then transition proteins, and finally highly basic protamines, resulting in up to 20 fold genomic compaction (Balhorn, 2007). To facilitate epigenetic transition, histone variants are loosely packaged and spurious transcription is initiated globally, including the expression of TEs (Soumillon *et al.*, 2013; Ernst *et al.*, 2016).

The silencing of these TEs and spuriously transcribed mRNAs including long-non-coding RNAs are predominantly mediated by a second class of piRNAs bound by MIWI, called pachytene piRNAs (Girard *et al.*, 2006; Grivna *et al.*, 2006; Lau *et al.*, 2006), but also esiRNAs (Watanabe *et al.*, 2008), via post-transcriptional gene silencing. Pachytene piRNAs are transcribed from large intergenic non-coding and 3'-UTRs of protein-coding gene loci, with an apparent dearth of repeat content (Girard *et al.*, 2006; Robine *et al.*, 2009; Vourekas *et al.*,

2012). Biogenesis of pachytene piRNAs are largely similar to pre-pachytene piRNAs, with notable differences including transcription directionality (Li *et al.*, 2013; Homolka *et al.*, 2015) and unique RNA G-quadruplex structures (Vourekas *et al.*, 2015). Interestingly, pachytene piRNAs not only help to silence TEs during the latter stages of spermiogenesis, but regulate meiotic gene expression (Goh *et al.*, 2015; Zhang *et al.*, 2015), elimination of mRNAs in post-meiotic elongating spermatids (Gou *et al.*, 2014), and mRNA storage for later translation following transcriptional shutdown (Vourekas *et al.*, 2012).

The Sperm Histone Code?

Most relevant to the mechanism of inter- and transgenerational epigenetic inheritance through the male germline, histone to protamine transition is incomplete, and about 5-10% of histones remain associated with the genome in mature sperm (Hammoud, David A Nix, *et al.*, 2009; Brykczynska *et al.*, 2010). The genomic locations of these residual nucleosomes remain highly controversial, with groups claiming either developmental gene promoter or gene-dessert enriched nucleosomes in mature sperm (Saitou and Kurimoto, 2014; Yamaguchi *et al.*, 2018; Yoshida *et al.*, 2018). It has been argued that the discrepancy in nucleosome position mapping results from MNase over-digestion (Carone *et al.*, 2014), and a recent study does support this line of reasoning (Yoshida *et al.*, 2018). In addition to technicalities, while it is plausible that these residual nucleosomes have the potential to mediate paternal memory and therefore act as a carrier of inter- and transgeneration epigenetic information in

the mammalian male germline, several counter-arguments have to be considered.

Firstly, the question of penetrance and the propagation of specific epigenetic marks across developmental time-scales in various tissues with divergent functions. As detailed above, two successive rounds of epigenetic reprogramming occur in the embryonic germline – first directly following fertilization, and then a few days later in PGCs. It is therefore unlikely that any later developing tissues of the embryo (F1) or even future generations of offspring from the developed embryo (transgenerational epigenetic inheritance) could “remember” its original paternal chromatin marks or DNA cytosine methylation. Indeed, recent genome-wide sequencing of DNA cytosine methylation, H3K4me3 and H3K27me3 in sperm, oocyte, and allelically resolved pre-implantation embryos suggest extensive genome-wide erasure and remodeling of these epigenetic marks quickly following fertilization (Wu *et al.*, 2016; Zhang *et al.*, 2016; Zheng *et al.*, 2016).

In addition, it is unclear how changes in the paternal environment could directly shape DNA cytosine methylation and histone modifications in the post-natal testis. First, as detailed above, *de novo* epigenetic marks are guided by pre-pachytene piRNAs, which essentially act as safeguards to prevent spermatocytes that have undergone inappropriate epigenetic silencing of TEs from continuing into spermiogenesis (Zamudio *et al.*, 2015). Next, during and following meiosis II, spurious transcription occurs genome-wide because of or in order to facilitate the histone to protamine transition. It is widely known that

histone replacement occurs most often over transcribed promoters (Dion *et al.*, 2007). It is therefore more likely that un-replaced histones reside in transcriptionally silent regions (*i.e.* gene desserts). In addition, even if histone marks can be retained at developmental promoters in sperm (Siklenka *et al.*, 2015b), it is hard to envision environmental cues that would allow for specific targeting of precise histone marks. For example, if paternal diet leads to a global increase or decrease in methylation metabolites such as SAM or α-ketoglutarate in the testis, there is no clear mechanism of how this would specifically affect H3K4 vs. H3K9 vs. H3K29 methylation, which would have conflicting roles in gene regulation in the developing sperm and in embryonic development.

Taken together, while a sperm histone code remains an attractive mechanistic model for paternal epigenetic inheritance, with convincing evidence backing up the basic requirements of this model (Hammoud, David A Nix, *et al.*, 2009; Siklenka *et al.*, 2015b), much work remains to address how specific environmental cues can fine tune and/or evade the sledgehammer that is embryonic and germline epigenetic reprogramming.

The Case for Sperm-borne Small-RNAs

After spermiogenesis, testicular sperm release into a long convoluted tubule known as the epididymis (Cornwall, 2009). Testicular sperm are transcriptionally inert and morphologically comparable to mature sperm, yet are nonfunctional gametes. It is during epididymal transit that spermatozoa undergo maturation, and acquire progressive motility and the ability to fertilize the oocyte.

The epididymis is grossly separated into three regions – the caput (head), corpus (body), and the cauda (tail). The caput and corpus carry out sperm maturation events, while the cauda mainly acts as a storage site (Robaire, Hinton and Orgebin-Crist, 2002). The epididymal lumen is a highly proteinaceous and metabolically active environment that subtly balances the pH, osmolarity, and secretory proteins that reshape the sperm proteome (Robaire, Hinton and Orgebin-Crist, 2002). The caput epididymis quickly resorbs all testicular proteins and secrets its own contents of proteins into the lumen. Of particular importance are proteins involved in not only sperm motility but also interactions with the zona pellucida of the oocyte, which are also delivered via membranous vesicles containing these proteins called epididymosomes (Yanagimachi *et al.*, 1985; Sullivan, 2015). Originating from the epididymal epithelial cells via apocrine secretion (van Niel, D'Angelo and Raposo, 2018a), these apical blebs bind to and modify the sperm proteome and supports it post-ejaculatory functions (Sullivan, Frenette and Girouard, 2007; Sullivan and Saez, 2013).

Recent studies have identified small RNAs as novel carriers of sperm epigenetic information in mammals (Rassoulzadegan *et al.*, 2006; Peng *et al.*, 2012a; Kiani *et al.*, 2013; Gapp *et al.*, 2014b; Chen, M Yan, *et al.*, 2016; Grandjean *et al.*, 2016; Sharma *et al.*, 2016, 2018; Guo *et al.*, 2017; Conine *et al.*, 2018; Zhang *et al.*, 2018), which has been a central carrier of germline epigenetic information in essentially all intensively studied organisms (Castel and Martienssen, 2013; Miska and Ferguson-Smith, 2016). Small RNAs implicated in male germline inheritance and early embryonic development in various models

include miRNAs (Rassoulzadegan *et al.*, 2006; Kiani *et al.*, 2013; Grandjean *et al.*, 2016; Conine *et al.*, 2018), tRNA fragments or tRFs (Chen, M Yan, *et al.*, 2016; Sharma *et al.*, 2016; Zhang *et al.*, 2018) and long-noncoding RNAs (Gapp *et al.*, 2018). However, other small and long RNAs can potentially mediate paternal information as well, as ribosomal RNA fragments (also abundant in mature sperm, unpublished data) (see Appendix I) have been shown to mediate cellular proliferation (Chen *et al.*, 2017).

Importantly, epididymosomes also contain abundant small RNAs (Belleannée *et al.*, 2013; Sharma *et al.*, 2016, 2018). Work from our lab has shown that these small RNAs are capable of being trafficked to epididymal sperm via epididymosomes (Sharma *et al.*, 2018). tRFs make up the majority of small RNAs gained during epididymal transit, although earlier figures of around 80-90% of total caudal sperm small-RNAs may be higher than what is actually gained (unpublished data). Nevertheless, these tRFs and other small RNAs overwhelm the initial pool of piRNAs, and it has been shown that tRFs carried by sperm can play important roles in early embryonic development (Chen, M Yan, *et al.*, 2016; Sharma *et al.*, 2016). In addition, the miRNAs that are gained during epididymal transit, possibly via epididymosome fusion or further processing during this period (Yuan *et al.*, 2016), regulate embryonic gene expression and are required for proper pre- and peri-implantation embryonic development (Conine *et al.*, 2018). Taken together, these studies support the idea that the sperm epigenome can be shaped once it exits the testis, while it is trafficked through the epididymis. Given that the epididymis is a highly metabolically active epithelial tissue, it could

potentially be a privileged environmental sensor that can modify the sperm epigenome, leading to inter- or transgenerational epigenetic inheritance (Sharma *et al.*, 2016).

Intriguingly, there is precedent in plants and worms for the modification of the germline epigenome by somatic tissue (Bourc'his and Voinnet, 2010; Castel and Martienssen, 2013). In plants, germline cells arise from somatic stem cells, and selective epigenetic reprogramming must occur particularly over TEs. *Arabidopsis* male gametophytes contain a supportive vegetative nucleus that expresses transposons, and can deliver small-RNAs antisense to these transposons to the two sperm nuclei to silence TE expression there (Slotkin *et al.*, 2009). In addition, *C. elegans* injected with or ingesting double stranded RNA leads to potent gene silencing for up to five generations (Fire *et al.*, 1998). While these organisms and their germline development are distinct from mammals, their existence suggests an evolutionary need for germline epigenetic pruning by somatic tissue, apparently in relation to silencing of transposons.

Functionally, it is unclear how epithelial cells in the epididymis may sense the environment, and/or sort small RNAs and proteins into epididymosomes. In other words, no signaling pathways have been identified that can cause specific epigenetic changes in sperm (whether it be small RNAs or histones). However, it is known that exosomal contents are fine-tuned via various intracellular trafficking processes, and that their contents differ starkly in normal and pathological states, and across cell types (van Niel, D'Angelo and Raposo, 2018b). It has also been shown that certain RNA binding proteins are specifically involved in sorting of

certain miRNAs via binding to cognate motifs (Villarroya-Beltri *et al.*, 2013). Therefore, it is plausible that a possible pathway for stress sensing and epididymosomal sorting exists. Lastly, the ribonuclease(s) responsible for tRNA fragment biogenesis in the epididymis is unknown. The RNase A family, of which Angiogenin is a part of (thought to be responsible for most tRNA fragment cleavage, see below), has undergone a huge expansion in mammals, with humans carrying 15 distinct RNase genes, 13 of which are predicted to be functional (Cho, Beintema and Zhang, 2005). These genes are located on chromosomal clusters. However, ANG (RNase 5) is not expressed in the epididymis, and instead RNases 4, 9-13 are highly expressed (Li *et al.*, 2010), although 9-13 do not appear to have intact RNase activity (Castella *et al.*, 2004; Cho, Beintema and Zhang, 2005; Cheng *et al.*, 2009). Intriguingly, RNase 9 itself appears to play a role in epididymal sperm maturation, despite its loss of catalytic activity (Westmuckett *et al.*, 2014). Much work remains to be done to understand the biogenesis and sorting of small RNAs into epididymosomes, and whether and how this tissue could potentially sense metabolic changes that could lead to modifications of the sperm epigenome.

Forms and Functions of tRNA Fragments

Of particular interest in the sperm epigenome are the large abundance of tRNA fragments, or tRFs, and their potential regulatory roles in the early embryo. tRFs represent a relative new-comer to the small-RNA (sRNA) field, while the three main classes – microRNAs (miRNAs), Piwi-interacting RNAs (piRNAs), and

small-interfering RNAs (siRNAs) have been extensively characterized with well-established functions (Ghildiyal and Zamore, 2009). These sRNAs all associate with the ancient Argonaute family proteins, and regulate gene expression and development in eukaryotes (Stefani and Slack, 2008). In contrast, while certain species of tRFs, such as those derived from the 3' end via cleavage at the T loop, appear to be cleaved by Dicer and associate with Argonaute family proteins (Haussecker *et al.*, 2010; Mary T. Couvillion *et al.*, 2012; Kumar *et al.*, 2014a; Andrea J. Schorn *et al.*, 2017; Martinez, Choudury and Slotkin, 2017), the majority of tRFs described in the literature display a different biogenesis and functional pathway. Perhaps because of this potential functional diversity, and intrinsic high copy number, studies on tRFs have expanded in the last decade (Keam and Hutvagner, 2015).

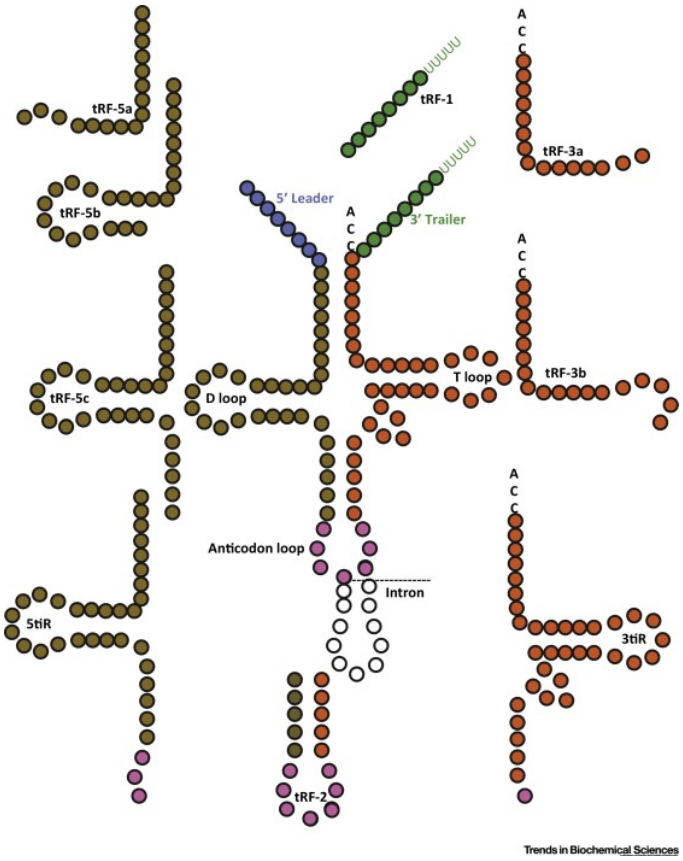


Figure 1.6. Classification of tRNA-Derived Fragments (tRFs). Fragments from tRNAs color coded to indicate area of origin. White circles indicate an intron present in some tRNAs (e.g., chr16.trna4-ProAGG) that are normally spliced out by TSEN and CLP1. *From (Kumar, Kuscu and Dutta, 2016). Reprinted with permission from Elsevier.*

Pre-tRNA fragments

tRNAs are heavily modified post-transcriptionally, including splicing, cleavage of leader and trailer sequences, addition of non-templated 'CCA' sequence at the 3' end, and chemical modifications (Torres, Batlle and Ribas de Pouplana, 2014). Consequently, a kaleidoscope of tRNA fragments has been found to be generated from tRNAs at various stages of their life cycle (**Fig. 1.6**). Following transcription of tRNAs by RNA Polymerase III, leader and trailer sequences are cleaved by RNase P and Z respectively, and a non-templated

CCA is added to the 3’ end of the trailer-free tRNA by tRNA nucleotidyl transferase (Aebi *et al.*, 1990). Lee and colleagues identified such a trailer derived 19nt tRF they called tRF-1001, derived from pre-tRNA$^{Ser\text{-}TGA}$ (Lee *et al.*, 2009). Found in high levels in proliferative cancer cells, knockdown of tRF-1001 impaired G2-M transition. The same tRF was identified by Haussecker and colleagues, which they reported to be associated with AGO3 and 4 (Haussecker *et al.*, 2010). However, they did not find inhibitory activity of this tRF on a reporter complementary to tRF-1001, in contrast to the presumed role of AGO-associated sRNAs in gene silencing. The exact function of these trailer tRFs derived from pre-tRNAs remains unknown.

Certain species of tRNAs also contain introns that need to be spliced out. In vertebrates, a kinase called CLP1 plays a key role in the tRNA endonuclease complex (TSEN) that catalyzes tRNA intron splicing (Schaffer *et al.*, 2014). Mutations in CLP1 causes neurodegeneration in zebrafish and mouse models, and humans (Hanada *et al.*, 2013; Karaca *et al.*, 2014; Schaffer *et al.*, 2014). Loss of efficient pre-tRNA splicing led to an accumulation of linear introns and tRFs derived from these introns, particularly from the tyrosine pre-tRNA (Hanada *et al.*, 2013). The authors also showed that these tRFs sensitize cells to oxidative stress induced p53 activation, and p53-dependent cell death. However, the exact signal that leads tRFs to induce p53-sensitized apoptosis is unclear.

Short miRNA-like 3’ tRFs

A myriad of tRFs are formed from mature tRNAs. Numerous studies have focused on tRFs derived from the 3' end, after the T-loop of the tRNA, with sizes around 18-22 nt, also known as tRF-3s (**Fig. 1.6**). The reported biogenesis and function of tRF-3's are mixed in the literature. While most studies show that tRF-3s are Dicer dependent (Babiarz *et al.*, 2008; Cole *et al.*, 2009; Yeung *et al.*, 2009; Couvillion, Sachidanandam and Collins, 2010; Maute *et al.*, 2013), others show these tRFs to be Dicer-independent (Li *et al.*, 2012; Kuscu *et al.*, 2018). In terms of function, tRF-3s play different roles in various biological contexts. For example, a growth essential *Tetrahymena* Piwi requires tRF-3s to properly recruit a nuclear exonuclease Xrn2 to stimulate its activity (Couvillion, Sachidanandam and Collins, 2010; Mary T. Couvillion *et al.*, 2012). tRF-3s also appear to stimulate cancer growth opportunistically: a 22-base tRF-3 from tRNA-Gly-GCC associates with AGO proteins to repress RPA1 in B Cell lymphomas (Maute *et al.*, 2013); Dicer-independent tRF-3s derived from tRNA-Leu can repress expression of a variety of genes in HEK293 cells, one of which specifically derived from tRNA-Leu-CAG, binds to ribosomal mRNAs to enhance translation (Kim *et al.*, 2017). Further work is needed to explore a unifying mechanistic understanding of tRF-3s.

5'tRNA halves and others

Another abundant and well-characterized class of tRFs, called tRF-halves or tiRs, are mostly cleaved in the anti-codon loop by RNase A superfamily related endonucleases such as Angiogenin (Yamasaki, Ivanov, G. F. Hu, *et al.*, 2009;

Emara *et al.*, 2010; Li *et al.*, 2012; Honda *et al.*, 2015), with longer lengths between 28-40 nt. These tRFs have been implicated in numerous contexts as well, including in phloem sap (Zhang, Sun and Kragler, 2009a), the archae *Haloferax volcanii* (Gebetsberger *et al.*, 2012b), and *Trypanosoma brucei* (Fricker *et al.*, 2019), although they are mostly associated with stress or in hormone-responsive tissues. Honda and colleagues demonstrated that in breast and prostate cancer cell lines, generation of tRF-halves involved in cellular proliferation are sex-hormone responsive and dependent on ANG, with expected cleavage ends of cyclic-phosphate on the 5'-tRFs and amino-acids on the 3'-tRFs (Honda *et al.*, 2015). Others show tRF-halves induced upon a diverse range of stress conditions, including etoposide and caffeine treatment, nutritional deficiency, hypoxia and oxidative stress, heat-shock and hypothermia, UV irradiation, and arsenite-stress (Fu *et al.*, 2009b; Yamasaki, Ivanov, G. F. Hu, *et al.*, 2009; Emara *et al.*, 2010; Ivanov, Mohamed M. Emara, *et al.*, 2011; Gebetsberger *et al.*, 2017a). A consensus emerges that during stress, ANG-induced tRNA cleavage occurs by translocation of ANG into the cytoplasm (Pizzo *et al.*, 2013) or via dissociation from its inhibitor ribonuclease/angiogenin inhibitor 1 (RNH1) (Thompson and Parker, 2009). How a variety of cellular stresses in various organisms can lead to dissociation of ANG from RNH1 or translocation of ANG is unknown. One explanation is based on the formation of stress granules, which are conserved RNP granules formed from pools of stalled untranslating mRNPs (Protter and Parker, 2016). First and most straightforwardly, the degradation of tRFs could potentiate translational inhibition by simply reducing

the pool of tRNAs. In addition, tRF halves could potentially act as a trigger to nucleate stress granules, leading to wide-spread translational inhibition observed in numerous cell types and organisms (Zhang, Sun and Kragler, 2009a; Ivanov, Mohamed M. Emara, *et al.*, 2011; Gebetsberger *et al.*, 2012b; Goncalves *et al.*, 2016). Here, the mechanism of how tRFs inhibit translation has been well worked out for a specific tRF derived from 5' half of tRNA-Ala and 5' half of tRNA-Cys. 5'tiRNAs derived from tRNA-Ala and tRNA-Cys specifically displace eIF4G/A from uncapped and capped mRNAs, and eIF4F from the $m7^G$ cap (Ivanov, Mohamed M. Emara, *et al.*, 2011). This displacement apparently occurs through binding of 5'tiRNA-Ala to YB-1/YBX1, a translational repressor known to displace eIF4G from RNA and eiF4F from $m7^G$ cap (Evdokimova *et al.*, 2001; Nekrasov *et al.*, 2003). This interaction could potentially be mediated by the G-quadruplex structure that 5' tRF-Ala and tRF-Cys can form (Ivanov *et al.*, 2014), although how widespread this capability of tRFs to form this structure is an open question, as is the *in vivo* validity of RNA G-quadruplex formation (Guo and Bartel, 2016).

Nevertheless, in terms of evidence supporting YBX-1 mediated tRF-regulation of cellular proliferation, Goodarzi and colleagues also identified a role for this particular interaction in breast cancer cell proliferation. However, their CLIP-seq data suggests that YBX-1 in breast cancer cells predominantly interact with a tRF-species that has not been identified elsewhere, derived from the center of the tRNA around the anti-codon loop, also known as tRF-2s (**Fig. 1.6**) (Goodarzi, Liu, Hoang C.B. Nguyen, *et al.*, 2015a). The biogenesis of tRF-2s is unknown, although they appear to be products resulting from cleavage at the T-

and D-loops simultaneously. Previous evidence suggests that these cleavages are Dicer dependent, although those authors did not detect tRF-2s (Babiarz *et al.*, 2008; Cole *et al.*, 2009). In any case, these YBX1-bound tRF-2s also appear to be stress-induced products, as it is common for cancer cells to experience hypoxia, and they act as natural tumor-suppressors through their interaction with YBX-1 leading to 3'UTR control of oncogenic target RNA expression. Together, the authors propose these tRF-2s to act through a "sponge"-like mechanism, soaking up YBX-1 which would otherwise be bound to oncogenic mRNAs to facilitate their mRNA stability and translation (Goodarzi, Liu, Hoang C.B. Nguyen, *et al.*, 2015a). It remains unclear however, how prevalent are tRF-2s in these cancer cells to effectively act as sponges, since the authors present no quantitative measurements such as northern blots or competition assays between tRFs and oncogenic mRNAs. It is also unclear how prevalent tRF-2s are in general, since it has not been observed in any other deep-sequencing study of tRFs.

Taken together, Angiogenin-dependent tRF-halves and those bound to YBX-1 appear to prevent cellular proliferation in cancer cells and during stress. Specifically, stress granules in part triggered by tRFs appear to have protective effects on cells in numerous cellular and disease contexts, and in a variety of organisms, possibly through inhibition of translation to prevent exhaustion of metabolites during stress. Indeed, this paradigm appears to hold true even in normal cellular contexts, as ANG-dependent tRFs appear to promote

hematopoietic stem cells quiescence and maintenance of cell fate through repression of translation (Goncalves *et al.*, 2016).

Interestingly, tRF-halves appear to have more diverse roles than translational inhibition in one specific cell type – the embryo. Recently, three separate studies confirmed that mammalian sperm contain very high levels of tRF-halves, in contrast to other tissue types which contain mostly miRNAs, or piRNAs in the testis for instance (Peng *et al.*, 2012a; Chen, Menghong Yan, *et al.*, 2016; Sharma *et al.*, 2016). Their particular function in sperm appears to be that of epigenetic inheritance – the levels of specific tRFs, particularly those derived from tRF-Glycine, Lysine, and Histidine, are consistently modulated in fathers fed a low-protein diet (Sharma *et al.*, 2016). One particular tRF-half, derived from tRNA-Gly-GCC, regulates the expression of a set of genes in the embryo associated with the endogenous retroelement MERVL (Sharma *et al.*, 2016), while larger gene expression changes are associated with modulation of the complete set of diet-responsive tRFs (Chen, Menghong Yan, *et al.*, 2016). Both groups show that injection of tRF-sized RNAs can induce the diet-associated phenotypic changes in the embryo, although Chen *et al.* goes further by showing these tRFs can actually lead to the same changes in offspring development observed through IVF.

The mechanism by which tRFs may act to regulate gene expression in the embryo, and how this can lead to changes in offspring development is still an open question. Our work did show that at least for tRF-Gly-GCC regulation of MERVL, translational inhibition does not appear to be the mechanism

(elaborated in Chapter II). However, this does not mean that the other abundant tRF halves do not impose translational limitation of the embryo, as ribosomal gene expression is a major gene expression signature for both data sets from both groups. In addition, the biogenesis of these tRFs in sperm is unknown. While we show that these tRFs are delivered via epididymosomes as sperm transit through the epididymis (Chapter II) (Sharma *et al.*, 2016, 2018), we do not know the detailed pathway by which the epididymis may sense dietary changes and generate changes in specific tRFs, or even the protein responsible for cleavage – the tRF-halves identified do show characteristics of cleavage by RNase A family members, however.

Intriguingly, the unique functions of tRF-5s in reproductive tissue is also observed for pollen in *Arabidopsis* and land plants, although these pollen tRF-5s are not halves but instead cleaved by Dicer in the T-loop with an average length of 19 nt, and act through AGO1 (Martinez, Choudury and Slotkin, 2017). Their function here in pollen, as in mammalian embryos, is also to silence TEs, albeit through unconventional mechanism particularly evident during relaxed epigenetic silencing conditions of loss of DDM1 SWI/SNF chromatin remodeler family protein. An important open question is therefore, whether tRF-Gly-GCC silences MERVL in a conserved mechanism in mammals.

The role of tRNA modifications in the generation of tRFs

A potential way that tRF-cleavage is modulated, whether it be paternal diet during sperm maturation or under specific stress conditions, is through changes

in the levels of tRNA modifications (Schaefer *et al.*, 2010; Tuorto *et al.*, 2012; Durdevic *et al.*, 2013; Blanco *et al.*, 2014; Guzzi *et al.*, 2018; Zhang *et al.*, 2018). Schaefer and colleagues were the first to demonstrate the role of tRNA modifications in tRF biogenesis (Schaefer *et al.*, 2010). Specifically, they show that cystone-5 methylation (m^5C) of tRNA-Gly-GCC, tRNA-Val-AAC, and tRNA-Asp-GTC is DNMT2 dependent, and protects them from cleavage by ANG during stress in *Drosophila* S2 cells and mouse embryonic fibroblasts. However, the functional relevance of m^5C protection against recombinant ANG is questionable – ANG is a mammalian member of the RNase A superfamily, and its overexpression in whole flies led to severe developmental defects (Genenncher *et al.*, 2018). Furthermore, overexpression of the *Drosophila* RNase family member RNase X25 did not affect tRF cleavage (Ambrosio *et al.*, 2014; Genenncher *et al.*, 2018). Importantly, DNMT2 or NSUN2 knockouts in whole adult flies did not result in increased tRFs, with NSUN2 mutants actually resulting in a *decrease* of tRFs in whole flies (Genenncher *et al.*, 2018).

The function of DNMT2 as a methyltransferase of tRNA m^5C is conserved in mice, as demonstrated by Zhang and colleagues, who in a follow-up to Chen *et al.,* identified DNMT2 as a key modulator of tRNA m^5C and m^2G modifications of Gly-GCC, Asp-GTC and Val-AAC (Zhang *et al.*, 2018). This follows on previous study from another epigenetic inheritance model of the *Kit* gene, where miRNAs in sperm were implicated, which also required DNMT2 (Kiani *et al.*, 2013). In terms of sperm-borne epigenetic signaling, loss of DNMT2 actually abolished the inheritance of high-fat diet induced phenotype in the offspring,

which the authors show is dependent on changes in the small RNAs fraction of sperm-borne RNAs. In addition, Zhang and colleagues show that cleavage of tRNAs carrying m^5C *in vivo* was dependent on DNMT2, although the results demonstrating RNase cleavage of oligos carrying m^5C at the DNMT2-dependent tRNA position *in vitro* was inconclusive at best (Zhang *et al.*, 2018). Furthermore, the data does not directly support DNMT2-dependent m^5C protection of tRF cleavage as a trigger of the downstream developmental phenotype induced by high fat diet, and how paternal high fat diet could lead to down-regulation of specific RNA methyltransferases or availability of metabolites involved in methylation of RNAs, and the location of where this regulation occurs, remains unclear. Perhaps in tissues that produce large amounts of tRFs, such as primary cells in the epididymis, RNA methyltransferases or methylation metabolites such as S-adenosylmethionine are depleted, leading to increased endonuclease cleavage of tRNAs into tRFs (Sharma and Rando, 2017).

On the other hand, NSUN2 has a much larger repertoire of tRNAs in mice, including tRNA-Leu-CAA, Met-CAT, Thr-TGT, Glu-CTC, and overlapping with DNMT2 targets Asp-GTC and Gly-CCC (Blanco *et al.*, 2014). *Nsun2* loss in both *Drosophila* and mice suffer translational inhibition and stress, particularly in neurons, but also in various stem cells (Tuorto *et al.*, 2012; Blanco *et al.*, 2014, 2016). Interestingly, reminiscent of DNMT2 catalyzed m^5C in tRNAs, NSUN2 catalyzed m^5C also appears to have a protective effect against ANG dependent tRF cleavage, although both *in vivo* tRF deep sequencing data and *in vitro* cleavage in of m^5C are far from convincing (Blanco *et al.*, 2014). In addition, as in

the DNMT2 studies, only circumstantial, but no direct evidence is presented demonstrating tRF-mediated effects on translational inhibition during stress response in stem cells or in neurons (Tuorto *et al.*, 2012; Blanco *et al.*, 2014, 2016). In any case, m^5C appears to help maintain tRNA stability to a certain extent, although the exact mechanism by which cleavage of various tRFs occurs and the degradation of tRFs carrying m^5C following cleavage remains unclear.

Another tRNA modification that has been studied in terms of mediating tRNA stability and tRF cleavage is pseudouridylation (Guzzi *et al.*, 2018). Here, Guzzi and colleagues implicated PUS7, a pseudouridine synthase, in the generation of tRFs derived from various tRNAs, particularly 18nt tRFs derived from the 5' end of tRNA-Ala/Cys/Val (mTOG) (Guzzi *et al.*, 2018). The authors claim that the level and function of these tRFs depends on the presence of the pseudouridine modification, and therefore on the presence of PUS7, and that loss of PUS7 lead to translational inhibition and imbalances in stem cell developmental niche. Again, as with m^5C, data supporting this claim is inconclusive at best. In addition, the particular prevalence of these tRFs is not explored, given that pseudouridylation is a widespread modification found throughout essentially every tRNA at multiple nucleotides. Nevertheless, these mTOGs were similar to those studied by Ivanov and colleagues in form and function (inhibition of translation), although those were mostly tRNA halves, but the regulatory region appears to be the same – the initial segment before the D-loop (Ivanov, Mohamed M. Emara, *et al.*, 2011). Pseudouridylation is also found in rRNAs and snoRNAs, and since these RNAs also produce sRNAs, it might be

interesting going forward to understand whether this modification and others also influence sRNA-generation from these other RNAs. This possibility appears likely since cystone-5 methylation has already been shown to determine processing of vault-RNA derived sRNAs (Hussain *et al.*, 2013), and the levels and functions of miRNAs (Squires *et al.*, 2012; Kiani *et al.*, 2013; Yuan *et al.*, 2014). In addition, given the diversity of post-transcriptional modifications on tRNAs (Lorenz, Lünse and Mörl, 2017), and the known function of these modifications to impart structural stability and translational fidelity (Chou *et al.*, 2017), important work remains to determine the potential roles that these modifications play in tRF biogenesis and function (Durdevic and Schaefer, 2013; Lyons, Fay and Ivanov, 2018).

tRF questions to consider

For tRF-3s, is its abundance in cancer altogether an existential accident, or does this abundant class of tRFs play an essential function in mammalian development? One possible answer is given by Schorn and colleagues, who showed that under relaxed epigenetic conditions, such as in *Setdb1* or *Dnmt1* knockout mouse embryonic stem cells (which the authors claim simulates early embryos or germ cells undergoing epigenetic reprogramming), tRF-3s are recruited to silence ERVs (Andrea J. Schorn *et al.*, 2017). However, the biogenesis pathway for these tRFs and effector protein involved in silencing are unclear. Nevertheless, their study argues that cellular reprogramming elicits derepression of LTR retroelements, and given the natural complementarity of

tRF-3's to the primer-binding site (PBS) of LTR ERVs, cells utilize them as esiRNAs to control rampant retrotransposition. A similar mechanism is described in *Arabidopsis,* although the tRFs Martinez and colleagues described in pollen to silence TEs are short tRFs derived from the 5' end, or tRF-5s (Martinez, Choudury and Slotkin, 2017). Then, the implication for cancer cells or growing *Tetrahymena* being that since these cell types do not experience relaxed epigenetic silencing, their utilization of tRF-3s in silencing of TEs is unlikely to be observed. However, since certain cancers do arise from loss of epigenetic silencing, particularly in H3K9 methylation which is important for transposon silencing in mammalian cells (Nguyen *et al.*, 2002; Yang, 2004; Kondo *et al.*, 2008), these cells present an important avenue for testing of the hypothesis proposed by Schorn et al. that during relaxed epigenetic silencing cells utilize tRF-3s as esiRNAs to silence transposon expression.

Another important issue to be resolved is whether all tRF-3s in the literature are derived from mature tRNAs (whereby they would end in a CCA), or pre-tRNAs, since both classes make up large fractions of total tRF-3s (see Appendix I). In terms of tRF biogenesis, certain species of tRF-5/3s may also be more amenable to Dicer cleavage under certain physiological conditions or in certain cell types, although how exactly this could occur is unknown.

One barrier to a comprehensive understanding of tRFs in various biological contexts is the current method of identifying and quantifying tRFs – deep sequencing, is not consistently applied across laboratories. Important recent work has shown that modifications and changes in cloning methodology

can impact whether the tRFs are clonable, and therefore, assumed to be "present" in the sample by standard sRNA deep sequencing methods (Cozen *et al.*, 2015) (see also Appendix 1). Future work will need to critically filter, clarify, and consolidate tRF deep sequencing data from various groups with the cloning strategy and mapping methodology in mind.

Regulation of Endogenous Retrovirus Family ERV-L

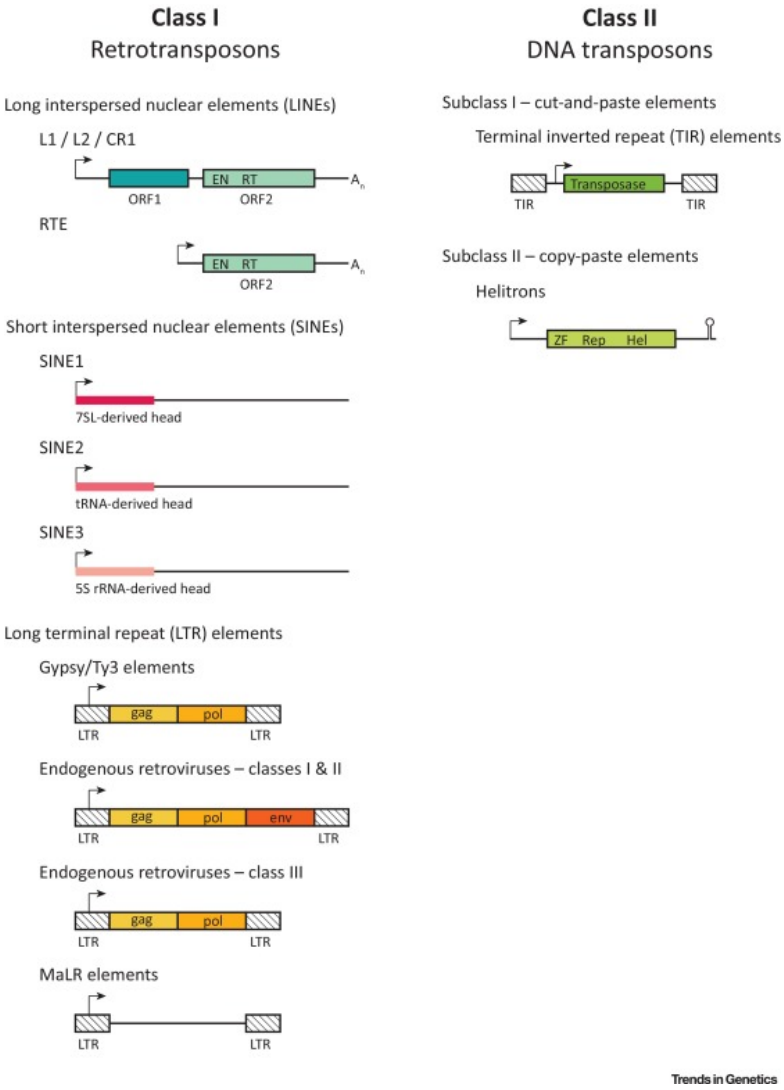


Figure 1.7. Schematic representations of the different transposable element groups found in mammalian genomes. Colored boxes demarcate coding

regions and the protein domains encoded within. *From (Rodriguez-Terrones and Torres-Padilla, 2018). Reprinted with permission from Elsevier.*

There are two main types of TEs – type I DNA transposons, and type II retrotransposons, which are then divided into classes of non-LTR retrotransposons such as LINE-1 and SINE-1 superfamilies, and LTR retrotransposons such as IAP and *Gypsy* superfamilies (**Figure 1.7**) (Kapitonov and Jurka, 2008). Together, TEs have been highly successful in shaping eukaryotic genomes. For example, 40% of the mouse genomes are fossils of TEs, with about 10% derived from LTR retrotransposons, also known as endogenous retroviruses or ERVs (Consortium, 2002). ERVs appear to be remnants of ancient viral infections of the germline that have now become endogenous threats to genome stability. Despite the evolution of potent host silencing mechanisms, ERVs continue to win the evolutionary arms race – ERVs remain highly active in mice both in terms of expression and retrotransposition, and continue to shape genomes in mostly spontaneous mutations. These lead to a variety of negative functional consequences, including direct gene mutation, repeat-driven non-allelic recombination leading to translocations, expression mis-regulation, and post-transcriptional processing defects (Kazazian, 2004; Slotkin and Martienssen, 2007a; Garcia-Perez, Widmann and Adams, 2016). In certain cases however, these novel transpositions can lead to genomic innovation (Sela *et al.*, 2007; Biémont and Voytas, 2010; Dubin, Mittelsten Scheid and Becker, 2018), and can even take on new regulatory roles during development (Jachowicz *et al.*, 2017; Percharde *et al.*, 2018). Here, I specifically focus on

introducing the silencing mechanisms of ERVs in mammals, and evidence of ERVs acquiring novel regulatory functions in mammalian development.

Transcriptional silencing

Potent silencing mechanisms exist in the nucleus, where various epigenetic machineries act in tandem to silence expression of ERVs. DNA methylation and chromatin state are two highly conserved mechanisms that act cooperatively to silence retrotransposon expression in all mammalian cell types (Yoder, Walsh and Bestor, 1997). CpGs are more than 70% methylated throughout the genome of mammalian somatic tissue, with half of all human CpGs residing in repeats (Rollins *et al.*, 2006). How exactly DNA cytosine methylation represses transcription is unclear however, but has been suggested to occur through two separate mechanisms – direct blocking of transcription factor binding, and recruitment of methyl-CpG binding MBD proteins and repressor complexes such as the NuRD complex (Li and Zhang, 2014). During preimplantation development and germline specification – where active and passive DNA cytosine demethylation leads to loss of genomic silencing capacity (Li and Zhang, 2014) – chromatin appears to take the lead (Walsh, Chaillet and Bestor, 1998; Messerschmidt, Knowles and Solter, 2014). In these tissues, ERVs are recognized and marked by repressive chromatin marks through a coordinated dance of epigenetic machineries. First, one of hundreds of KZFPs recognize specific sequences in TEs, particularly ERVs, with some KZFPs recognizing multiple classes of ERVs (Imbeault, Helleboid and Trono, 2017).

Importantly, KZFPs are rapidly evolving and appear to bind to evolutionarily young ERVs (Jacobs *et al.*, 2014). The binding of KZFPs then leads to recruitment of the transcriptional regulator TRIM28 (KAP1) (Wolf and Goff, 2007, 2009; Rowe *et al.*, 2010), and associated methyltransferases of H3K9 (such as SET domain containing proteins G9A/EHMT2, SETDB1, and SUV39H) (Schultz *et al.*, 2002; Matsui *et al.*, 2010; Maksakova *et al.*, 2013; Liu *et al.*, 2014; Yeung *et al.*, 2019). KZFPs also recruit *de novo* DNA cytosine methyltransferases (DNMT3A/B/L) (Walsh, Chaillet and Bestor, 1998; Liang *et al.*, 2002; Quenneville *et al.*, 2012; Turelli *et al.*, 2014), and these epigenetic machineries together drive heterochromatin formation by binding of HP1 (Lachner *et al.*, 2001), finally leading to silencing of ERV transcription.

Within this system, H3K9me2/3 appears to mark and repress specific subsets of ERVs, with H3K9me2 preferentially acting to repress class III ERVLs, and H3K9me3 marking class II ERVKs (Maksakova *et al.*, 2013; Walter *et al.*, 2016). This epigenetic place holding potentially demarcates the evolutionary age of LTR-ERV classes, and whether KZFPs have evolved to specifically recognize and recruit H3K9me3 machinery to the genomic region. Interactions between various chromatin-associated factors and their associated marks is crucial, as depletion of histone lysine demethylase KDM1a (Macfarlan *et al.*, 2011), histone deacetylase HDAC1 (Reichmann *et al.*, 2012), and SUMO (Yang *et al.*, 2015; Cossec *et al.*, 2018) all lead to derepression of various classes of ERVs. The various histone marks also compensate for each other following loss of specific marks over certain classes of ERVs (Walter *et al.*, 2016). Last but not least,

chromatin building blocks play a major role in repression of ERV expression, as ERVs appear to exploit both globally under assembled chromatin (Ishiuchi *et al.*, 2015b; Yang *et al.*, 2015), and even reductions of a specific histone variant H3.3 (Elsässer *et al.*, 2015). Taken together, the mammalian nucleus contains a dense network of transcriptional repressors centered on the packaging of chromatin and methylation of DNA to suppress ERV expression, ranging from the specific (KZFPs) to the general (histone levels). Given that histones (and their variants) also appear to be general repressors of zygotic genome activation (Almouzni and Wolffe, 1995; Amodeo *et al.*, 2015; Joseph *et al.*, 2017a), a large proportion of which involves transcription of ERVs (Peaston, A. V Evsikov, *et al.*, 2004; Macfarlan *et al.*, 2012a), whether expression of ERVs during early embryonic development is consequential or essential remains a fascinating open question.

Post-transcriptional silencing

Despite extensive transcriptional silencing, certain classes of ERVs remain highly expressed, particularly during early embryogenesis and germline specification. Paradoxically, this expression leads to their successful silencing by small-RNAs, as in plants, where transcription is necessary for small-RNA mediated *de novo* DNA methylation (Law and Jacobsen, 2010). Perhaps this system allows for more efficient targeting of the highly polymorphic families of retroelements by relying on their own sequences to target gene-silencing machinery, instead of solely relying on the evolution of KZFPs to chase down

rapidly evolving ERV sequences and possibly new invasions of exogenous retroviruses.

As mentioned above, the post-transcriptional silencing pathway for ERVs during germline specification involves pre-pachytene piRNAs that target expressed ERVs via binding to slicers MILI, resulting in the ping-pong cycle that enriches for highly expressed ERVs, such as IAPs, which in turn targets MIWI2 in the nucleus for epigenetic silencing of genomic ERV sequences (Aravin *et al.*, 2007, 2008; Kuramochi-Miyagawa *et al.*, 2008).

While the piRNA system functions well in male primordial germ cells, it is absent in the oocyte and early embryonic development. Instead, endogenous siRNAs generated by Dicer and loaded onto AGO2 appear to be the key silencer of ERVs during oogenesis and early embryonic development (Svoboda *et al.*, 2004; Murchison *et al.*, 2007; Tam *et al.*, 2008; Watanabe *et al.*, 2008), while miRNAs instead appear to be subdued during these two stages of development (Ma *et al.*, 2010; Suh *et al.*, 2010).

During early embryonic development, ERVs become severely derepressed (Peaston, A. Evsikov, *et al.*, 2004; Macfarlan *et al.*, 2012a). Surprisingly, the expression of these ERVs does not appear to be detrimental, but are instead prominent markers of early cleavages, and help orchestrate synchronous gene activation (Göke *et al.*, 2015; Franke *et al.*, 2017) and chromatin organization (Jachowicz *et al.*, 2017). Well studied examples include ERVL activation during embryonic genome activation (EGA) in mouse and human embryos (De Iaco *et al.*, 2017; Hendrickson *et al.*, 2017), and LINE1

elements marking 4-8 cell transition in mouse (Jachowicz *et al.*, 2017), which apparently also act in part to silence the MERVL transcription program (Percharde *et al.*, 2018). The silencing of these ERVs, both transcriptionally and post-transcriptionally, is not well understood in the early embryo. For example, in terms of MERVL silencing after the 8-cell stage, there is evidence that silencing is partially mediated by RNAi via *Dicer* (Svoboda *et al.*, 2004), and LINE-1 itself recruits KAP1/TRIM28 to mediate transcriptional silencing (Percharde *et al.*, 2018). However, these programs appear insufficient to fully suppress MERVL expression, since each pathway only partially represses the expression of MERVL. In addition, the mechanism that is deployed to silence other ERVs, including LINE-1 itself, and IAP elements, remains obscure.

tRNA-fragment based silencing of ERVs in early embryos?

Could there be alternative pathways to silencing of ERVs during early embryonic development? Interestingly, several studies have implicated tRFs in the silencing of endogenous or exogenous retroviruses. Wang and colleagues showed that respiratory syncytial virus (RSV), a non-Long Terminal Repeat (LTR) RNA virus, induced tRF levels in human A549 cells from 1.7% of normal sRNA library to 36.5% during infection (Wang *et al.*, 2013). Curiously, the majority of these tRFs were derived from the 5’ half of tRNAs (tiRs) in an ANG-dependent manner, with some, such as those of the 5’ half of Glu-CTC being up-regulated more than 400 times upon RSV infection. Surprisingly, the authors claim that tRF-GluCTC promotes RSV replication by suppressing the expression

of APOER2 through complementarity to the 3'UTR of APOER2, in a miRNA-like pathway. This result is surprisingly for many reasons: 1. None of the tRFs derived from GluCTC are miRNA length (22 nt). 2. The link between APOER2 and RSV replication is circumstantial. 3. The authors do not demonstrate how RSV infection could lead to ANG cleavage of tRNAs.

Quite separately, it has been shown that DNMT2 and NSUN2 mutants cannot repress heat-shock induced LTR-retrotransposon expression in *Drosophila* (Genenncher *et al.*, 2018). As mentioned above, the same group previously showed that loss of DNMT2 mediated m^5C leads to increased tRF biogenesis in S2 cells (Schaefer *et al.*, 2010), and independent studies linked DNMT2 and NSUN2 catalyzed m^5C in tRF biogenesis in mice (Blanco *et al.*, 2014, 2016; Zhang *et al.*, 2018), with the implication being that maybe tRFs could possibly play a role in silencing of TEs in general. However, Genenncher and colleagues actually show that in whole flies, DNMT2 and NSUN2 mutation does not influence tRF biogenesis, possibly even leading to loss of tRFs (Genenncher *et al.*, 2018), and could not demonstrate a direct link between tRF biogenesis and TE silencing. However, ERVs are known to demonstrate tissue specific expression, particular in the testis and ovary, and further study needs to focus on these tissues to definitively rule out the role of tRFs in silencing of endogenous retroviruses in *Drosophila*. Curiously, a catalytically dead mutant of DNMT2 that does not reconstitute tRNA methylation was able to repress retrotransposon expression in the presence of loss of endogenous DNMT2, suggesting that the RNA-binding activity of DNMT2 provides TE silencing. One potential mechanism

by which this could occur via tRNA quenching – implying that tRNAs are used by ERVs for expression, and the binding of DNMT2 and NSUN2 to tRNAs acts as a sink for tRNAs, in affect silencing retrotransposition and expression.

This hypothesis is plausible given the fact that tRNAs are used as a primer for reverse transcriptases of exogenous and endogenous LTR retroviruses (Marquet *et al.*, 1995). Indeed, as discussed above, Schorn and colleagues show that 18 and 22nt tRFs derived from the 3' terminal of mature tRNAs target the tRNA primer binding site (PBS) in the LTR of endogenous retroviruses to silence their retrotransposition (Andrea J. Schorn *et al.*, 2017). However, it is unclear whether inhibition of retrotransposition itself silences expression, as data from other groups in *Drosophila* nurse cells suggest no correlation between expression and retrotransposition ability of ERVs (Wang *et al.*, 2018). Along these lines, tRF-5s apparently also target transposons in plant pollen via complementarity, although whether these tRF-5s could act through their counterpart tRF-3s as in mammalian cells is an open question (Martinez, Choudury and Slotkin, 2017).

In yet another mechanism, work from our lab showed that tRF-5s derived from Gly-GCC, or tRF-GG, represses transcriptional silencing of MERVL and its target genes specifically (Sharma *et al.*, 2016). In Chapter II and III, I will go into the details of what we uncovered regarding the mechanism of tRF-GG repression of MERVL in the early embryo, but to understand this mechanism, I will first give a more in depth introduction to what we know about the regulation of MERVL expression and its function in the early embryo.

From Escapee to Functional Contributor – the story of ERVL

The first invasion of MERVL (MuERVL) was more than 80 million years ago before the split of placental mammals, while apes and mice lineages both experienced individual major expansions (Bénit *et al.*, 1999; Franke *et al.*, 2017). MERVL no longer actively retrotransposes, although it produces virus-like particles (Bénit *et al.*, 1999; Ribet *et al.*, 2008). The majority of ERVL genomic inserts are solo LTRs, accounting for 60-70% of the thousands of annotated copies throughout the mouse genome (Franke *et al.*, 2017). Murine specific ERVL LTRs are annotated MT2_Mm, and share significant homology to MaLR and MLT LTRs, most of which are highly transcribed in mouse oocytes and early embryos (Peaston, A. Evsikov, *et al.*, 2004; Veselovska *et al.*, 2015). Human ERVL (HERVL) LTRs, annotated as MLT2A1, function much the same way, with specific activation during EGA at the eight cell stage (Hendrickson *et al.*, 2017).

Importantly, ERVL LTRs have co-opted and act as alternative promoters to regulate the EGA specific expression of >1500 in mouse and >100 genes in humans (Macfarlan *et al.*, 2012a; Franke *et al.*, 2017; Hendrickson *et al.*, 2017). In mice, these include hundreds of protein coding genes, long noncoding RNAs, and pseudogenes, potentially leading to *de novo* evolution and functionalization of these genes during embryogenesis (Franke *et al.*, 2017). The exact functions of MERVL co-opted genes are vaguely defined, with the exception of the well-studied transcription factor Zscan4, which plays a role in promoting global demethylation by inducing ubiquitination and degradation of Uhrf1 and Dnmt1

(Dan *et al.*, 2017), and maintaining chromosome stability by preventing telomere shortening (Zalzman *et al.*, 2010). Strikingly, the expression of MERVL and its co-opted genes (also referred to as MERVL target genes) collectively regulate the developmental potential of early embryos (Huang *et al.*, 2017). However, its spurious expression in adult tissue is detrimental, causing the genetic disorder facioscapulohumeral muscular dystrophy (FSHD) (Tassin *et al.*, 2013). The elucidation of ERVL target gene functions altogether remains an important future direction of study.

The regulation of ERVL and its associated genes have become hotly studied in recent years due to its association with the totipotent cell fate (Baker and Pera, 2018), the stabilization and creation of which in culture could enable more efficient generation of chimeric animals for research, and organ production for transplantation. These studies have concentrated in mouse embryonic stem cells (mESCs), where surprising, a minor fraction of these pluripotent cells spontaneously cycle into a cell state that highly expresses ERVL, and importantly, demonstrate increased developmental potency (Macfarlan *et al.*, 2012a; Ishiuchi *et al.*, 2015a).

A myriad of regulators of ERVL expression and its associated totipotent-like (or 2-cell like) state have been identified, which has also bettered our understanding of the molecular basis of totipotency itself. First, the uniquely accessible and dynamic early embryonic chromatin state permits the spurious transcription of ERVs, including ERVL elements (Boskovic *et al.*, 2014; Wu *et al.*, 2016; Schulz and Harrison, 2019). Indeed, depletion of numerous chromatin

factors involved in condensation and compactness of chromatin, such as CAF1 (Cheloufi *et al.*, 2015; Ishiuchi *et al.*, 2015a), NuRD complex members (Matsui *et al.*, 2010; Macfarlan *et al.*, 2011; Yang *et al.*, 2015), and other deacetylase and polycomb associated complexes (He *et al.*, 2019) can all lead to derepression of MERVL target genes.

Following initial exploration of gene expression, transcription of ERVL starts in earnest at the major EGA via direct binding of the ERVL LTR by DUX family transcription factors, itself a ERVL target gene in mouse (De Iaco *et al.*, 2017; Hendrickson *et al.*, 2017; Whiddon *et al.*, 2017). Other MERVL target genes also appear capable of reinforcing ERVL expression and propagation of the 2-cell like state (Iaco *et al.*, 2019). The expression of DUX itself appears to be regulated by small DNA binding proteins from the DPPA family (Huang *et al.*, 2017; Eckersley-Maslin *et al.*, 2018; Iaco *et al.*, 2018). The expression of MERVL and its target genes also coincides with genome-wide demethylation (active in the paternal and passive in the maternal genome) (Messerschmidt, Knowles and Solter, 2014), although the exact role that MERVL plays in the initiation or facilitation of this process is unclear (Eckersley-Maslin *et al.*, 2016; Dan *et al.*, 2017). The expression of ERVL and its target genes continues unabated for a few cell divisions until the 8 cell stage in mice, where they are finally silenced (Peaston, A. Evsikov, *et al.*, 2004). As mentioned above, the repression of ERVL is both transcriptional and post-transcriptional, mediated TRIM28 guided deposition of heterochromatic marks in the nucleus (Guallar *et al.*, 2012; Schoorlemmer *et al.*, 2014; Percharde *et al.*, 2018), by RNAi in the cytoplasm

(Svoboda *et al.*, 2004), and also apparently indirectly by miR34 (Choi *et al.*, 2017). Whether these silencing mechanisms are conserved in human embryos is unclear. It is also unclear whether any other nuclear processes play a role in the silencing of ERVL, since its expression disrupts chromatin organization globally in the early embryo (Wu *et al.*, 2016; Kruse *et al.*, 2019).

Cajal Body and Histone Locus Body Functions

Our study of the mechanism underlying tRF-Gly-GCC regulation of MERVL also led us to an unexpected biological process – Cajal bodies and histone biogenesis, which I will briefly introduce here.

The Cajal body (CB) is a nuclear organelle present in all eukaryotes studied so far, and is identified most reliably with the silver impregnation technique originally used by its founder, Ramon y Cajal (Gall, 2000; Nizami, Deryusheva and Gall, 2010). It is also identified by the marker protein p80 coilin, as the CB was first named coiled bodies in human and mouse cells (Monneron and Bernhard, 1969), and by CB-specific RNAs (scaRNAs). Surprisingly, the conservation of coilin is relatively low, as its orthologs have not been identified in non-vertebrate model organisms such as *S.cerevisae* and *C.elegans*, even though organelles with parallel functions exist in these organisms. In *Arabidopsis, Drosophila,* and mice, in which coilin mutations have been studied, coilin is required for CB formation, but neither coilin nor a typical CB is essential for viability (Tucker *et al.*, 2001; Jady *et al.*, 2003; Collier *et al.*, 2006; Deryusheva and Gall, 2009; Liu *et al.*, 2009). Nevertheless, coilin plays an important

developmental role in mice and presumably mammals, as knockout of coilin in mice leads to a semi-lethal phenotype, where homozygotes die as embryos in a semi-penetrant manner, and adults have fertility and fecundity defects (Tucker *et al.*, 2001). CB loss also leads to formation of “residual” bodies each with a subset of CB components (Jády, Bertrand and Kiss, 2004). Therefore, coilin appears to coalesce the formation of CB rich in dense interactions between RNAs and proteins, reminiscent of phase-separating organelles such as P-granules and nucleoli (Machyna, Heyn and Neugebauer, 2013). In turn, this “phase-separation” and therefore concentration of snRNP machinery may speed up snRNP formation (Novotný *et al.*, 2011).

For decades, the CB has been presumed to be the location of assembly and/or modification of the splicing snRNPs, since they are enriched in proteins and RNAs (U1 to U6) involved in splicing (Carmo-Fonseca *et al.*, 1992; Matera and Ward, 1993; Staněk *et al.*, 2003; Schaffert *et al.*, 2004; Staněk *et al.*, 2008). In addition, scaRNAs such as U85, closely related and identical in structure and function to snoRNAs, are complexed with Fibrillarin and dyskerin, to guide 2’-O-methylation and pseudouridylation respectively, of other snRNAs in the CB (Darzacq, 2002; Kiss *et al.*, 2002; Liu *et al.*, 2009). Another prominent component of CB is the survival motor neuron protein (SMN) (Matera and Frey, 1998; Carvalho *et al.*, 1999), which is required for proper functioning of motor neurons in the spinal cord (Miguel-Aliaga *et al.*, 2000; Chan *et al.*, 2003; Rajendra *et al.*, 2007), and also appears to be involved in assembly of snRNPs with the Sm ring (Pellizzoni, J and G., 2002; Shpargel and Matera, 2005; Battle *et al.*, 2006).

However, CB function does not appear to be limited to snRNP maturation, as proteins involved in a variety of other nuclear functions, including transcription and cell signaling, are also enriched here (Machyna, Heyn and Neugebauer, 2013). A major future challenge will be to remediate the dichotomy between seemingly essential functions of the CB and the lack of phenotype upon CB loss. Indeed, lack of CBs in *Drosophila* devoid of coilin do not suffer decreases in snRNA levels nor their modifications (Deryusheva and Gall, 2009).

In experiments designed to identify *Drosophila* CB in tissues of the fly, U7 snRNP and U85 scaRNAs surprisingly detected not one, but two independent nuclear organelles. One of these organelles was invariably associated with the locus containing all histone genes in *Drosophila* (Liu *et al.*, 2006, 2009). Subsequently named the histone locus body (HLB), these organelles are distinct from CB in mammalian cells (Bongiorno-Borbone *et al.*, 2008; P.N. Ghule *et al.*, 2008). Instead of coilin, the nucleator and marker of HLB appears to be NPAT (Ye *et al.*, 2003; White *et al.*, 2011), and the dedicated function of the HLB is the transcription and processing of replication-dependent histone pre-mRNAs (Marzluff and Koreski, 2017). Replication-dependent, or canonical, histone genes are only required as the name implies, during DNA replication to form chromatin, and therefore have to be regulated temporally. Consequently, these genes evolved multiple unique characteristics – they are intronless, and generate non-poly-adenylated pre-mRNAs associated with its own unique set of processing machinery (**Fig. 1.8**) (Marzluff and Koreski, 2017).

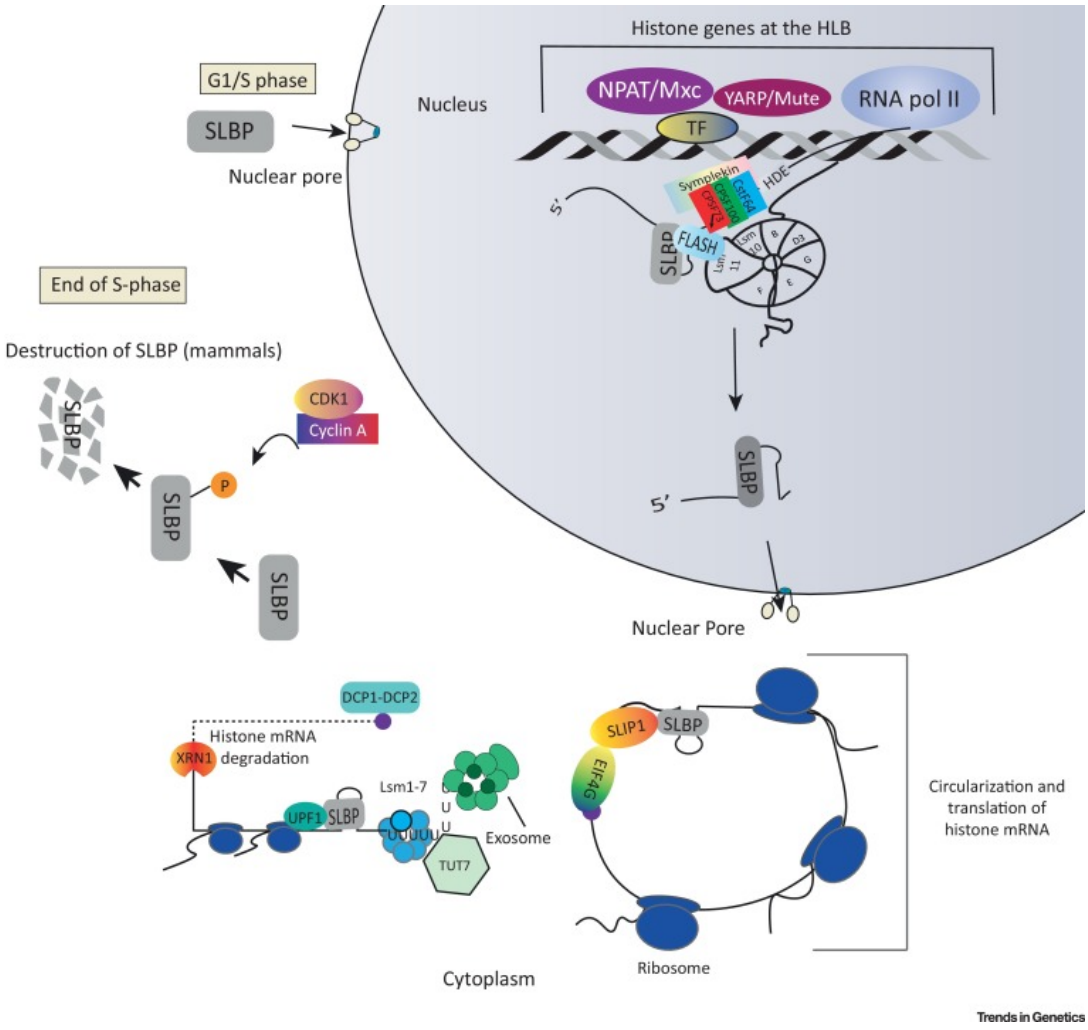


Figure 1.8. Life Cycle of Histone mRNA. A schematic of the life cycle of histone mRNAs and SLBP in mammalian cells. Note that SLBP participates in every part of the histone mRNA life cycle, and is also cell cycle-regulated. Abbreviations: SLBP, stem-loop binding protein; P, phosphorylation; Pol II, polymerase II. *From (Marzluff and Koreski, 2017). Reprinted with permission from Elsevier.*

Canonical histone pre-mRNAs carry a 3' extension containing a stem loop bound by stem-loop binding protein (SLBP), that is cleaved via guidance by the U7 snRNP, in concert with a variety of other factors, including Lsm10/11 as part of the SM ring, FLASH (Frey *et al.*, 1995; Bongiorno-Borbone *et al.*, 2008; Prachi N Ghule *et al.*, 2008; Brooks *et al.*, 2015), MUTE (Bulchand *et al.*, 2010), and the histone-cleavage complex (HCC) (Gick *et al.*, 1987; Sullivan, Steiniger and

Marzluff, 2009; Romeo, Griesbach and Schümperli, 2014). Two distinct sites within the pre-mRNA guide cleavage to form the mature 3' end, the stem-loop, and the histone-downstream element (HDE), which basepairs with U7 snRNP just 3' of the cleavage site (Strub, Galli and Bimstiel, 1984; Schaufele *et al.*, 1986; Bond, Yario and Steitz, 1991). U7 snRNA is a <70nt RNA containing a 2,2,7 trimethyl G cap (Cotten *et al.*, 1988). The core U7 snRNP is assembled in the cytoplasm, and consists of U7 snRNA bound to five Sm proteins, B, D3, E, F, and G, found in all spliceosomal snRNPs, with Lsm10 and 11 replacing the normal SmD1 and 2 proteins in the ring (Pillai *et al.*, 2001). Lsm11 binds to FLASH, which is essential for proper processing of histone pre-mRNAs (Burch *et al.*, 2011). The active form of the U7 snRNP, or the "holo" form, is assembled in the nucleus during S phase, where the core complex recruits the HCC, which consists of symplekin, CstF64, CPSF100, and CPSF73, the endonuclease that cleaves the pre-mRNA (Kolev and Steitz, 2005; Sullivan, Steiniger and Marzluff, 2009).

Surprisingly, despite the coordinated fashion by which histone genes are expressed, they lack common transcription regulatory elements. There is however, a dedicated transcription factor activated at S-phase entry by Cdk2 called NPAT, which is necessary for transcription of all replication-dependent histone genes (Ma *et al.*, 2000; Zhao *et al.*, 2000; Ye *et al.*, 2003). NPAT is also required for HLB formation, which coordinates the transcription of entire clusters of histone genes – multiple tandem repeats of each copy of canonical histone gene as in *Drosophila* and sea urchin, or jumbled clusters on separate

chromosomes as in mammals. However, NPAT is an atypical transcription factor, as it is a very large unstructured protein that does not directly bind to DNA (Terzo *et al.*, 2015). Transcriptional elements associated with individual histone genes are also found very widely in the genome, such as Oct-1 or Pou2f1 for H2b (Zheng, Roeder and Luo, 2003). A comprehensive ChIP-seq study also identified E2f, Smad, and YY1 as transcription factors that bind to specific subsets of histone genes, in a somewhat cell-cycle and cell-type dependent way (Gokhman *et al.*, 2013). Of these, E2f transcription factors are enriched in G1-S transition, and are rapidly degraded following S-phase (Marti *et al.*, 1999), and were thus proposed as a “master regulator” of histone gene expression. However, E2f knockout MEFs only showed a moderate decrease in histone gene expression (Gokhman *et al.*, 2013). NPAT does however appear to interact extensively with itself and other constitutive HLB proteins, such as FLASH, itself a large unstructured protein (Yang *et al.*, 2014; Tatomer *et al.*, 2016).

The 3’-end processing of canonical histone genes is tightly coupled to transcription in a cell-cycle dependent manner. First, transcriptional elongation factors such as NELF (Narita *et al.*, 2007) and Ars2 (Kiriyama *et al.*, 2009; Gruber *et al.*, 2012) result in misprocessing of histone mRNA. Additionally, while U7 snRNP and FLASH are constitutively present in the HLB, they interact with the HCC in a cell-cycle dependent manner (Tatomer *et al.*, 2016). This leads to activation of the U7 snRNP to drive 3’-end processing of histone pre-mRNAs. Levels of SLBP are also cell cycle regulated (Whitfield *et al.*, 2000). Together, these studies support a model where NPAT, when phosphorylated at S-phase

entry, changes conformation to allow for rapid recruitment of a range of transcription factors and processing machinery required for efficient and rigorous transcription of cell-cycle dependent histone genes (Marzluff and Koreski, 2017). Additionally, YY1 and CTCF, which have been shown to bind at the HLB at specific border regions, may act in coordinate to drive strong boundary formation to isolate the HLB domain (Gokhman *et al.*, 2013; Nora *et al.*, 2017; Weintraub *et al.*, 2017).

Of note, there is one study that links small-noncoding RNAs with histone mRNA processing (Köhn *et al.*, 2015). Y3 noncoding RNA, a RNA polymerase III transcript conserved in eukaryotes, associates with Ro60 and may plays a role in DNA replication and RNA quality control (Köhn *et al.*, 2015). Knockdown of Y3 with anti-sense oligos disrupts HLB formation and increase in misprocessed histone pre-mRNAs. The authors proposed that Y3 imposes this regulation on histone processing through its interactions with CPSF, an important pre-mRNA processing factor.

At the end of S-phase, when cells no longer require new histone production, histone mRNAs are rapidly degraded through regulation of the stem-loop (Whitfield *et al.*, 2000) and requires translation of the mRNAs (Slevin *et al.*, 2014). It appears that after nuclear processing, the mature mRNA associated with the histone 3’ exonuclease (3’hExo), which shortens the mRNA by 2-3nt, while the full length of the mRNA is restored by uridylation (Yang *et al.*, 2006; Lackey, Welch and Marzluff, 2016). When degradation is initiated by inhibition of DNA replication, 3’hExo triggers the process by removing 5-7nt into the step loop

(Hoefig *et al.*, 2013). How this process occurs in unknown. Following this chewing, 3'hExo no longer binds to the degrading mRNA, TUT7 polyuridylates the transcript (Lackey, Welch and Marzluff, 2016). This process leads to SLBP dissociation, possibly through the loss of stem-loop stability, leading to 3' to 5' degradation of the histone mRNA by the exosome (Hoefig *et al.*, 2013).

While it has been largely assumed that most of these processes regulating CBs and HLBs are conserved from *Drosophila* to mammals, the dynamics and regulation of CB and HLB in mammalian embryos remains unexplored. Given the unique cell cycle, chromatin and transcriptional mechanisms that occur in the early stages of mammalian embryonic development, and the expanding tools for genomic tools targeting miniscule sample sizes, this body of research is ripe for exploration. For example, the proper regulation of these loci, particularly in terms of regulation of histone expression, during early embryogenesis is crucial – mounting evidence suggests that the pool of maternally deposited histones relative to genomic DNA determines the timing of EGA (Amodeo *et al.*, 2015; Jevtić and Levy, 2017; Joseph *et al.*, 2017b).

CHAPTER II

Repression of endogenous retrovirus transcription by a tRNA fragment

Abstract

Paternal diet has been shown to influence offspring metabolism, although the mechanisms by which paternal dietary information is transmitted are unknown. Sharma *et al*. observed that protein restriction in mice affects small RNA levels in mature sperm, in particular 5' fragments derived from glycine tRNAs. These tRNA fragments are apparently scarce in testicular sperm, but increases in abundance during epididymal transit. I show that epididymosomes, vesicles that fuse with sperm during maturation through the epididymis, have a concordant profile of tRNA fragments with sperm, and can deliver these fragments to immature sperm *in vitro*. Functionally, tRNA-glycine-GCC fragments repress genes associated with the endogenous retrovirus MERVL, in mouse embryonic stem cells and embryos. Surprisingly, this tRNA fragment represses the synthesis of MERVL transcripts, establishing a novel regulatory role for a tRNA fragment in the expression of an endogenous retroelement active in preimplantation embryos.

Introduction

Increasing evidence supports the idea that ancestral environmental conditions can influence offspring phenotype via epigenetic inheritance mechanisms (Heard and RA Martienssen, 2014; Bošković and Rando, 2018). Within this burgeoning field spanning all well-studied organisms, a number of

mammalian models point to epigenetic inheritance of ancestral stress information, including nutritional limitations or excesses from both parental lineages (Dunn and Bale, 2009; Jimenez-Chillaron *et al.*, 2009; Carone *et al.*, 2010; Ng *et al.*, 2010a; Ding *et al.*, 2012; Radford *et al.*, 2012, 2014). Together, these works support the appealing, yet controversial, hypothesis that transgenerational epigenetic inheritance may contribute to human epidemiology of disease (Pembrey *et al.*, 2014).

Previous work in the Rando lab showed that male mice fed a Low Protein diet sire offspring with altered hepatic cholesterol biosynthesis (Carone *et al.*, 2010). Subsequently, it was shown that the diet induced intergenerational phenotype could be recapitulated via IVF, therefore supporting the hypothesis that the carrier of epigenetic information co-purifies with sperm and is delivered into the fertilized oocyte (Sharma *et al.*, 2016). Studies in the underlying mechanisms of imprinting, position-effect variegation, epivariation and other epigenetic phenomena have revealed three major classes of epigenetic information carriers: DNA-cytosine methylation, chromatin state, and RNA, in particular small RNAs or sRNAs (Heard and RA Martienssen, 2014). Since previous efforts of epigenomic profiling studies in sperm argued against DNA-cytosine methylation (Shea *et al.*, 2015) or histone occupancy (Carone *et al.*, 2014) as functionally relevant in our model, we profiled the small RNA (sRNA) payload of sperm. Consistent with a previous report (Peng *et al.*, 2012b), sequencing of sRNAs < 40nt isolated from cauda, in mature sperm, revealed that 80% of all sRNAs with cloneable 3’ ends were derived from the 5’ ends of tRNAs

(**Fig. 2.1**). These tRNA fragments, also known as tRFs or 5'tiRs (Kumar, Kuscu and Dutta, 2016), are between 28-32 nt in length, with a series of predominant 3' ends around the anti-codon loop, presumably derived from either degradation or alternative cleavage/processing of tRNAs (**Fig. 2.1B-D**). tRFs derived from Glu-CTC, Gly-GCC, and Val-CAC were particularly abundant. Importantly, comparison of Low Protein to Control sperm revealed consistent changes in several small RNAs across eight paired replicate sperm samples (**Fig. 2.1E-F**), with 5' fragments of tRNA-Glycine isoacceptors, Lys-CTT, and His-GTG being consistent up-regulated 2-3 fold, while miRNAs from the let-7 family were consistently down-regulated in Low Protein sperm. Together, the high abundance of specific tRFs derived from various isoacceptors, and their consistent misregulation in Low Protein sperm, suggests that these tRFs could act as epigenetic information carriers in our model.

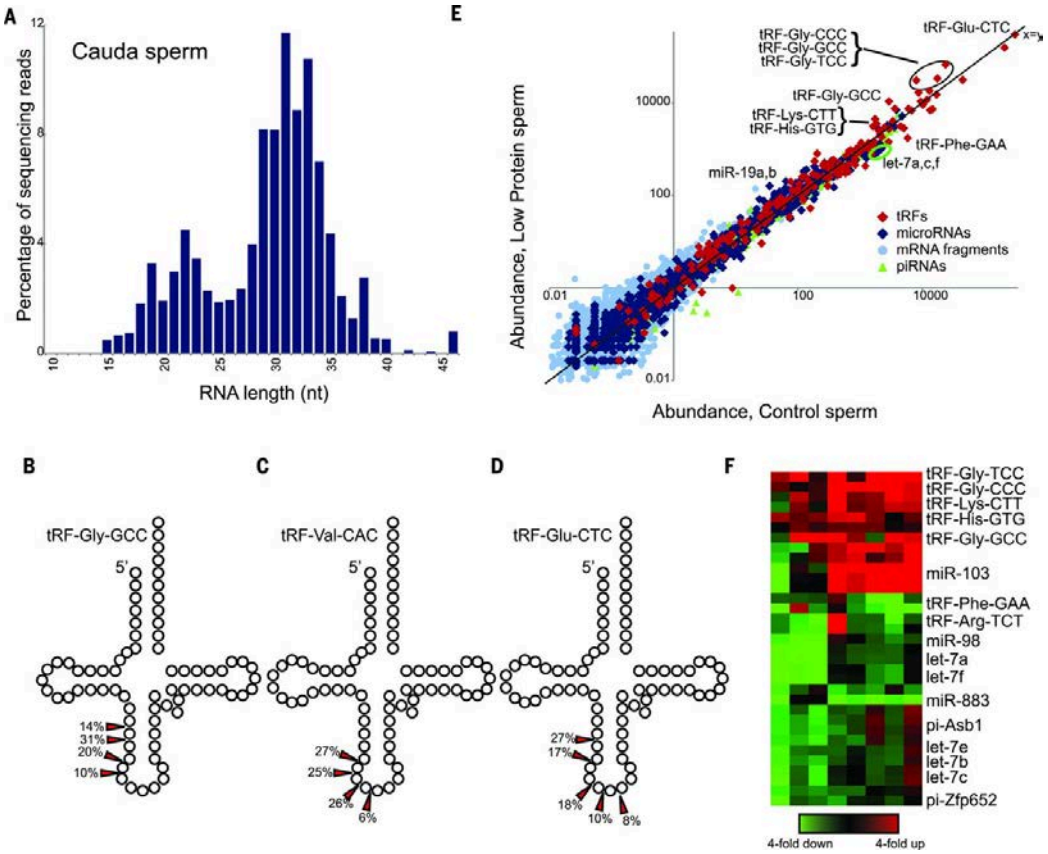


Fig. 2.1. Dietary effects on small RNAs in sperm. (A) Size distribution of mature or cauda sperm sRNAs. **(B-D)** Arrowheads indicate predominant 3' ends. **(E)** Dietary effects on sperm sRNA content. The scatterplot shows RNA abundance (in parts per million) for sperm isolated from control animals versus low-protein sperm, with various RNA classes indicated. Multiple points for tRFs result from sequence differences between genes encoding a given tRNA isoacceptor. **(F)** Heatmap showing RNAs responding to diet across eight paired sperm samples. Reprinted from (Sharma *et al.*, 2016).

Here, I set out to test the origin and explore the function of tRFs in mature sperm. I show that, consistent with deep sequencing and northern blotting data, the majority of tRFs in cauda sperm are gained after sperm exit the testis, within the epididymis, and that tRFs may be gained through the fusion of sperm with extracellular vesicles secreted from the epididymis called epididymosomes. In an attempt to probe potential functions of tRFs in the fertilized embryo, we identified one tRNA fragment – tRF-Gly-GCC, or tRF-GG, as a repressor of a subset of genes regulated by the long terminal repeat (LTR) of the endogenous

retroelement MERVL, in mouse embryonic stem cells and embryos. Given that tRFs have been shown to repress translation in various organisms and cell types, I investigated the possibility that tRF-GG regulates the translation of known MERVL repressors. Surprisingly, I find that impairing tRF-GG function had no effect on translation of MERVL regulators, and very subtle effects on translation in general. Instead, I show that tRF-GG represses nascent transcription of MERVL, rather than through the expected sRNA-mediated RNA decay pathway. I discuss how changes in the level of tRF-GG, and other sRNAs delivered via sperm and modulated by the epididymis, may lead to alterations in metabolic status in the offspring.

Results

Epididymosomes can shape the RNA payload of sperm

tRNA cleavage occurs mostly in response to stress (Yamasaki, Ivanov, G.-F. Hu, *et al.*, 2009; Ivanov, Mohamed M. Emara, *et al.*, 2011a), and tRFs could therefore be merely degradation products with little directed biogenesis or function. Two observations are inconsistent with the hypothesis that tRFs in sperm are degraded remnants of RNAs present during spermatogenesis: dietary effects on levels of intact tRNAs in the testis were uncorrelated with tRF abundance changes in cauda sperm, and sRNAs from testicular tissue (**Fig. 2.2**) and even post-meiotic spermatids had a very low abundance of tRNA fragments according to deep sequencing and Northern blotting (Sharma *et al.*, 2016). Instead, sRNAs isolated from the epididymis, where sperm continue to mature for several days after they exit the testis, exhibit high levels of various 5' tRFs according to Taqman qRT-PCR (**Fig. 2.2**), deep sequencing and northern blotting. Particularly striking is the increase in tRF abundance from the proximal epididymis (caput) to distal region (cauda), where sperm has matured (**Fig. 2.2**). This raised the surprising possibility that tRFs in cauda sperm might originate from the epididymal luminal epithelium.

During epididymal transit, sperm gain proteins via fusion with small extracellular vesicles known as epididymosomes (Sullivan, Frenette and Girouard, 2007; Sullivan and Saez, 2013; Dacheux and Dacheux, 2014). Since extracellular vesicles carry functional RNAs in many cell types (Valadi *et al.*, 2007; van Niel, D'Angelo and Raposo, 2018a), we hypothesized that

epididymosomes fuse with sperm to reshape its RNA payload during epididymal transit.

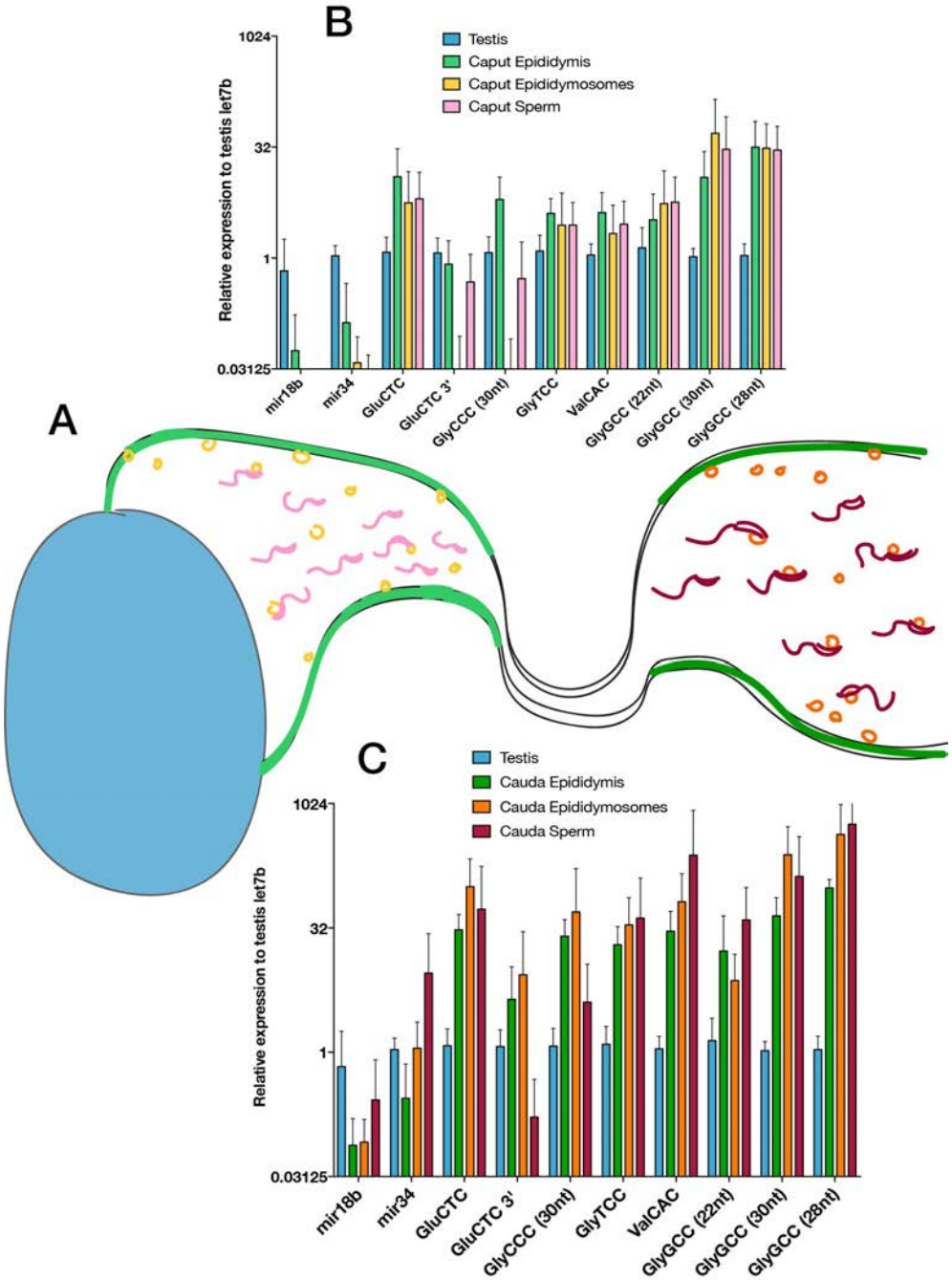


Fig. 2.2. Sperm gain tRFs during epididymal transit, potentially through fusion with epididymosomes. (A) The epididymis is a long convoluted tubule separated anatomically into three sections, the proximal to testis (caput, light green), middle (corpus, not highlighted), and distal (caudal, dark green) epididymis, which leads mature sperm into the vas deferens. As sperm transit through the epididymis, it picks up extracellular vesicles called epididymosomes

(yellow for caput and orange for cauda), which at least in part facilitate its maturation. All graphs are color-coded as in this cartoon. **(B)** Expression of various tRFs and miRNAs relative to testis let7b, for testis, caput epididymis, epididymosomes, and sperm, as measured by Taqman qRT-PCR. **(C)** Expression of the same tRFs and miRNAs relative to testis let7b, for testis, cauda epididymis, epididymosomes, and sperm.

Epididymosomes purified from caput and cauda epididymis by differential centrifugation were subject to deep sequencing, Northern blotting, and qRT-PCR for select tRFs and miRNAs (**Fig. 2.2**) (Sharma *et al.*, 2016). Surprisingly, the levels of various tRFs of both caput (**Fig. 2.2B**) and cauda epididymosomes (**Fig. 2.2C**) show a striking correlation with the levels of tRFs present in sperm from these epididymal regions. The correlation was particularly high for 5' tRFs derived from glycine and valine isoacceptors, while certain 3' tRFs such as those derived from GluCTC were relatively depleted in sperm (**Fig. 2.2B-C**). In contrast, sperm-borne miRNAs did not increase in abundance during epididymal transit, and were poorly correlated with epididymosome-bound miRNAs levels. These qRT-PCR data agree remarkably well with both deep sequencing of sRNAs and northern blotting of specific tRFs (Sharma *et al.*, 2016). Together, these results show that the sRNA payload of caput and cauda epididymosomes resembles that of sperm within each compartment of the epididymis, and that sperm can potentially gain sRNAs as they transit through this tissue to maturation.

To directly test the hypothesis that epididymosomes can reshape the RNA payload of sperm, we stringently purified sperm from the caput epididymis over a Percoll gradient, briefly incubated them with cauda epididymosomes, then pelleted and washed the "reconstituted" sperm (**Fig. 2.3A**). Fusion of cauda

epididymosomes was sufficient to deliver tRF-Val-CAC and tRF-Gly-GCC to caput sperm (**Fig. 2.3B**), confirming that tRF-bearing epididymosomes either are capable of fusing with sperm to deliver their small RNA cargo (Belleannée *et al.*, 2013), or adhere to caput sperm strongly enough to resist removal by several consecutive washing steps. These results were repeated using the more abundant caput sperm samples obtainable from the bull *B. taurus*, with cauda epididymosome fusion with caput sperm (n=4) resulting in delivery of tRF-Val-CAC and other tRFs to relatively immature caput sperm (Sharma *et al.*, 2016).

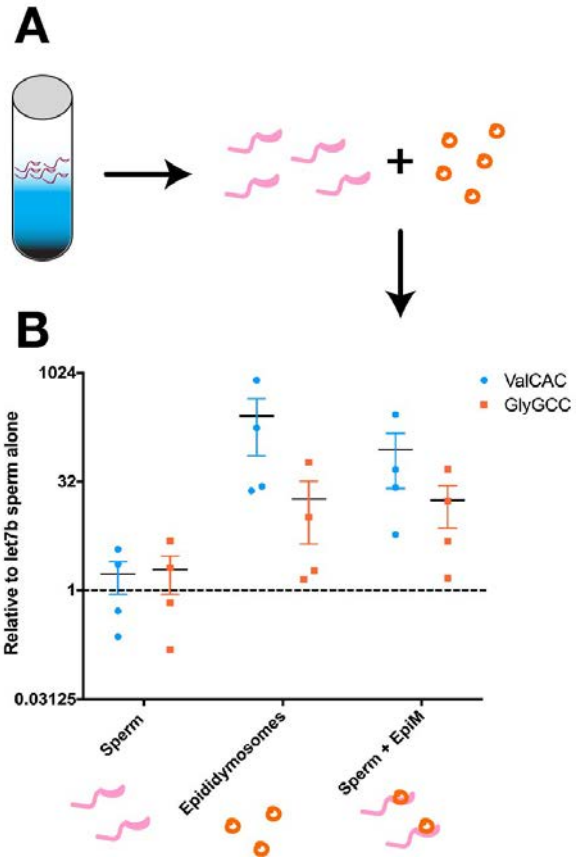


Fig. 2.3. Epididymosomal fusion can deliver tRFs to sperm. (A) Caput sperm was stringently purified via percoll gradient centrifugation, then incubated with cauda epididymosomes at 30C for ***. The sperm was pelleted briefly, which does not pellet epididymosomes, and washed. **(B)** Taqman qRT-PCR of tRF-Val-CAC and tRF-Gly-GCC, two 5' tRFs that are gained from caput to cauda, for sperm alone, epididymosomes alone, and sperm fused with epididymosomes.

Taken together, these experiments are most consistent with a mechanism of RNA biogenesis in mammalian sperm in which tRFs generated in the epididymis are trafficked to sperm in epididymosomes, although we note that they do not rule out the alternative hypothesis: that intact tRNAs in immature sperm are cleaved in maturing sperm to generate tRFs during

epididymal transit. Definitive identification of the tissue of origin of tRFs in mature sperm will require epididymis-specific expression of polymorphic “tracer” tRNAs (Sharma *et al.*, 2018), or the inactivation of tRNA cleaving enzymes in either testes or epididymis.

tRF-GG regulates a subset of the MERVL-driven totipotency program

What are the potential downstream targets of the diet-regulated RNAs in sperm? As mechanistic experiments are technically challenging in the desired cell type (the early mouse embryo), we first use embryonic stem (ES) cells as an experimental system amenable to mechanistic analysis. We interfered with the function of specific tRFs using antisense LNA-containing oligonucleotides in ES cell culture, and assayed mRNA abundance as readout of tRF inhibition effects. Most antisense oligos had no effect on mRNA abundance (**Fig. 2.4A**), suggesting that the targeted tRFs are not functional in ES cells, or that they exert regulatory effects that are not assayed by mRNA abundance. In contrast, interfering with tRF-Gly-GCC function resulted in dramatic up-regulation of ~70 genes, with several genes being up-regulated over 10-fold (**Fig. 2.4B**). Up-regulation of these genes was consistently observed by microarray (**Fig. 2.4A,** n=7), and further confirmed by RNA-Seq (**Fig. 2.4C,** n=4). These genes were unaffected by antisense LNA oligos directed against the 5’ ends of other tRNA-Gly isoacceptors or against the middle or the 3’ end of tRNA-Gly-GCC (**Fig. 2.4A**). This last finding strongly suggests that changes in gene expression caused by interfering with tRF-Gly-GCC are unlikely to be an artifact of interfering with the function of the intact tRNA.

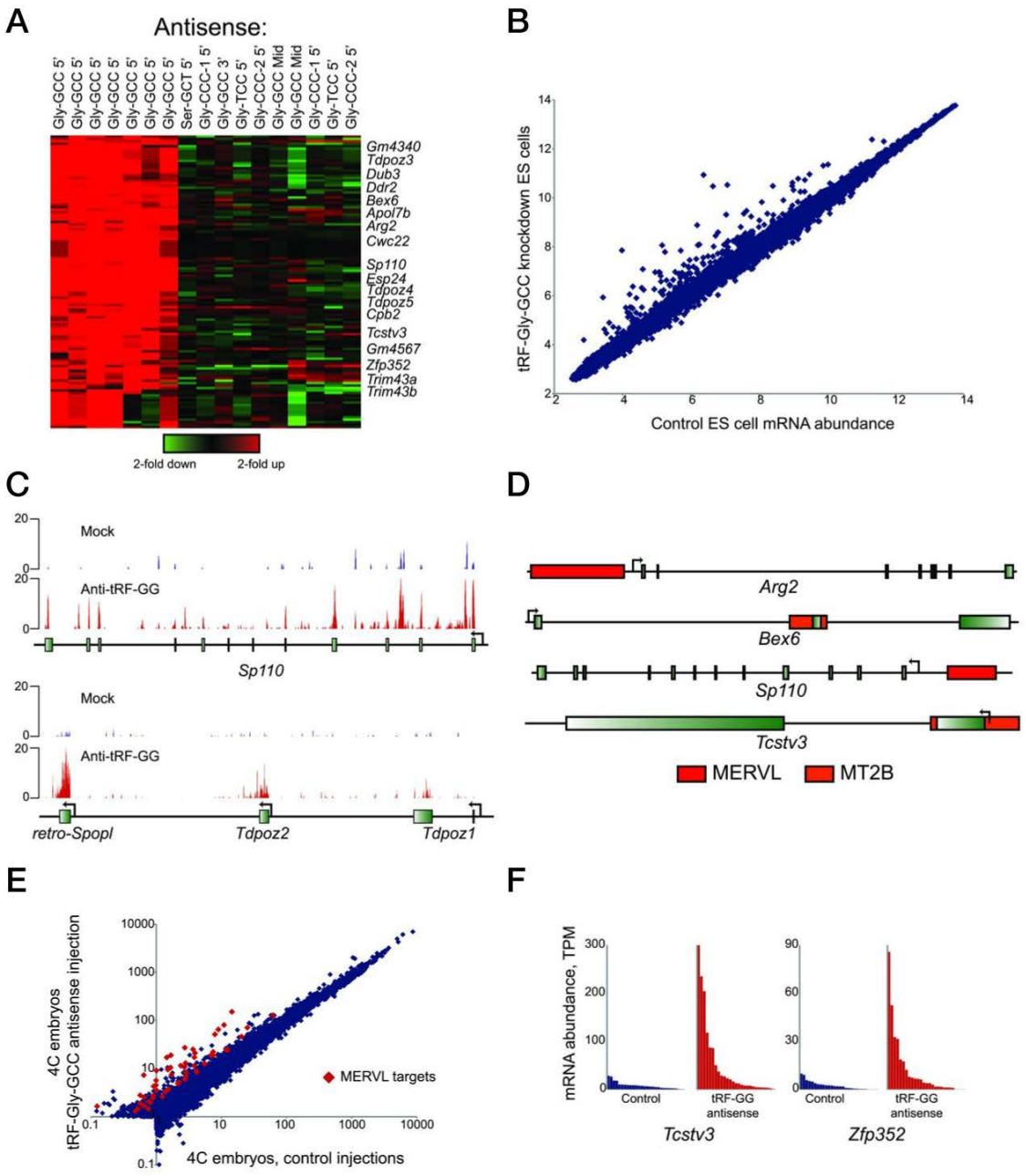


Fig. 2.4. tRF-GG represses MERVL transcriptional program in mESCs and mouse embryos. (A) Affymetrix microarray data for mRNA abundance in embryonic stem cells transfected with indicated LNA antisense oligo. Note specificity for MERVL associated genes in tRF-Gly-GCC knockdown but not other isoacceptors. **(B)** Microarray data for mRNA abundance in embryonic stem cells transfected with an LNA antisense oligo targeting the 5' end of tRF-Gly-GCC. Scatterplot shows mRNA abundance in in anti-GFP knockdown cells (x axis) vs. tRF-Gly-GCC inhibition (y axis). Data represent average of seven replicates. **(C)** RNA-Seq data was pooled from four replicate samples of ES cells transfected with shRNA against GFP, or with the anti-tRF-Gly-GCC LNA oligo, as

indicated. **(D)** Schematic showing genomic context for four tRF-Gly-GCC target genes, showing MERVL LTRs associated with all target genes. Some additional target genes, such as the *Tdpoz* cluster, are not as closely associated with MERVL LTRs but instead are located in large MERVL-rich genomic clusters, and have also been shown to be part of the MERVL-regulated gene expression program (Macfarlan *et al.*, 2012b). **(E)** Inhibition of tRF-Gly-GCC affects MERVL target expression in 4-cell embryos. Averaged single embryo RNA-Seq data for control (n=28) or tRF-inhibited (n=27) embryos. Among genes upregulated at least 2-fold on average, those previously described as MERVL targets are indicated separately. **(F)** Examples of single embryo data for two MERVL targets. Here, each bar represents mRNA abundance from a single embryo, with embryos ordered from highest to lowest expression for each condition.

Surprisingly, all the genes up-regulated in tRF-Gly-GCC knockdowns are highly expressed in 2-cell and 4-cell embryos, and have been shown to be regulated by the long terminal repeat (LTR) of the endogenous retroelement MERVL (Macfarlan *et al.*, 2011, 2012b) (**Fig. 2.4D**). To determine whether the effects of tRF-Gly-GCC inhibition observed in tissue culture also hold in a more physiological context, we microinjected zygotes (n=27) with an antisense oligo directed against tRF-Gly-GCC. These embryos were then allowed to develop to the 4-cell stage and subjected to single embryo RNA-Seq (**Fig. 2.4E**). Strikingly, we observe significant up-regulation of 72 transcripts in embryos subject to tRF-Gly-GCC inhibition compared to control embryos (n=28), with the majority of up-regulated genes having previously been identified as MERVL targets (Macfarlan *et al.*, 2012b) (**Fig. 2.4F**), and overlapping the same genes we observed as tRF-Gly-GCC targets in mESC experiments.

Paternal dietary effects on gene regulation in preimplantation embryos

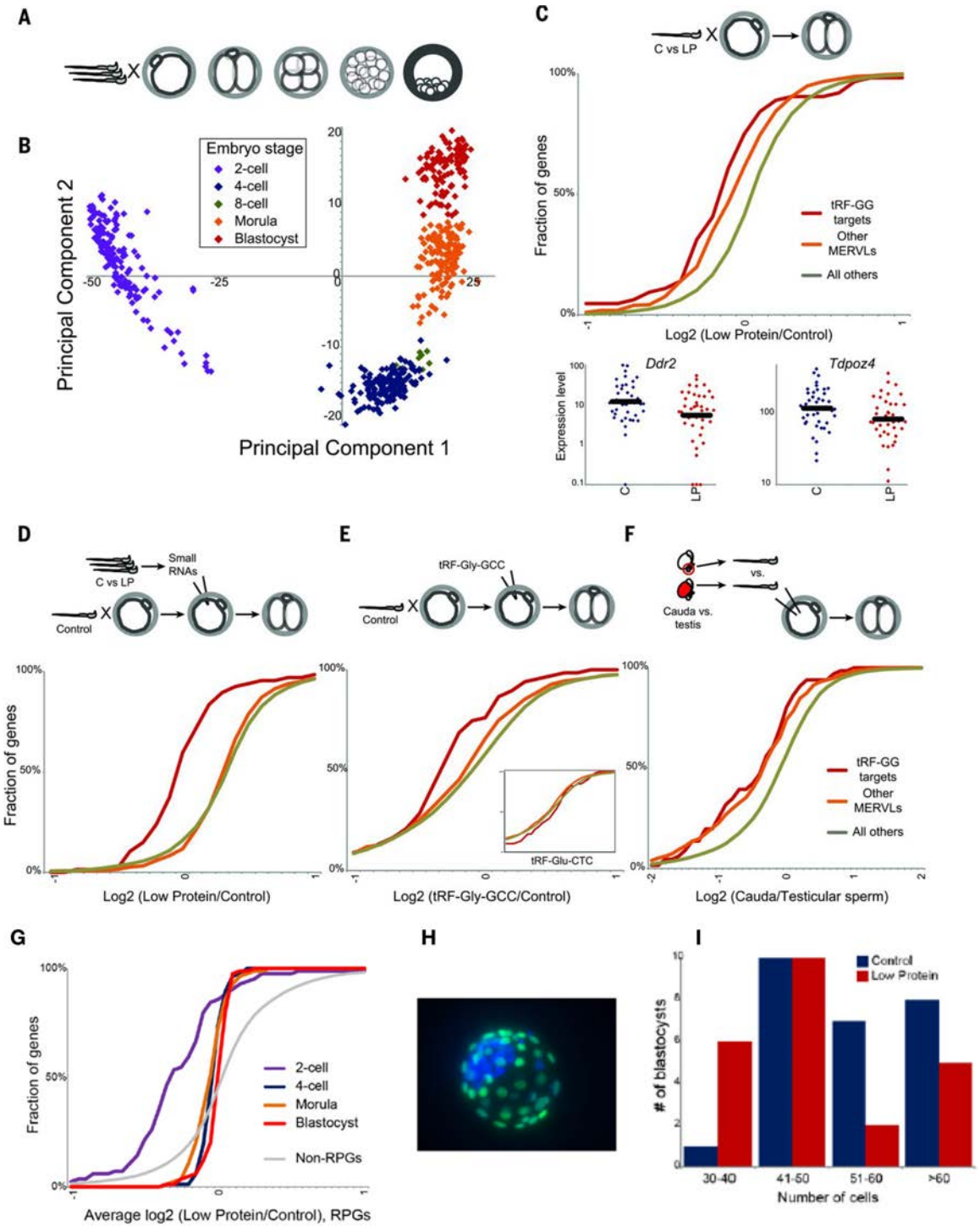


Fig. 2.5. Paternal diet affects embryonic gene expression through small-RNAs in sperm. (A) Embryos generated by IVF were cultured for varying times, then subject to single embryo RNA-Seq. **(B)** Single-embryo RNA-Seq data for preimplantation embryos represented via principal component analysis. The first

two principal components explain 74% of the total variance in the dataset, and cleanly separate the five embryonic stages shown here. **(C)** Abundance of mRNA in two-cell embryos generated via IVF using control versus low-protein sperm (*n* = 41 C and 39 LP embryos). Cumulative distribution plots for tRF-Gly-GCC targets ($P = 4.5 \times 10^{-7}$, Kolmogorov-Smirnov test), other MERVL targets (Macfarlan *et al.*, 2012b) ($P = 2.5 \times 10^{-13}$), and all remaining genes, showing the percentage of genes with the average $\log_2$(LP/C) indicated on the *x* axis. Low-protein embryos exhibit a significant shift to lower expression of MERVL targets. Bottom panels show individual embryo data for two targets. **(D)** Small RNAs isolated from control or low-protein cauda sperm were microinjected into control zygotes. RNA-seq (*n* = 42 C and 46 LP embryos) reveals down-regulation of tRF-Gly-GCC targets ($P = 4.8 \times 10^{-14}$) driven by low-protein RNAs. **(E)** Effects of synthetic tRF-Gly-GCC on two-cell gene regulation, showing significant (*P* = 0.0001) down-regulation of target genes in embryos injected with tRF-Gly-GCC (*n* = 26) versus GFP controls (*n* = 11). The inset shows effects of tRF-Glu-CTC (*n* = 6). **(F)** Effects of epididymal passage on embryonic gene regulation. Intact sperm isolated from the rete testis (*n* = 12) or cauda epididymis (*n* = 9) were injected into control oocytes, and mRNA abundance was analyzed as described above. **(G)** Dietary effects on ribosomal protein gene expression at four embryonic stages. Graph shows cumulative distribution for Low Protein effect on all ribosomal protein genes at the indicated stages. Grey line shows distribution of dietary effects on all non-RPG genes, for all four stages. **(H)** Blastocyst stained with DAPI (blue) and anti-CDX2 (green) to image total cell number and trophectoderm cells. **(I)** Low Protein diet reproducibly alters developmental tempo. For embryos generated via IVF using Control or Low Protein sperm, plot shows the number of blastocysts with the indicated number of cells.

Given the robust connection between a diet-regulated small RNA and a highly specific set of target genes, we asked whether targets of tRF-Gly-GCC are regulated in preimplantation embryos generated using sperm from animals consuming Control or Low Protein. We generated embryos via *in vitro* fertilization (IVF), and then carried out RNA-Seq of individual embryos cultured to the 2-cell (n=80), 4-cell (n=82), 8-cell (n=4), morula (n=73), and blastocyst (n=41) stages of preimplantation development (**Fig. 2.5A**). Principal component analysis of all embryos analyzed revealed robust clustering of embryos by developmental stage (**Fig. 2.5B**), with the first two principal components representing oocyte-derived transcripts (PC1), and products of embryonic genome activation (PC2).

We next turned to dietary effects on mRNA abundance. As single embryo RNA-Seq data are not suitable for identification of modest changes in individual mRNA targets, we sought to identify consistent changes in larger gene sets, primarily focusing on the subset of MERVL targets identified in our ES cell studies, and on the remaining ~500 MERVL targets defined by MacFarlan et al. At the 2-cell stage both tRF-Gly-GCC targets, as well as remaining MERVL targets, were downregulated in Low Protein embryos relative to Control (**Fig. 2.5C**, $p=1.1 \times 10^{-16}$ KS test). We carried out several independent tests of the hypothesis that tRF-Gly-GCC in sperm affects expression of MERVL targets in early embryos. First, to test the hypothesis that small RNAs are responsible for the effect of diet on MERVL targets, <40 nt RNA populations from Control and Low Protein sperm were purified and injected into IVF-derived zygotes, and gene expression was analyzed in resulting 2-cell embryos. Compared to Control RNA injections, Low Protein RNA recapitulated the modest but systematic inhibition of tRF-Gly-GCC target genes in 2-cell embryos (**Fig. 2.5D**), supporting the hypothesis that paternal diet affects preimplantation gene expression via its effects on the RNA payload of sperm. Second, to further define the relevant RNA species from Low Protein sperm, microinjection of a 30 nt oligonucleotide corresponding to tRF-Gly-GCC resulted in lower levels of MERVL genes in a subset of 2-cell embryos (**Fig. 2.5E**). Finally, as our data show that most tRFs in sperm are gained during epididymal transit, we generated embryos via intracytoplasmic sperm injection (ICSI) using either testicular spermatozoa or cauda sperm. Consistent with the higher levels of tRF-Gly-GCC present in cauda

sperm, we find that embryos generated using cauda sperm expressed MERVL targets at lower levels than those generated using testicular sperm (**Fig. 2.5F**). Together, the effects of diet on IVF-derived embryos, the small RNA injections, and ICSI using various sperm populations all support the hypothesis that tRF-Gly-GCC in sperm is capable of delaying or repressing MERVL targets in 2-cell stage embryos.

Finally, we note that tRF-Gly-GCC is only one of several high-abundance RNAs regulated by Low Protein diet, and MERVL-driven genes are not the only diet-responsive genes in preimplantation embryos. Most notably, ribosomal protein genes (RPGs) were expressed at lower levels in Low Protein embryos relative to Controls at several stages (**Fig. 2.5G**). Given that RPG expression is often linked to cell proliferation, we counted cell number after four days of embryo culture – at the blastocyst stage – finding that paternal diet affected the developmental clock (Mitchell, Bakos and Lane, 2011; McPherson *et al.*, 2013), with Low Protein embryos (n=23) developing slower relative to Control embryos (n=26) (**Fig. 2.5H-I**). Future studies will be needed to determine whether altered preimplantation growth kinetics, or regulation of the MERVL program, are responsible for eventual metabolic consequences in offspring. However, as the MERVL program is linked to totipotency (Macfarlan *et al.*, 2012b), we speculate that regulation of MERVL-linked genes by tRF-Gly-GCC might affect placental size or function, leading to downstream effects on metabolism known to be secondary to altered placental function (Rando and Simmons, 2015).

tRF-GG represses nascent synthesis of MERVL RNAs

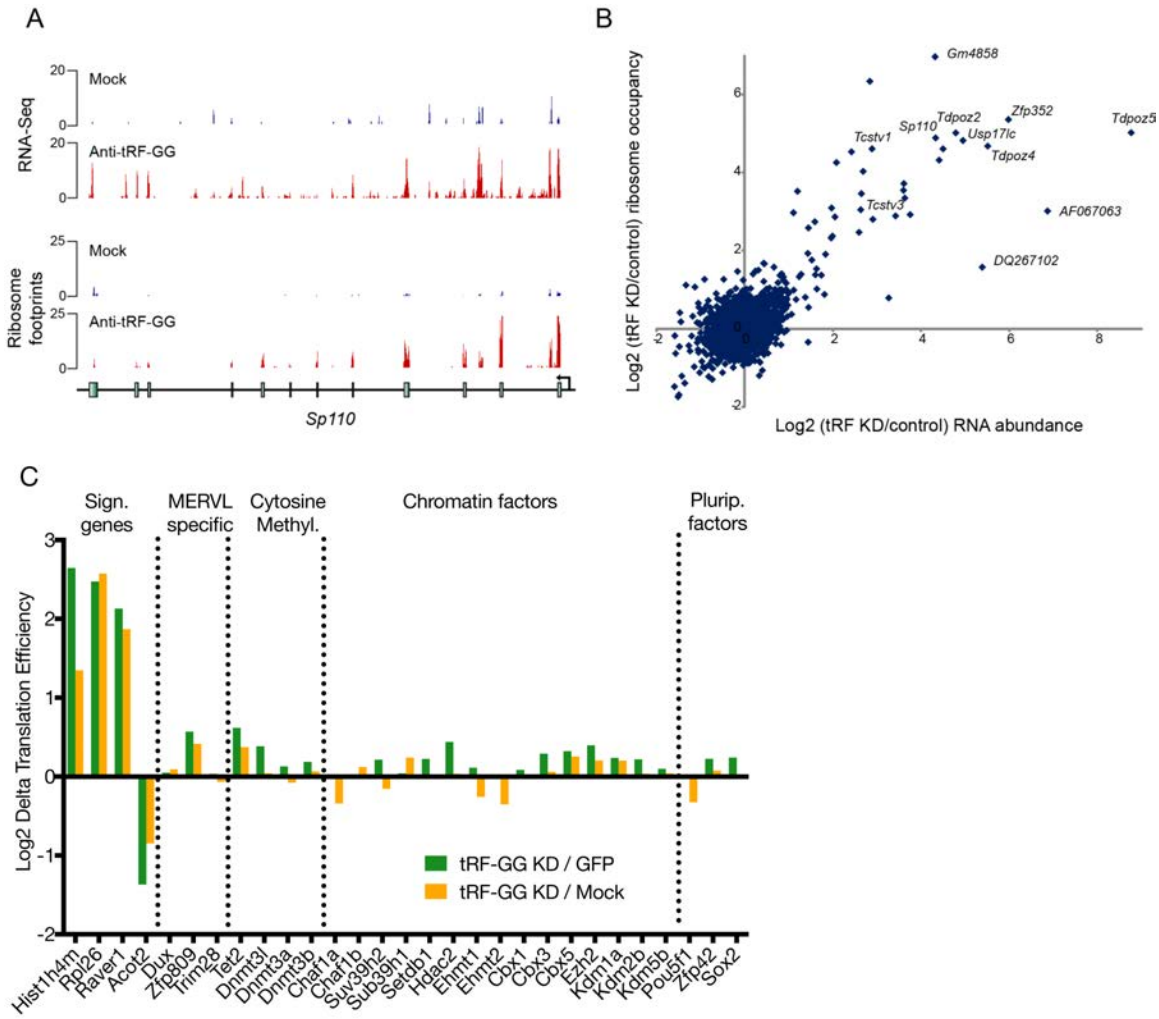


Fig. 2.6. tRF-GG does not function in translation. (A) Genome browser tracks showing increased expression of MERVL target gene *Sp110* for both RNA and ribosome footprints upon tRF-GG KD. **(B)** tRF-GG has little or no role in global translation of mRNAs. Note the high level of correlation between change in ribosome footprints and RNA levels upon tRF-GG KD. Note the highlighted MERVL target genes fall close to the diagonal (i.e. no change in translation efficiency). **(C)** Very few genes had consistent changes in translation efficiency upon tRF-GG KD, and no changes were observed for known MERVL regulators or other chromatin factors that play a role in repressing endogenous retroelement expression.

Finally, I turn to the mechanism by which tRF-Glycine-GCC regulates MERVL expression in mouse embryonic stem cells and preimplantation embryos.

Since tRFs repress translation in various cellular and organismal contexts

(Yamasaki, Ivanov, G.-F. Hu, *et al.*, 2009; Ivanov, Mohamed M. Emara, *et al.*, 2011a; Bąkowska-Żywicka *et al.*, 2016; Goncalves *et al.*, 2016; Gebetsberger *et al.*, 2017b; Guzzi *et al.*, 2018; Luo *et al.*, 2018), I first investigated the possibility that tRF-GG represses the translation of MERVL activators or related machinery. To study the effects of tRF-GG on translation, I carried out ribosome profiling (Ingolia, Lareau and Weissman, 2011) along with RNA-seq in mouse ES cells to measure changes in ribosome occupancy genome-wide upon tRF-GG KD (**Fig. 2.6**). Consistent with our previous microarray data, tRF-GG inhibition led to the derepression of a subset of MERVL target genes at the RNA level, with a concurrent increase in ribosome occupancy over the same transcripts (**Fig. 2.6A-B**). Surprisingly, inhibition of tRF-GG had little to no effect on translation efficiency of mRNAs genome-wide. While there were a few genes that were consistently changing in mRNA abundance and ribosome occupancy, notably histone genes (see Chapter III), most changes in translation efficiency could not be statistically verified. In particular, we could not identify any genes that could be plausibly linked to MERVL expression demonstrating consistent changes in RNA abundance, ribosome occupancy, or translation efficiency, upon tRF-GG inhibition (**Fig. 2.6C**). From this data, we conclude that tRF-GG does not globally repress translation, and plays no role in translational repression of specific factors that could be involved in either activation or repression of MERVL expression.

tRNA fragments have in some cases been reported to function in complex with Argonaute proteins and act effectively like microRNAs or endo-siRNAs

(Deng *et al.*, 2015; Andrea J. Schorn *et al.*, 2017; Martinez, Choudury and Slotkin, 2017). However, the repressive activity of tRF-GG on MERVL expression does not appear to be a consequence of sequence homology between tRF-GG and the MERVL element, based on three observations. First, there is no significant sequence homology between tRF-GG and MERVL (which uses tRNA-Leu for replication), and many of the tRF-GG target genes are in any event regulated by “solo” LTRs, which have lost the MERVL primer binding sequence. Second, LNAs targeting the 3’ end of tRNA-Gly-GCC have no effect on MERVL target expression (**Fig. 2.4A**). Third, transfection of ES cells with various synthetic 3’ tRNA fragments of potential relevance to either tRF-Gly-GCC or to MERVL has no significant effect on MERVL target gene expression (**Fig. 2.7A**), arguing against tRF-GG affecting levels or function of other tRNA fragments that might have more direct roles in MERVL control. Taken together, these data do not support a role for Argonaute-mediated targeting of MERVL mRNA (or DNA) sequences in tRF-GG regulation of this gene set.

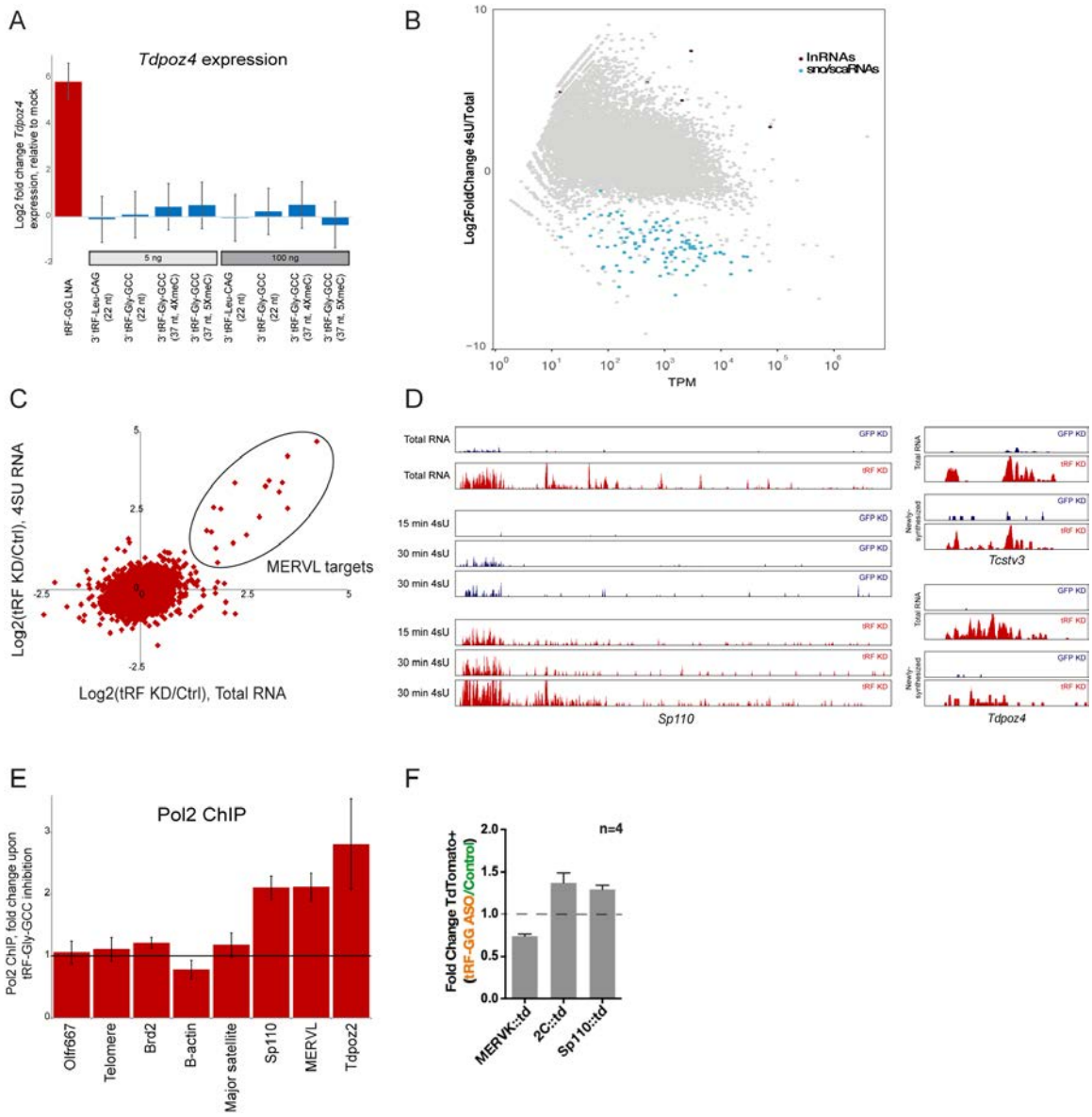


Fig. 2.7. tRF-GG repress MERVL transcription. (A) Expression of the MERVL target gene *Tdpoz4* is unaffected by several potentially relevant 3' tRNA fragments. ES cells were transfected either with the antisense LNA targeting the 5' tRF-Gly-GCC, or with either 5 or 100 ng of one of four synthetic 3' tRFs. Tested oligos include two short 22 nt 3' fragments of mature (CCA-tailed) tRNA-Gly-GCC and tRNA-Leu-CAG – the former has the potential to form a complex with 5' tRF-Gly-GCC and so its activity could be modulated by manipulating 5' tRF availability, while the latter was chosen due to the use of this tRNA in priming MERVL replication. We also included two longer (37 nt) 3' tRFs derived from tRNA-Gly-GCC which differ in whether they carry 4 or 5 5meC nucleotides, as a recent study reported significant differences between these oligonucleotides in both RNA stability and structure (Zhang *et al.*, 2018). Importantly, with the exception of the LNA targeting 5' tRF-GG, none of the other oligos, even at elevated concentrations, had any significant effect on MERVL target gene

expression in ES cell cultures. **(B)** Metabolic labeling of RNA with 4sU libraries demonstrate expected behavior of long-noncoding RNAs (fast transcription rate) and sno/scaRNAs nascent RNA processing (slow transcription rate due to splicing and post-transcriptional processing) (Windhager *et al.*, 2012). **(C)** Metabolic labeling scatterplot showing effects of tRF-GG inhibition on total RNA (x axis) vs. newly-synthesized 4SU-labeled RNA (y axis). **(D)** Metabolic labeling reveals transcriptional derepression upon tRF-Gly-GCC inhibition. Genome browser tracks show total RNA levels, and newly-synthesized RNAs obtained after 15 or 30 minutes of 4-thiouridine (4SU) labeling, for ES cells transfected with esiRNAs targeting GFP, or an LNA oligonucleotide antisense to tRF-Gly-GCC. Effects of tRF inhibition on previously-described MERVL-associated target genes (Sharma *et al.*, 2016) are nearly identical for total RNA as well as newly-synthesized RNA. **(E)** Pol-II ChIP qPCR shows increased Pol-II occupancy on MERVL and MERVL target gene promoters, but not other types of ERV promoters. **(F)** Increased MERVL-LTR driven TdTomato signal upon tRF-GG knockdown. Two independent MERVL LTRs, or a MERVK LTR as control, were cloned into a vector driving TdTomato expression.

Alternatively, tRF-GG may repress nascent synthesis of MERVL target genes. Consequently, we set out to distinguish between tRF-GG regulation of target gene synthesis vs. RNA decay. Metabolic labeling with 4-thiouridine (4sU) (Rabani *et al.*, 2011) was used to label newly synthesized RNAs in murine E14 ESC cultures following inhibition of tRF-GG function using an LNA-containing antisense oligo. The 4sU method efficiently captured nascent RNA synthesis, and recapitulated previously reported characteristics of 4sU libraries, including depletion of snoRNAs due to slow intron processing at snoRNA host genes (**Fig. 2.7B**) (Windhager *et al.*, 2012). Replicating our prior findings, inhibition of tRF-GG resulted in dramatically increased levels of total MERVL target mRNA abundance (**Fig. 2.4, 2.6, 2.7C-D**). Importantly, we also observed increased levels of MERVL target genes following tRF-GG inhibition in the purified 4sU-labeled, newly-synthesized, mRNA fraction, indicating that tRF-GG represses target gene synthesis rather than stability (**Fig. 2.7C-D**). In addition, MERVL target derepression was accompanied by increased RNA Pol2 levels on target

genes (**Fig. 2.7E**), and could be recapitulated using two independent MERVL LTR-driven tdTomato fluorescent reporters (**Fig. 2.7F,** see methods). We conclude that tRF-Gly-GCC plays a role in *transcriptional* repression of MERVL LTRs.

Next, we asked how transcription of MERVL-driven genes is controlled. Like many retroelements, MERVL LTRs are packaged into, and repressed by, heterochromatin (Macfarlan *et al.*, 2012b). Interfering with chromatin assembly, for instance via knockdown of the CAF-1 histone chaperone (Ishiuchi *et al.*, 2015b), results in derepression of MERVL-driven transcripts. To investigate the effects of tRF-Gly-GCC on chromatin architecture, we carried out ATAC-Seq (Buenrostro *et al.*, 2015) in mouse ES cells to measure changes in chromatin accessibility genome-wide upon tRF-GG KD (**Fig. 2.8**). Our ATAC-seq libraries recapitulated previously described characteristics of Tn5-transposed genomes (**Fig. 2.8A-C**). Consistent with the enhanced transcription observed at MERVL LTRs, we find that inhibition of tRF-GG resulted in a broad increase in chromatin accessibility over MERVL elements and throughout heterochromatin (**Fig. 2.8D-F**), with minimal changes in ATAC-Seq signal over euchromatic transcriptional start sites (**Fig. 2.8G-I**). To extend these findings to a more developmentally relevant scenario, we microinjected IVF-derived zygotes with a synthetic tRF-GG oligonucleotide to mimic the process of sperm delivery of tRFs to the zygote. We assayed chromatin accessibility in control and injected embryos by ATAC-qPCR, confirming that increasing tRF-GG levels resulted in a decrease in accessibility at

MERVL elements (**Fig. 2.8J**). Thus, tRF-Gly-GCC manipulation alters chromatin accessibility in both ES cells and in preimplantation embryos.

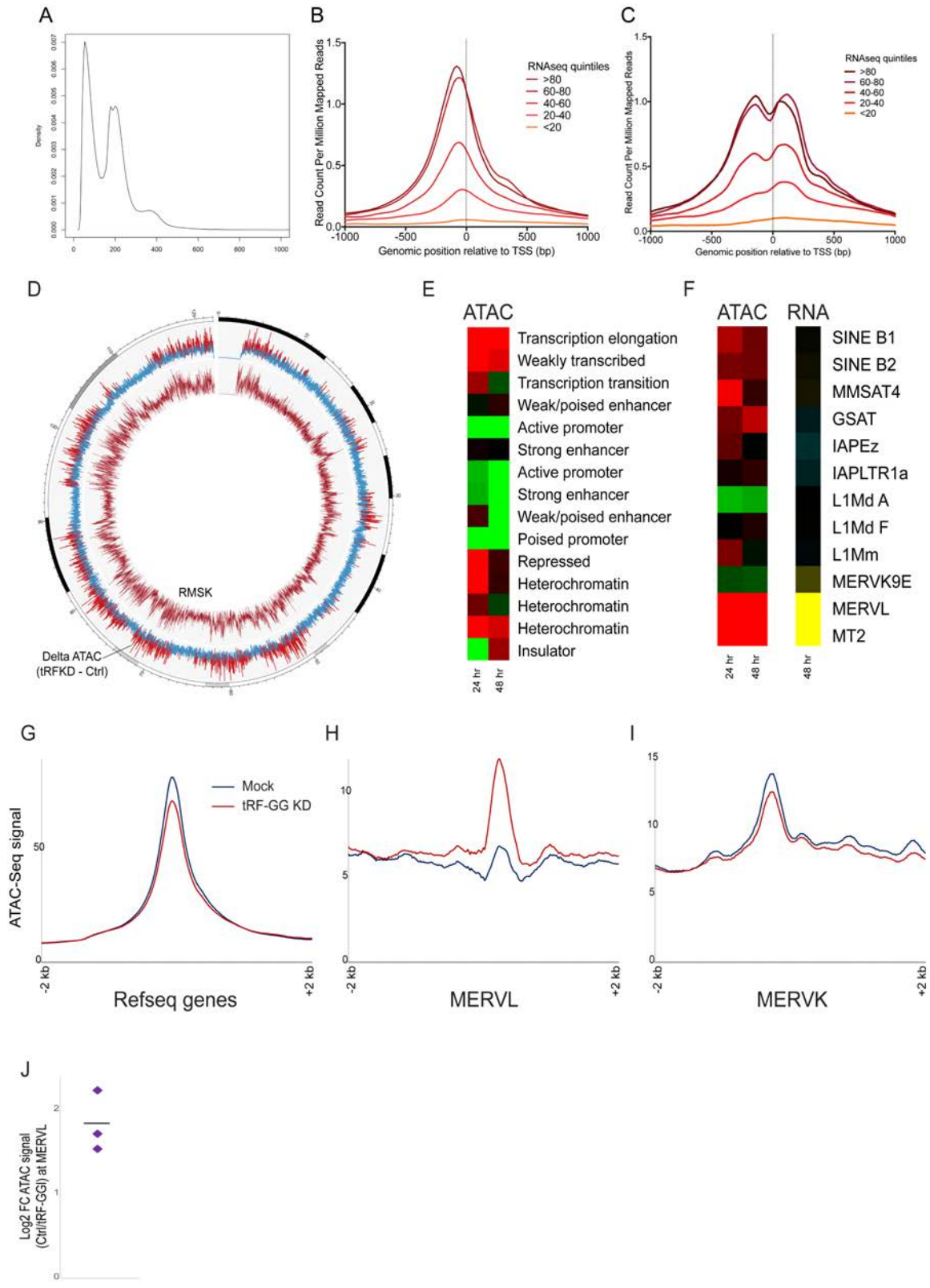


Fig. 2.8. tRF-GG regulates chromatin compaction. (A) Fragment size distribution of ATAC-seq libraries show characteristic enrichment of transcription

factor sized, mono- and di-nucleosome periodicity fragments. **(B)** Fragments less than <140nt in size demonstrate an enrichment over the expected nucleosome depleted region of transcriptionally active TSS's genome-wide. **(C)** Mono-nucleosome sized fragments demonstrate a bimodal enrichment over actively transcribed TSS's genome-wide. **(D)** Circos plot showing ATAC-Seq data for control and tRF-GG-inhibited ES cells across chromosome 11. Inner circle shows Repeatmasker density, and blue/red trace in the outer ring shows the change in ATAC signal between tRF KD and control ES cells, with red indicating more than 1.5 fold increased chromatin accessibility following tRF-GG inhibition. **(E)** Increased accessibility at heterochromatin and weakly-transcribed regions in tRF-GG-inhibited ES cells. Heatmap shows log2 fold change in ATAC-Seq reads following tRF-GG inhibition, aggregated across the indicated types of chromatin (Bogu *et al.*, 2015). **(F)** As in (E), with tRF-GG effects on ATAC-Seq occupancy and RNA abundance averaged across the indicated repeat elements. **(G-I)** Examples showing average ATAC-Seq signal across the indicated genomic elements – Refseq genes (G), MERVL elements (H), or MERVK elements (I). **(J)** tRF-GG injection decreases chromatin accessibility at MERVL elements in preimplantation embryos. Zygotes were microinjected with a synthetic tRF-GG oligonucleotide, or control injected (H3.3-GFP mRNA), and developed to the 2-cell stage. Groups of 5 embryos were then pooled and subject to Tn5 transposition as in ATAC-Seq, then assayed by q-PCR for both MERVL and *a-tubulin*. Data show the fold decrease in accessibility for tubulin-normalized MERVL in tRF-GG-injected embryos compared to control embryos in three replicates – horizontal line shows average of the three replicates.

Taken together, our data strongly supports the conclusion that tRF-GG represses the nascent synthesis of MERVL RNA by regulating the chromatin architecture surrounding MERVL in mouse embryonic stem cells and preimplantation embryos. Surprisingly, this repression of chromatin accessibility by tRF-GG was not limited to MERVL specifically, but heterochromatin in general.

Discussion

This study focused on exploring the biogenesis and function of diet-regulated tRNA fragments in sperm. Surprising, I observed that epididymosomes could shape the RNA payload of sperm via fusion. The RNA payload of sperm is

heavy in 5’ fragments derived from tRNAs, or tRFs, and is modulated by paternal diet. As to the function of these tRFs in preimplantation embryos, we focused on one particularly abundant diet-regulated tRF, tRF-Gly-GCC, because it is the most abundant sRNA in mature sperm that is consistently up-regulated in response to paternal diet. We found that tRF-Gly-GCC represses the expression of endogenous retroelement MERVL, which is active in the initial stages of embryonic development. I show that instead of acting through the well-understood Argonaute-mediated RNA degradation pathway, or the plausible repression of translation, tRF-Gly-GCC represses the nascent synthesis of MERVL transcripts by regulating chromatin compaction surrounding these elements. Together, these results illuminate a pathway by which environmental conditions experienced by the father can influence the very earliest stages of preimplantation development.

tRF-Glycine-GCC controls chromatin architecture and expression of MERVL

Several aspects of the finding that tRF-Gly-GCC controls chromatin architecture surrounding not only MERVL, but also heterochromatin in general, and yet represses the nascent synthesis of MERVL elements specifically, are unexpected.

First, the transcriptional silencing of transposons and associated repression of chromatin surrounding transposons by tRF-GG are more reminiscent of the piRNA pathway in *Drosophila* and RNA-induced transcriptional silencing (RITS) is fission yeast *S. pombe* (Sienski, Dönertas and Brennecke, 2012; Mohn *et al.*, 2014; Holoch and Moazed, 2015). Neither of these processes

are thought to be active in our system, as RITS requires RNA-dependent RNA polymerase, which is absent from mammals (Stein *et al.*, 2003), and piRNAs are active in the mammalian testes, not mouse embryonic stem cells or preimplantation embryos. The lack of homology between tRF-GG, or the 3' end of tRNA-Gly-GCC, and the MERVL LTR or any part of MERVL, also argue against an Argonaute mediated RNA-silencing pathway. tRF-GG may therefore regulate heterochromatin and MERVL transcription through a potentially novel pathway.

Second, our ATAC-seq data suggests that tRF-GG participates in maintaining heterochromatin compaction in general, rather than the expected limited genomic regions surrounding MERVL (**Fig. 2.8D-E**). MERVL LTR elements are interspersed throughout the genome, in around 650 full length copies or 37000 solo-LTRs (Schoorlemmer *et al.*, 2014), but only make up a small fraction of all transposable elements in the mouse genome (Nellåker *et al.*, 2012). However, heterochromatin in mouse embryonic stem cells encompass much larger genomic regions, sometimes megabases long (**Fig. 2.8D**), which are enriched in heterochromatic histone marks H3K27me2/3, H3K9me2/3, and DNA cytosine methylation (Bogu *et al.*, 2015; Walter *et al.*, 2016; Tosolini *et al.*, 2018). These regions also appear to be the first to become replicated during S phase (data not shown), in apparent contrast to their penchant to becoming under-assembled. MERVL LTRs are not the only transcriptionally active retroelements in mouse embryonic stem cells, as LTR retroelements such as IAP and MERVK, and many long interspersed nuclear elements (LINEs) are all expressed

(Elsässer *et al.*, 2015; Walter *et al.*, 2016). Therefore, it is extremely surprising that inhibition of tRF-GG derepresses the synthesis of MERVL and its target genes specifically. This specificity of expression from generality of chromatin openness could potentially result from the tendency for mouse embryonic stem cells to express MERVL target genes in its transient exploration of the 2-cell like state (Macfarlan *et al.*, 2012b; Ishiuchi *et al.*, 2015b; Eckersley-Maslin *et al.*, 2016). In fact, it has become apparent that undercompaction of chromatin leads to derepression of specific retroelements in many model systems (Qian *et al.*, 1998; Ishiuchi *et al.*, 2015b). In addition, the transcription factor that is required for expression of MERVL and its target genes appears to be ready, if not primed to be expressed in mouse embryonic stem cells (De Iaco *et al.*, 2017; Hendrickson *et al.*, 2017; Whiddon *et al.*, 2017; Iaco *et al.*, 2019). Detailed molecular understanding of how MERVL and its target genes can efficiently exploit chromatin under-assembly, and the molecular roles that these oftentimes retro-transposed or highly duplicated genes play in early embryonic development, will help elucidate the biological consequences of tRF-GG regulation of this gene set. Future work will also explore the mechanistic nature of how tRF-GG helps maintain chromatin architecture in the early embryo, and the metabolic consequences that this early modulation may impose on post-implantation development.

Sperm small RNAs influence pre-implantation development

Due to the small cytoplasmic volume of mammalian sperm, it is commonly assumed that sperm epigenetic information must be carried in the nucleus, either

via DNA cytosine methylation or chromatin state (Hammoud, David A Nix, *et al.*, 2009; Rando and Simmons, 2015). However, our study reveals important regulatory roles that sperm sRNAs can play in the early embryo, potentially by altering preimplantation development kinetics and placentation (**Fig. 2.5G-I**). While we demonstrate the regulatory potential of one particularly abundant diet-regulated sRNA, tRF-GG, we found that paternal diet also alters a myriad of other sRNAs, including tRFs and miRNAs (**Fig. 2.1**). Of these, *let7* family miRNAs play particularly pleiotropic roles in shaping animal development (Büssing, Slack and Grosshans, 2008), and further study will elucidate the potential regulatory roles *let7* and other sperm-borne sRNAs in embryonic development. Furthermore, our study further solidifies evidence that either purified sperm RNAs or synthetic RNA mixtures microinjected into the embryo are capable of recapitulating paternally induced phenotypes in offspring (Gapp *et al.*, 2014b; Rodgers *et al.*, 2015; Chen, Menghong Yan, *et al.*, 2016; Grandjean *et al.*, 2016), and therefore, that sperm-borne RNAs represent a viable source of epigenetic information that can shape offspring development. However, it is yet unclear how changes in different sRNAs (individual miRNAs vs gel-purified tRFs, for example) can lead to converging phenotypes (such as control of glucose levels) (Chen, Menghong Yan, *et al.*, 2016; Grandjean *et al.*, 2016). It will be key to address how different sRNAs delivered by sperm can influence preimplantation development in a pleiotropic manner to influence offspring metabolism.

Epididymosomes can reshape the RNA payload of sperm

Finally, a profound conundrum in paternal epigenetic inheritance has been the signaling pathway by which stressful environments experienced by the father can lead to changes in the sperm epigenome, whether via chromatin state (Siklenka *et al.*, 2015b; Ben Maamar *et al.*, 2018b), cytosine methylation (Radford *et al.*, 2014), or sRNAs (Gapp *et al.*, 2014a; Rodgers *et al.*, 2015). Our data suggest an alternative pathway that bypasses gametogenesis *per se*, but instead involves the post-developmental maturation of sperm during its transit through the epididymis (**Fig. 2.1-3,** Sharma et al. 2015). Our data uncover the temporal dynamics of small RNA biogenesis during post-testicular maturation, and strongly suggest a role for epididymosomes in transmitting small RNAs from somatic cells of the epididymis to maturing gametes. This is reminiscent of soma-to-germline communication previously observed in other organisms, although mostly associated with oocyte development (Castel and Martienssen, 2013). Future studies will shed further light on the role of the epididymis in sensing environmental conditions and the factors responsible for biogenesis and modulation of tRNA cleavage and/or sorting into epididymosomes in response to environmental conditions.

Materials and Methods

Mouse husbandry

Mice used in this study were primarily FVB/NJ strain background, obtained from Jackson Laboratories. All animal care and use procedures were in accordance with guidelines of the Institutional Animal Care and Use Committee.

Animals were raised on one of two diets – defined Control diet (Bioserv AIN-93g) or a Low Protein diet based on AIN-93g (10% of protein rather than 19%, remaining mass made up with sucrose) – as previously described (Carone *et al.*, 2010). Importantly, as we have found in natural matings that paternal dietary effects are substantially less penetrant when using females from our long-term mouse colony, we restricted all experiments here to females whose parents or grandparents had been obtained from the animal vendor.

As we have previously found that siblings, whatever their diets, are more epigenomically similar than are Control animals that are not siblings, all analyses are restricted to paired siblings. In other words, all dietary effects on small RNAs in sperm, testis, etc., were assessed only for pairs of littermates split to Control or Low Protein diet, or for pools of animals split to diets in which all animals in a given pool are matched one to one with littermates in the other pool. For example, in **Fig. 2.1**, the eight pairs of sperm samples include seven pairs of individual sibling males, and one pair of sperm pools with two animals in each pool, with all four animals in the two matched pools being from the same litter. Similarly, paternal dietary effects on offspring metabolism or preimplantation gene regulation always utilize male siblings on different diets as the sperm donors. Finally, the majority of analyses of early embryonic gene expression were carried out using clutches of oocytes from individual females (typically ~20-25 oocytes per female) that had been split into two groups for Control or Low Protein IVF (using sperm from sibling males), or for Control IVF embryos that were treated with and without tRF-Gly-GCC injection (using the same Control

sperm sample in both cases for IVF), and so forth. Experimental data were then combined for multiple such paired experiments.

Testis, epididymis, and sperm collection

Testes were dissected from 10-12 week mice fed on Control or Low Protein diet, directly frozen in liquid nitrogen and stored at -80 °C until RNA extraction. Cauda and Caput epididymides were dissected from mice and placed in Whitten's Media pH 7.4 (100 mM NaCl, 4.7 mM KCl, 1.2 mM KH2PO4, 1.2 mM MgSO4, 5.5 mM Glucose, 1 mM Pyruvic acid, 4.8 mM Lactic acid (hemicalcium), and HEPES 20 mM) at 37 °C. To collect caput sperm two incisions were made at the distal end of caput and using a 26G needle holes were poked in the rest of the tissue to let the caput sperm ooze out. For cauda sperm collection, cauda epididymides were gently squeezed to allow the caudal fluid to ooze out. After incubation for 15 minutes at 37 °C, sperm containing media was transferred to a fresh tube where they were incubated for another 15 minutes and the sperm-free epididymis tissues were directly frozen in liquid nitrogen and stored at -80 °C. After incubation at 37 °C for total of 30 minutes, sperm were collected by centrifugation at 2000 x g for 2 minutes, followed by a 1X PBS wash, and a second wash with lysis buffer (0.1% SDS and 0.5% Triton-X) for 10 minutes on ice to eliminate somatic cell contamination, and finally washed with 1X PBS before freezing down. Unlike cauda sperm, which are motile, caput sperm cannot swim up to allow a pure sperm sample collection. For the primary caput sperm dataset, sperm were obtained from 8 pairs of caput epididymis, pelleted, washed with PBS, and washed again with lysis buffer (0.1% SDS and 0.5% Triton-X) for

10 minutes on ice to eliminate somatic cell contamination. Sperm sample purity was confirmed by microscopic examination of the samples. For sperm "reconstitutions" caput sperm were purified over a Percoll gradient, washed with PBS, then either mock incubated at 37C or incubated with purified cauda epididymosomes at 37C as described in (Krapf *et al.*, 2012). After fusion reactions, sperm were pelleted, washed with PBS, pelleted again and snap frozen.

Epididymosome preparations

Epididymosomes were prepared as previously described (Belleannée *et al.*, 2013). In brief, gently extruded contents of the epididymis were centrifuged at 2000 × g to remove sperm, and the resulting supernatants were centrifuged at 10000 x g for 30 minutes to get rid of cellular debris. Next the supernatants were subjected to an ultracentrifugation at 120 000 × g at 4 °C for 2 h (TLA100.4 rotor; Beckman). Pellets were washed in cold PBS and subjected to a second ultracentrifugation at 120 000 × g at 4 °C for 2 h. These pellets were then resuspended in 50 µl of PBS and used for RNA extraction. A small aliquot was used for performing electron microscopy and Nanosight analysis of the epididymosomes.

RNA Extraction

For epididymosome RNA extraction, samples are thawed and the total volume of the sample was adjusted to 60 µl with filtered water (Macron). 33.3 µl of lysis buffer (6.4 M Guanidine HCl, 5% Tween 20, 5% Triton, 120 mM EDTA, and 120 mM Tris pH 8.0), 3.3 µl Proteinase K (>600 mAU/ml, Qiagen 19131),

and 3.3 µl water was then added to the sample. The sample was then incubated, with shaking, at 60 °C for 15 minutes on an Eppendorf thermomixer. One volume of water (100 µl) was then added and the sample transferred to a phase lock column (5 PRIME). For phase separation 200 µl of TRI Reagent (MRC inc) and 40 µl BCP (1-bromo-2 chloropropane, MRC inc) were added. The samples were then mixed by inversion 10-15 times, followed by centrifugation at 14,000 RPM for 4 minutes (a second addition of TRI reagent and BCP can also be performed for more pure RNA). The aqueous phase was then removed and transferred to a low binding RNase/DNase free microcentrifuge tube (MSP), followed by the addition of 20 µg of glycoblue (Ambion) and 1 volume (~200 µl) of Isopropanol. The RNA was then precipitated for 30 minutes or greater at -20°C, followed by centrifugation at 14,000 RPM for 15 minutes at 4°C, and one wash with 70% cold ethanol followed by centriguation at 14,000 RPM for 5 minutes at 4°C. Finally, the RNA was reconstituted in water (10 µl for gel size selection for small RNA cloning). For sperm RNA extraction, the same procedure was followed, apart from the addition of 3.3 µl 0.1M DTT with the lysis buffer and Proteinase K rather than the water. Also, prior to the incubation at 60 °C on the thermomixer the sperm pellet was disturbed physically using a pipette tip, and repeated pipetting. Finally, after the addition of the TRI reagent, sperm samples were vortexed for 5 minutes to ensure complete breakdown of the sperm. For testis and epididymidis samples, tissues were vortexed in 3-5 volumes TRI reagent with glass beads for 30 minutes at 4°C. The samples were then briefly spun and removed from the

beads and BCP was added, followed by phase separation, and Isopropanol precipitation (as above).

Small RNA Cloning

Small RNA cloning was carried out as in (Gu *et al.*, 2012). Total RNA was combined with an equal volume of Gel Loading Buffer II (Ambion), loaded onto a 15% Polyacrylamide with 7M Urea and 1X TBE gel, and run at 15W in 1X TBE until the dye front was at the very bottom of the gel (~25 minutes for Criterion minigels). After staining with SYBR Gold (Life Technologies) for 7 minutes, and destaining in 1X TBE for 7 minutes, gel slices corresponding to 18-40 nucleotides were then cut from the gel. Gel slices were then ground (using a pipette tip or plastic pestle) and 750 µl of 0.3 M NaCl-TE pH 7.5 was added and incubated with shaking on a thermomixer overnight at room temperature. The samples were then filtered using a 0.4 µm Cellulose Acetate filter (Costar) to remove gel debris. The eluent was transferred to a new low binding microcentrifuge tube and 20 µg of glycoblue and 1 volume of Isopropanol (~700 µl) were added. Samples were precipitated for 30 or more minutes at -20 °C. Size selection of the small RNAs was then followed by the ligation of a 3’ adaptor and then a barcoded 5’ adaptor as described in (Gu *et al.*, 2009). The libraries were then converted to DNA using Superscript III (Invitrogen) and amplified by sequential rounds of PCR, to first add short primer tails and then longer primer tails, providing the products with the correct adaptor sequences for deep sequencing. Libraries were subsequently sequenced by an Illumina HiSeq 2000 at the Umass Deep Sequencing Core.

Normalization and data analysis

For each small RNA library, rRNA-mapping reads (which were highly abundant in testis and epididymis samples, but rare in epididymosome and sperm samples) were removed. Remaining reads were mapped to murine tRNAs, to the unique sequences present in the 467 defined pachytene piRNA clusters (Li *et al.*, 2013), to Repeatmasker (tRNA entries from Repeatmasker were deleted to avoid duplicating tRNA-mapping reads), to miRbase, and to Refseq, yielding 27385 data points for each library (Sharma et al. 2015 **Table S1-S2**). piRNAs are defined as reads mapping to Repeatmasker along with reads mapping to the unique piRNA clusters. In this analysis pipeline, all reads mapping to a given entity – all fragments of the Lcn5 mRNA, for example, or all the various size fragments of tRNA-Gly-GCC – are accumulated into a single data point. Importantly, in the case of multiple tRNA fragment species – 23 vs. 27 nt tRF-Gly GCC, for example – various length species from the same tRNA typically behaved concordantly in comparisons, justifying this grouping. Non rRNA-mapping reads were normalized to parts per million mapped reads for subsequent analyses. For scatterplots, individual tRNA genes were grouped based on the uniqueness of the 5' 30 nucleotides – of the 14 tRNA-Gly-GCC genes in the mouse genome, 11 have indistinguishable 5' ends, so data for these 11 tRNA-Gly-GCC genes were merged for all scatterplots. This results in 190 unique 5' tRF points for most analyses. For analysis in **Fig. 2.4**, we accumulated all reads for a given anticodon for the sake of visual clarity. In **Fig. 2.4E**, individual samples with values more than 3 standard deviations away from the

mean abundance of that RNA across 16 sperm samples were adjusted to 3 standard deviations away from sperm mean value prior to summing read counts for the scatterplot, to minimize the visual impact of a small number of RNA species with extreme outlier values. For heatmaps **Fig. 2.4F**, only RNAs with a mean abundance >50 ppm are shown.

ATAC-seq libraries were mapped to mm10 using Bowtie2 (v2.3.2) with the following parameters: -D 15 -R 2 -N 1 -L 20 -i S,1,0.50 --maxins 2000 --no-discordant --no-mixed. Fragment lengths were separated using Python, and coverage of reads in various chromatin states was analyzed in R (v3.4.1) using data from ChromHMM (https://github.com/guifengwei/ChromHMM_mESC_mm10). All coverage data was normalized by global read depth prior to further analysis. Circos plots (v0.69-2) were generated from coverage data calculated by Bedtools (v2.25.0).

RNA-seq libraries were demultiplexed using Novobarcode (v3.02.08). Single end libraries were trimmed of 3' adapters using Fastx-toolkit (v0.0.14). Quantification was done using RSEM (v1.2.29) to RefSeq GTF annotation, mapped with Bowtie (v1.0.0) to mm10 using default parameters.

TaqMan assays

tRF and miRNA quantification was performed using custom designed TaqMan MicroRNA Assays according to manufacturer's recommended protocols (Applied Biosystems). 10 ng of total RNA isolated from mice as described above were reverse transcribed using the TaqMan MicroRNA reverse transcription kit. q-RT-PCR was performed in 15 µL reactions using TaqMan Universal PCR

Master Mix, following standard program (10 min at 95 °C, then 15 sec at 95 °C and 1 min at 60 °C for 40 cycles). Serial dilutions of template were run to confirm amplification efficiency of all TaqMan probes.

ES cell culture and transfection

E14 ESC lines were grown in DMEM (Gibco), and transfections were carried out in in OptiMEM in 6 well plates (28), with 9.5cm2 wells of ES cells seeded at a density of 3.5X105 cells/mL. 1 ng of antisense LNA containing oligonucleotides (Synthesized by Exiqon) were transfected using Lipofectamine 2000 (Invitrogen) for 16 hours, then ESCs were allowed to recover for 32 hours. Controls included lipofectamine only (Mock) and anti-GFP shRNA transfections. RNA extraction was performed at the end of 48 hours using standard Trizol protocol. RNA extracted from mouse ES cells was prepared for hybridization on Mouse GeneChip 2.0 ST arrays (Affymetrix) using the GeneChIP WT PLUS kit from Affymetrix.

Interfering with tRF-Gly-GCC function

Two distinct oligonucleotides were used in this study to interfere with tRF-Gly-GCC function. For the majority of ES transfections, we used an LNA-containing antisense oligonucleotide with the following sequence: 5'-ACCACTGAACCACCAA-3'. We also noted that an unmodified antisense oligonucleotide – 5'-GCG AGA AUU CUA CCA CUG AAC CAC CAA UGC-3' – could derepress MERVL targets in ES cells (not shown). We therefore used this antisense in zygote microinjections, as the different lengths and ends of this oligo relative to the LNA make MERVL repression unlikely to be a gain of function of

both antisense oligos. In addition, preliminary microinjections (n=3 embryos each) using the LNA antisense also resulted in upregulated MERVL target expression in 4-cell embryos (not shown), giving us confidence that both approaches to tRF-Gly-GCC inhibition work in both ES cells and intact embryos.

MERV-L reporter construction and FACS

The 2C::tdTomato construct described in (Macfarlan *et al.*, 2011) was obtained from Addgene, and contains the MERVL LTR nt 1-730 ligated into pcDNA3 hygro tdTomato plasmid. Sp110::tdTomato and MERVK::tdTomato were created from digested 2C:tdTomato using the same restriction sites, and the inserts were made by specific PCR amplification of the MERV-L LTR upstream of Sp110, and a MERVK LTR element located 5' of Tdpoz2, using nested PCR primers to the genomic sequences. To derive stable cell lines, E14 9 mESCs were transfected with 2 µg of plasmid using Lipofectamine 2000 (Invitrogen), and selected using 150 µg/mL of hygromycin for 7 days. Surviving cells were re-plated into a 100 mm dish at low density, and tdTomato+ single colonies were picked and expanded, and confirmed using PCR. Stable cell lines were transfected with either LNA antisense to tRF-Gly-GCC (Exiqon) or anti-GFP esiRNA (described above), then sorted by FACS 48-hours post-transfection at the UMass Medical School Flow Cytometry Core, using a FACSAria II Cell Sorter (BD).

Pol2 chromatin immunoprecipitation

Pol-II antibodies used: sc-899X (Santa Cruz, n=3 replicates) and ab5095 (Abcam, n=2 replicates). 2 µg of antibodies were conjugated to 50 µL of

Dynadeads M-280 overnight with 0.1% BSA in PBS. 1X107 mESCs transfected with anti-tRF-GG LNAoligos and anti-GFP esiRNA were fixed in 1% formaldehyde for 5 mins at RT. Cells were lysed in 650 µL of nuclei lysis buffer (50 mM Tris-HCl pH 8, 10 mM EDTA pH 8, 1% SDS, 1X protease inhibitor cocktail) for 10 mins on ice. Chromatin was sheared on a Covaris S220 using the following settings: peak power 105, Duty Factor 2, 200 cycles, 300 seconds per 130 µL microtube. 130 µL of chromatin (~2X106 cells) was diluted with 900 µL of ChIP dilution buffer (50 mM Tris-HCl pH 8, 0.167 M NaCl, 1.1% TritonX100, 0.11% sodium deoxycholate) and 500 µL RIPA-150 (50 mM Tris-HCl pH8, 0.15 M NaCl, 1 mM EDTA pH 8, 0.1% SDS, 1% Triton-X100, 0.1% sodium deoxycholate). An appropriate volume was also saved for input DNA, and extracted as below. Chromatin was added to Pol-II antibody conjugated Dynabeads and gently rotated for 4 hrs at 4 °C. Beads were washed once in RIPA-150, twice in RIPA-500 (50 mM Tris-HCl pH 8, 0.5 M NaCl, 1 mM EDTA pH 8, 0.1% SDS, 1% Triton-X100, 0.1% sodium deoxycholate), twice in RIPA-LiCl2 (50 mM Tris-HCl pH 8, 1 mM EDTA pH 8, 1% NP40, 0.7% sodium deoxycholate, 0.5 M LiCl2), and twice with TE buffer pH 8, 5 mins at 4 °C on rotator per wash. Beads were eluted with 200 µL elution buffer (10 mM TrisHCl pH 8, 0.3 M NaCl, 5 mM EDTA pH 8, 0.5% SDS) with 1 µL RNase A (Qiagen) at 65 °C for 4hrs. Supernatant was removed from beads then incubated overnight at 55 °C with 1 µL Proteinase K (20 mg/mL). Finally, DNA was extracted using phenolchloroform isoamyl alcohol, and precipitated with isopropanol and glycogen. DNA concentrations were determined using Qubit (Thermo-Fisher), then 0.5 ng of

DNA was loaded for each qPCR reaction. Data in Fig. S9C were normalized to IgG control ChIPs, then the ratio was calculated for tRF-GG inhibition vs. GFP controls.

Metabolic labeling

E14s were labeled in 500mM 4sU containing media for 15 or 30mins, then RNA was isolated using Trizol and isopropanol precipitation. 50 µg of total RNA was mixed with 0.2mg/ml of EZ-Link Biotin-HPDP (Themo Fisher) in 500 µL reaction then incubated for 2hr at 37°C in a shaking thermomixer (750 rpm). Biotinylated RNA was then extracted using phenol:chloroform:isoamyl alcohol (PCI) with phaselock gels and precipitated using isopropanol. RNA pellet was resuspended in 10µL water, and mixed with 30µL of washed Dynabeads MyOne Streptavidin C1 beads (Invitrogen) in binding buffer (10 mM Tris-HCl pH7.5, 300 mM NaCl, 0.1% Triton-X). The slurry was rotated for 20 mins at room temperature to immobilize biotin-tagged RNA, then placed on magnetic stand and washed with 500µL high salt buffer (50 mM Tris-HCl, 2M NaCl, 0.5% Triton-X). The supernatant from the first high-salt wash is the unlabeled total RNA. Beads were then stringently washed in high salt buffer, then two times in binding buffer, then once in low salt buffer (5 mM Tris-HCl, 0.1% Triton-X). The biotin-tagged RNA were extracted from the beads with 100 mM DTT at 65°C for 5 mins twice. Finally, labeled and unlabeled fractions were PCI extracted and RNA was isopropanol precipitated, and used to construct RNA-seq libraries.

RNA-seq for 4sU-labeled Libraries

5 µg of total RNA was depleted of ribosomal RNA using Ribo-Zero rRNA Removal Kit (Human, Mouse, Rat, Illumina). Less total RNA was used as input from metabolic labeling experiments for both 4sU labeled and unlabeled fractions. Illumina deep sequencing compatible libraries were constructed from rRNA-depleted RNA using an optimized version of a protocol described by Heyer et al. adding a purification using the RNA Clean and Concentrator (Zymo Research) in between procedures. Ribosome profiling data was published previously (GSE74537). Briefly, cell extract was incubated with Rnase A+T1, separated on a sucrose gradient to obtain 80S ribosomes, RNA was isolated by phenol-chloroform and ran on a gel to acquire 26-32nt RNAs. RNAseq libraries were made from RNA isolated using Trizol + isopropanol precipitation. Libraries were quantified, multiplexed and either single-end or paired-end sequenced on Illumina NextSeq 500 sequencer.

ATAC-Seq

E14 mES cells were transfected with antisense tRF-Gly-GCC LNA or mock-transfected and grown for 24 or 48 hours prior to harvesting and counting. ATAC-seq protocol was done essentially the same as described in Buenrostro et al. Briefly, after titration, 4 µL of TDE1 was determined as sufficient for 50000 cells. Tagmented DNA was amplified using Kapa HiFi Hotstart polymerase, and libraries were cleaned up using Ampure XT DNA beads. Libraries were quantified, multiplexed and paired-end sequenced on an Illumina NextSeq 500 sequencer.

In vitro fertilization

In vitro fertilization was performed according to “Manipulating the Mouse Embryo” Second Edition (Hogan *et al.*, no date). FVB/NJ mice were used as egg donors and sperm was isolated from males fed dietary regimes as above. Fertilization took place in 250 µL HTF media covered in mineral oil, pre-gassed in 5% CO2 at 37 °C.

Embryo RNA microinjection experiments

Zygotes for microinjection studies were generated by IVF using sperm from mice fed Control diet. After four hours of IVF, the zygotes were washed three times in HTF medium and placed in a drop of KSOM medium at 37 °C in 5% CO2 5% O2 for 2 hours. Embryos were then transferred to FHM medium containing 0.1% PVA, and subjected to 10 micromanipulation. Embryos were microinjected with either H3.3-GFP mRNA alone (Santenard *et al.*, 2010) (control group) or H3.3-GFP mRNA plus one of several synthetic tRNA fragments (experimental group), or with H3.3-GFP mRNA plus gel-purified small RNAs (18-40 nts) isolated from Control vs. Low Protein sperm. RNA injections were carried out using a Femtojet (Eppendorf) microinjector at 100 hPa pressure for 0.2 seconds, with 7 hPa compensation pressure. RNAs used for microinjections and their concentrations were: 100 ng/µl of H3.3- GFP mRNA, 200 ng/µl of tRF-Gly-GCC antisense RNA (5’GCG AGA AUU CUA CCA CUG AAC CAC CAA UGC 3’), 200 ng/µl of tRF-Glu-CTC sense RNA (5’ UCC CUG GUG GUC UAG UGG UUA GGA UUC GG 3’), and 200 ng/µl of tRF-Gly-GCC sense RNA with modified residues (5’ GCA JUL GUG GUU CAG UGG DAG AAU UCU CGC 3’ where J=2’-O-methyluridine, L=N2-methylguanosine, D=dihydrouridine). For sperm small

RNA microinjections, total RNA was extracted from either Control or Low Protein sperm, 18-40 nucleotides RNA was size selected on a gel and purified for microinjections, used at either 0.5 ng/µl (Fig. 2.5D). After the microinjections, embryos were placed back into culture and H3.3-GFP fluorescence was verified at the 2-cell stage. GFP-positive injected embryos were cultured until the late 2-cell stage (32 hours post IVF for most injections, 28 hours for the tRF-Gly-GCC injections in Fig. 2.5E) or 4-cell stage (tRF-Gly-GCC antisense RNA microinjections, Fig.2.4E-F), at which point embryos were collected and processed for single-embryo RNA sequencing.

Intracytoplasmic Sperm Injection (ICSI)

The testes and epididymis of 8-12 week old FVB/NJ mice were dissected into PBS. For rete sperm isolation, the efferent duct leading from the rete testis to the caput epididymis was located and the testis was removed under a stereo microscope and placed into a fresh dish containing PBS. After 2 washes with PBS, several cuts in the Rete testis were made, releasing its contents including the rete sperm. Released rete testis contents were transferred into an eppendorf tube, spun at 14,000 RPM for 2 minutes, and washed twice with modified nuclear isolation medium (NIM) with 1% polyvinyl alcohol. Sperm were finally resuspended in 500 µl NIM 1% PVA for use in ICSI. For cauda sperm isolation, the cauda epididymis was isolated and placed in PBS. The cauda sperm were then released by making an incision in the cauda epididymis followed by squeezing to release the epididymal contents. The cauda sperm were then spun and washed as the Rete sperm. For sperm heads (both rete and cauda), after

collecting the sperm in PBS in an eppendorf tube, the sperm were spun at 14,000 RPM and washed once with PBS. The sperm were then resuspended in 500 µl PBS and then drawn through a 26G needle on a 1 ml syringe between 20-30 times. The shearing force from being drawn in and out of the needle removes the sperm head from tail for the majority of sperm. The sperm were then washed twice in NIM 1% PVA and finally resupended in 100-500 µl NIM 1% PVA for use in ICSI. Females were superovulated by an intraperitoneal (i.p) injection of pregnant mare's serum gonadotropin (PMSG; 5 IU) followed by an i.p. injection of human chorionic gonadotropin (hCG; 5 IU) 48 hours later. Eggs were then collected from the oviducts of the females 13-16 hours later by placing the dissected ampulla of the oviduct 11 into KSOM containing 3 mg/ml hyaluranidase to digest the cumulus cells away from the eggs. After several minutes in hyaluranidase the eggs were washed 4-5 times in KSOM, finally being placed in KSOM in a 37 °C incubator until injected. For ICSI, plates were made with drops of NIM 1% PVA for washing the injection needle, drops of NIM 1% PVA with sperm, drops of FHM with 0.1% PVA for the eggs to be added to for injection, and finally covered with mineral oil. 10-15 eggs at a time were placed into the FHM + 0.1% PVA drops on the injection plate for subsequent injection. Sperm (heads or whole) were then picked and injected into the eggs. After completion of the 10-15 injections, the injected eggs were maintained at room temperature for 5 minutes, washed 4 times in KSOM, and then placed in 50 µl KSOM in a 37 °C 5% O2 incubator for development. The process was repeated for a total of 80-120 injections per day. 28 hours post-injection, 2-cell embryos were collected into

5 µl TCL buffer with 1% βME and then stored at -80°C for processing into single embryo RNA-sequencing libraries.

Single embryo RNA-Seq

Single embryo RNA-Seq was carried out using the SMART-Seq protocol, as described in (Ramsköld *et al.*, 2009; Shalek *et al.*, 2013). Data were mapped using RSEM, and were normalized to parts per million mapped reads after removing microRNA, snoRNA, and rRNA-mapping reads. Embryos with fewer than 10,000 detectable transcripts were removed from the dataset. For a given condition (Control, Low Protein, testicular sperm ICSI, etc.) we calculated the geometric mean of the mRNA abundance across all relevant embryos to minimize the influence of outliers, and filtered for genes with a geometric mean abundance >5 ppm. The cumulative distribution of the log2 fold change between different conditions is then plotted for various comparisons in Fig2.5. For analysis of MERVL target gene expression, we used all genes changing at least 2-fold in our ES cell Affymetrix and RNA-Seq data (Sharma et al. 2015 Tables S4-S5) for tRF-Gly-GCC targets, and we used the union of tables S2 and S6 from MacFarlan et al for remaining MERVL targets.

Immunofluorescence experiments and analysis of embryos

To determine the effects of different diets on cell fate in the mouse embryo, IVF experiments were performed as described above, using sperm from sibling males either on Control or Low Protein diets. 4 hours after IVF, embryos were washed in HTF medium and placed in KSOM drops for long-term culture, in a 5% CO2 5% O2 incubator. Embryos were collected at embryonic day 4 (E4) at

mid-blastocyst stage and fixed in 4% PFA for Cdx2 staining. Immunofluorescence staining was performed as described in (Torres-Padilla *et al.*, 2006). Primary antibody used was Anti-CDX2 (BioGenex) at 1:100. Secondary antibody used was AlexaFluor 488 goat anti-mouse IgG (Molecular Probes) at 1:500. After the final wash, stained blastocysts were mounted in a gradient of Vectashield mounting medium with DAPI, in drops to retain the three-dimensional structure of the blastocyst. Microscopy was performed on a Zeiss Axiovert 200 inverted microscope with Orca-ER camera (Hamamatsu, NJ), using a 40x/1.4 NA oil objective. Z-sections were taken every 3 μm through the entire embryo. Analysis and cell counting was performed using Zeiss AxioVision4.9.1 software.

CHAPTER III

The molecular function of tRNA fragment Gly-GCC

Abstract

Small RNAs derived from mature tRNAs, referred to as tRNA fragments or "tRFs", are an emerging class of regulatory RNAs with poorly understood functions in cellular regulation. We recently identified a role for one specific tRF – 5' tRF-Gly-GCC, or tRF-GG – in repression of genes associated with the endogenous retroelement MERVL, but the mechanistic basis for this regulation was unknown. Here, we show that tRF-GG plays a role in production of a wide variety of noncoding RNAs normally synthesized in Cajal bodies. Among these noncoding RNAs, tRF-GG regulation of the U7 snRNA modulates heterochromatin-mediated transcriptional repression of MERVL elements by supporting an adequate supply of histone proteins. Importantly, the effects of inhibiting tRF-GG on histone mRNA levels, activity of a histone 3' UTR reporter, and ultimately on MERVL regulation could all be suppressed by the U7 RNA. We show that the related RNA-binding proteins hnRNPF and H bind directly to tRF-GG, and are required for Cajal body biogenesis. Together, our data reveal a conserved mechanism for 5' tRNA fragment control of noncoding RNA biogenesis and, consequently, in global chromatin organization.

Introduction

It has been known for some time that mature tRNAs can be cleaved in response to cellular stressors (Lee and Collins, 2005), but only recently have the resulting cleavage products – broadly known as tRNA fragments, or tRFs – been

appreciated as potential regulatory molecules in their own right (Keam and Hutvagner, 2015). Although tRNA fragments have in some cases been reported to function in complex with Argonaute proteins and act effectively like microRNAs or endo-siRNAs (Deng *et al.*, 2015; Andrea J Schorn *et al.*, 2017; Martinez, Choudury and Slotkin, 2017), they have also been reported to have Argonaute-independent regulatory functions ranging from inhibition of translation to control of apoptosis (Elbarbary *et al.*, 2009; Zhang, Sun and Kragler, 2009b; Ivanov, Mohamed M. Emara, *et al.*, 2011; Gebetsberger *et al.*, 2012a; Mary T Couvillion *et al.*, 2012; Sobala and Hutvagner, 2013; Goodarzi, Liu, Hoang C B Nguyen, *et al.*, 2015; Molla-Herman *et al.*, 2015; Kim *et al.*, 2017). The diversity of proposed mechanisms for tRF function in part reflects the multitude of types of tRNA fragments that have been identified – 22 nt fragments derived from the 3’ ends of mature tRNAs have been found associated with Argonaute proteins (Kumar *et al.*, 2014b; Kuscu *et al.*, 2018) and have been suggested to direct cleavage of retrotransposon RNAs (Andrea J Schorn *et al.*, 2017; Martinez, Choudury and Slotkin, 2017), whereas longer (28-32 nt) fragments arising from tRNA 5’ ends appear to play more diverse mechanistic roles. For instance, 5’ fragments of valine tRNAs serve as global repressors of translation in archaea, yeast, and mammals, and in some cases appear to act by interfering with translational initiation (Gebetsberger *et al.*, 2012a; Bąkowska-Żywicka *et al.*, 2016; Guzzi *et al.*, 2018; Luo *et al.*, 2018).

We previously showed that interfering with a 5’ fragment of tRNA-Gly-GCC (hereafter, tRF-GG) using an antisense LNA oligonucleotide resulted, in

both murine ES cell culture and in preimplantation embryos, in derepression of ~50 genes associated with the long terminal repeat (LTR) of the endogenous retroelement MERVL (Sharma *et al.*, 2016). This functional link between a tRNA fragment and LTR element control is particularly interesting given the ancient and widespread role for tRNAs in LTR element replication – tRNAs almost universally serve as primers for reverse transcription of LTR elements (Marquet *et al.*, 1995) – as well as recent studies reporting that 3' tRNA fragments can interfere with multiple stages of the LTR element life cycle (Deng *et al.*, 2015; Andrea J Schorn *et al.*, 2017; Martinez, Choudury and Slotkin, 2017). In the case of tRF-GG-mediated control of MERVL elements, however, we find no identifiable homology between the 5' tRF-GG and either the LTR or the primer binding sequence of MERVL (which is primed by homology to Leucine tRNAs), making it unlikely that MERVL regulation occurs through homology-directed RNA targeting.

Here, we set out to uncover the mechanistic basis for repression of MERVL-associated genes by tRF-Gly-GCC. To our surprise, we find that control of MERVL elements is a downstream result of an evolutionarily-conserved function for tRF-GG in supporting noncoding RNA production. Manipulation of tRF-GG levels in human and mouse ES cells affects the levels of a wide range of noncoding RNAs, including snoRNAs, scaRNAs, and various U RNAs, all of which are normally produced in a subnuclear organelle known as the Cajal body. One such RNA, the U7 noncoding RNA, is essential for 3' UTR processing of histone pre-mRNAs. We show that tRF-GG control of U7 levels has downstream effects on histone mRNA and protein levels, as well as global chromatin

compaction, and that the effects of tRF-GG on histones and on MERVL target gene transcription can be suppressed by manipulating U7 snRNA levels. Finally, we identify the related proteins hnRNP F/H as direct binding partners for tRF-GG, and show that these RNA-binding proteins are required for normal Cajal body biogenesis and for repression of MERVL-driven gene expression. Taken together, our data reveal a novel pathway for tRNA fragment function in mammals, linking tRNA cleavage to regulation of noncoding RNA production.

Results

tRF-GG is a positive regulator of histone genes

To explore the mechanistic basis for global heterochromatin opening during tRF-GG inhibition, we turned to RNA-Seq and ribosome footprinting data (Sharma *et al.*, 2016) to identify potential effects of tRF-GG inhibition on expression of key chromatin regulators such as CAF-1. In addition, given the species-specific genomic locations of many ERVs such as MERVL, we also gathered RNA-Seq data in H9 human ES cells subject to tRF-GG inhibition, to identify conserved and divergent transcriptional consequences of tRF-GG inhibition. In both mouse and human ESCs, tRF-GG inhibition did not affect expression of any known chromatin regulators of MERVL such as CAF-1, Kdm1a, or Ehmt1. Intriguingly, tRF-GG inhibition in human ESCs had minimal effects on HERV expression, indicating that ERV regulation by this tRNA fragment is confined to mouse ES cells.

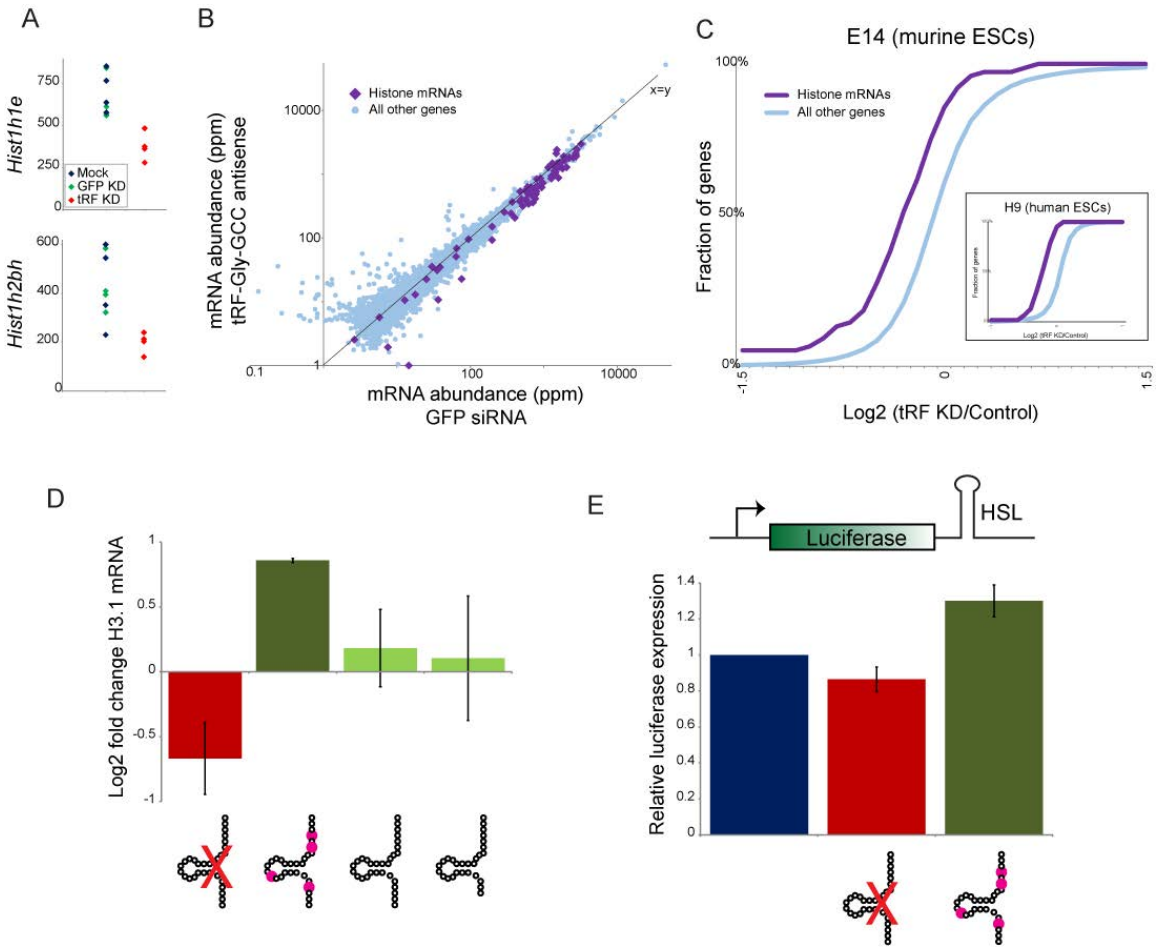


Figure 3.1. tRF-Gly-GCC represses expression of histone genes via the 3' histone stem loop. A) mRNA abundance for two example histone genes – *Hist1h1e* (top) or *Hist1h2bh* (bottom) – in four replicates of mock-transfected, GFP KD, and tRF KD ES cells, as indicated. **B)** Scatterplot comparing RNA abundance for histone genes (purple diamonds) and all other genes in GFP KS ES cells (x axis) and tRF-inhibited ES cells (y axis). Note that nearly all histone genes fall below the x=y diagonal here. **C)** Cumulative distribution of the effects of tRF inhibition on histone mRNA expression, with y axis showing cumulative fraction of genes exhibiting any given log2 fold change in expression (x axis). Main panel shows data from murine ES cells (n=4 replicates), while inset shows data for human ESCs. See also **Figure S3.2 A-C**. **D)** q-RT-PCR for *Hist2h3b* showing effects of transfecting the anti-tRF LNA, a synthetic tRF-Gly-GCC oligonucleotide bearing most of the expected nucleotide modifications present in tRNA-Gly-GCC, and two synthetic nucleotides without any modified nucleotides. However, although the modified synthetic tRF was more active in this assay, we found no significant effect on tRF modification in several other assays (not shown), so this was not pursued in further detail. **E)** tRF-Gly-GCC regulates the histone 3' UTR. We generated stable ES cell lines carrying a luciferase reporter bearing the 3' UTR of *Hist2h3b* (**Figure 3.2F** shows data for an independent cell

line bearing the *Hist1h4j* 3' UTR). Bar graph shows average changes to reporter activity in response to control KD, tRF LNA, or the modified tRF oligo.

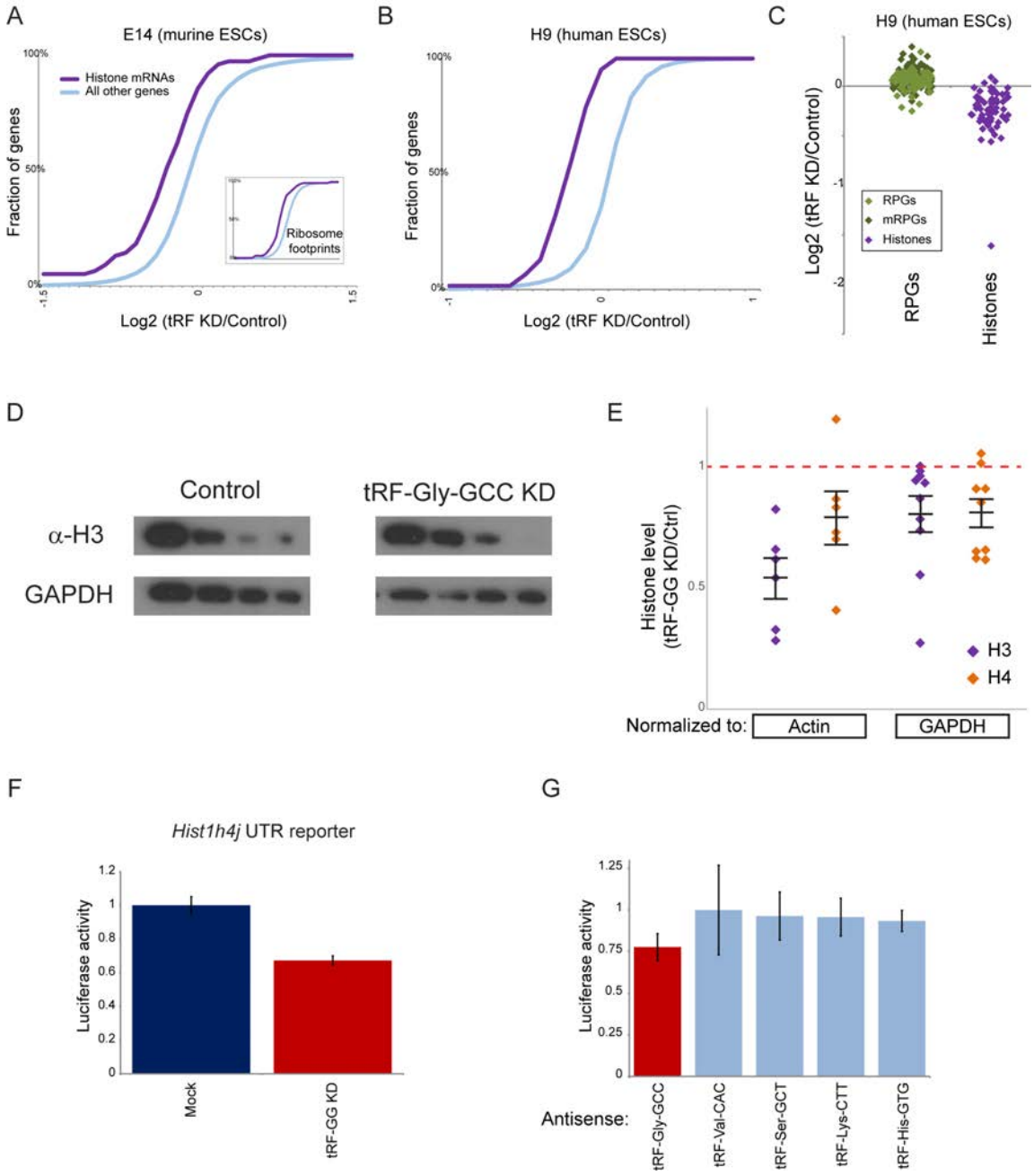


Figure 3.2. tRF KD effects on histone levels in human and mouse ESCs.

A-B) CDF plots for tRF-GG KD effects on histone gene expression in mouse (A) and human (B) ESCs. Inset in (A) shows effects of tRF-GG KD on ribosome occupancy of histone genes. Comparable effects on histone RNA and RBF levels suggest that tRF-GG does not directly regulate histone regulator SLBP, which regulates both histone mRNA processing/stability as well as its translation

{Marzluff, 2017 #2869}. **C)** Dot plot showing effects of tRF-GG knockdown on individual histone genes (these data are replicated in **Figure 3.1C**). **D)** Example Western blots for histone H3 and GAPDH in ES cells subject to control or anti-tRF-GG transfection. The four lanes show successive 2-fold dilutions. **E)** Quantitation of histone levels. Dots show H3 and H4 levels for individual experiments, normalized either to β-Actin or to GAPDH, as indicated. Associated lines show mean and s.e.m. for each experiment. **F)** As in **Figure 3.1E**, for stable ES lines carrying a dual luciferase reporter with Renilla luciferase fused to the *Hist1h4j* 3' UTR. **G)** Luciferase activity for the *Hist2h3b* UTR reporter, showing reporter activity in ES cells transfected with LNA antisense oligos targeting the indicated 5' tRFs. Bars show average and standard deviation, n=4 replicates.

Instead, we identified two conserved molecular phenotypes resulting from tRF-GG inhibition in both human and mouse ES cells: repression of histone mRNAs (**Figures 3.1A-C, Figure 3.2**), and decreased expression of a variety of noncoding RNAs (see below). Given our finding of global heterochromatin decompaction in tRF-inhibited cells (previous chapter), and the common derepression of ERV elements in undercompacted genomes (Lenstra *et al.*, 2011; Ishiuchi *et al.*, 2015a), we focus first on tRF-GG effects on histone genes. We confirmed by qRT-PCR that tRF-GG inhibition causes a decrease in histone mRNA abundance (**Figure 3.1D**), and that reduced histone mRNA levels are accompanied by a decrease in histone protein levels (**Figures 3.2D-E**). Importantly, we find that direct tRF "overexpression" via transfection of ES cells with a synthetic 28 nt tRF-Gly-GCC also resulted in *increased* histone mRNA abundance (**Figure 3.1D**), demonstrating that repression of histone genes observed in response to tRF-GG inhibition does not represent an off-target gain of function exhibited by our anti-tRF-GG LNA oligo. Together, our gain and loss of function studies demonstrate that tRF-Gly-GCC plays a conserved role in

histone mRNA expression, with the MERVL LTR representing a sensitized reporter for chromatin assembly specifically in murine ES cells.

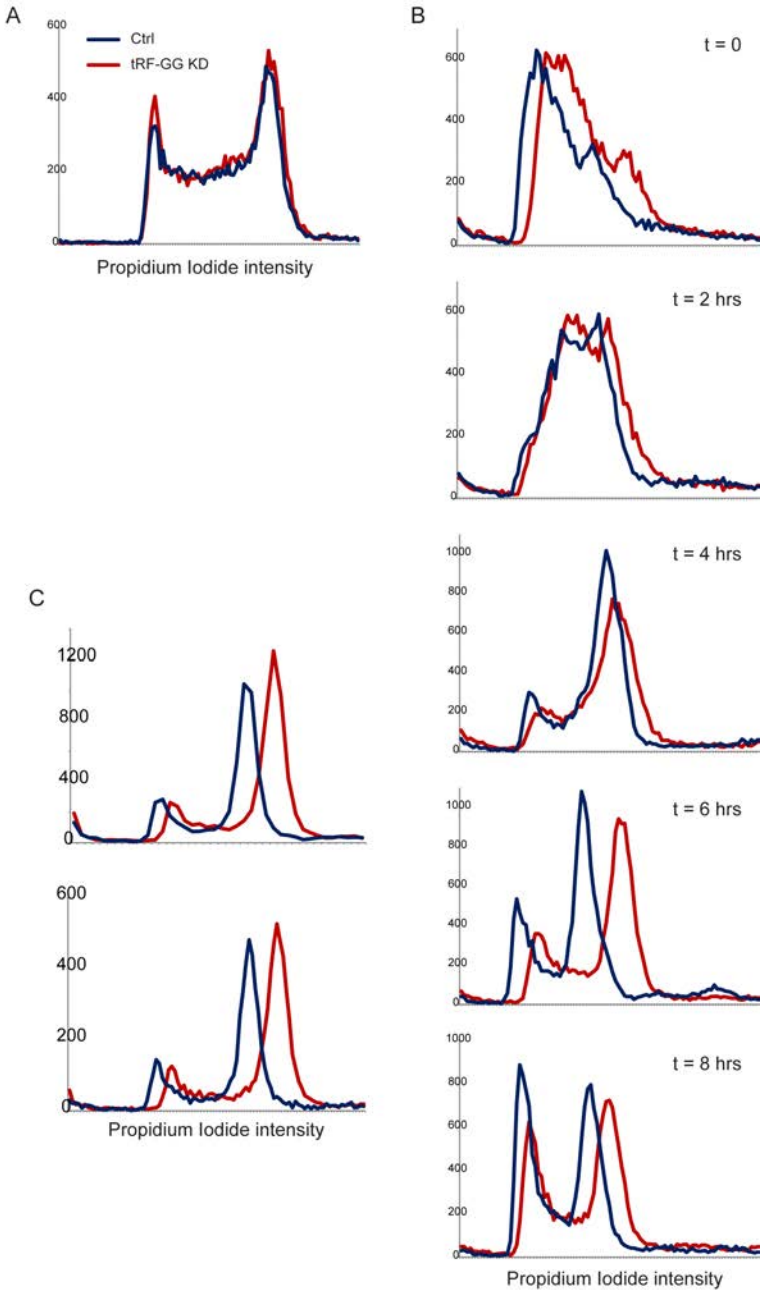


Figure 3.3 Cell cycle effects of tRF-Gly-GCC inhibition. A) Unsynchronized ES cells were either control or anti-tRF-GG LNA transfected, and 2 days later were characterized by FACS using propidium iodide staining. Nearly identical cell cycle profiles reveal no effect of tRF-GG inhibition on the fraction of S phase cells. **B)** Time course of S phase progression in the presence and absence of tRF-GG inhibition. ES cells were transfected either with anti-GFP siRNAs or with the LNA antisense to tRF-Gly-GCC, then arrested in G1 phase via thymidine block for 16 hours. FACS plots show DNA content at the indicated times after release from thymidine block. tRF KD cells release into S phase and proceed to G2/M with identical kinetics, although there is a subtle G2/M exit delay in the tRF KD cells reflected in the lower frequency of G1 cells at 6 and 8 hours post-release. Note that tRF KD cells exhibit a right-shifted peak of fluorescence intensity in G2/M – this is likely due to the greater availability of naked DNA for the DNA intercalator propidium iodide. Consistent with this, reanalysis of previous FACS data for SLBP deficient HCT cells (Jimeno-González *et al.*, 2015) also reveals increased fluorescence intensity in G2 cells. Taken together, our data do not support the hypothesis that

increased histone levels in tRF-inhibited cells arise from an increased fraction of S phase cells. **C)** Additional replicates for t=6 hours after release from thymidine block.

What is the mechanistic basis for tRF-GG-mediated repression of the histone genes? Histone expression is largely confined to the S phase of the cell cycle and could thus report on changes in cell cycle profile. However, FACS analysis of tRF-GG-inhibited ES cells revealed no change in the fraction of cells in S phase (**Figure 3.3**). Histone expression is highly regulated at levels from transcription to translation; perhaps the most unique feature of histone expression is the role of several cis-acting RNA elements in the histone 3' UTR – a short stem loop known as histone stem loop (HSL) that binds to stem loop binding protein (SLBP), and the histone downstream element (HDE) that binds to the U7 noncoding RNA and the associated U7 snRNP – in regulation of histone pre-mRNA processing (Dominski and Marzluff, 1999; Marzluff and Koreski, 2017). To separate the effects of tRF-GG manipulation on the histone 3' UTR from effects on the histone promoter or coding sequence, we generated stable ES cell lines carrying luciferase reporters fused to one of two histone UTRs (**Figure 3.1E, Figure 3.2F**). Transfection of synthetic tRF-GG drove increased luciferase activity (30%, $p = 0.0002$), while tRF-GG inhibition resulted in decreased luciferase levels (with values ranging from 14% to 32% in five separate experiments – each in at least triplicate – with p values ranging from 0.038 to 0.000019). tRF inhibition had no effect on a stable ES cell line carrying the wild-type luciferase reporter (data not shown), indicating that the histone 3' UTR is necessary to confer regulation. Moreover, loss of histone reporter activity

was specific to tRF-GG inhibition, as it was not observed in response to four other tRF-directed antisense LNA oligonucleotides (**Figure 3.2G**). We conclude from these data that tRF-GG regulates histone mRNA abundance via the histone 3' UTR.

tRF-GG affects histone expression and MERVL repression via control of U7 noncoding RNA

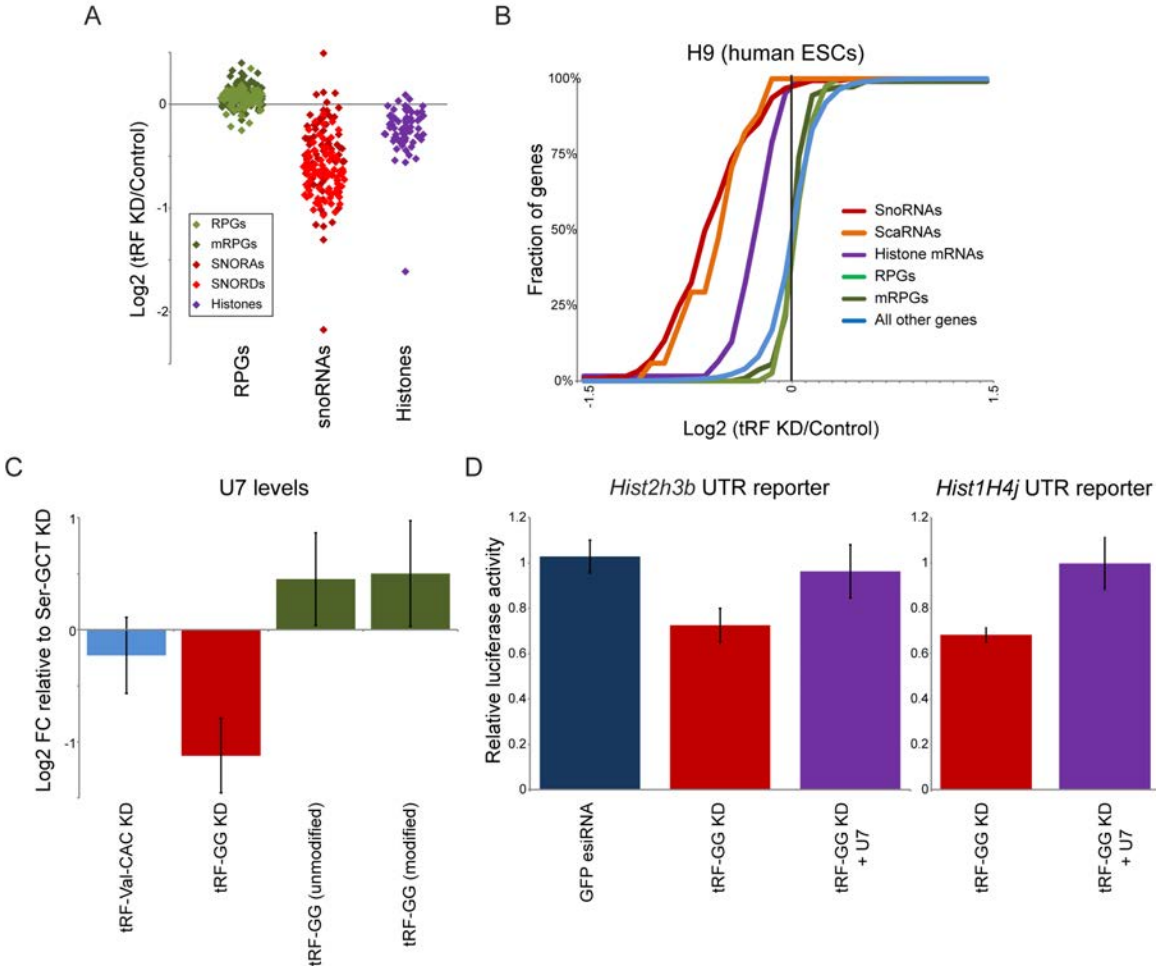


Figure 3.4. tRF-Gly-GCC supports production of U7 and other noncoding RNAs. A) Effects of tRF inhibition on various gene families in human H9 ES cells. Individual dots show individual species for the indicated families, illustrating the widespread downregulation of histone and snoRNA genes in response to tRF-Gly-GCC inhibition. Ribosomal protein genes are shown as a representative

highly-expressed but tRF insensitive gene family for comparison. Importantly, the role for tRF-GG in control of histone mRNAs and various noncoding RNAs was previously missed because prior analyses focused on polyadenylated mRNA abundance, while the current study analyzed rRNA-depleted total RNA samples (Methods). See also **Figure 3.5. B)** CDF plots showing the distribution of tRF effects on the indicated gene families, as in **Figure 3.1C**. **C)** Manipulating tRF-GG levels affects U7 ncRNA production. ES cells were transfected either with LNA antisense oligos targeting tRF-Ser-GCT, tRF-Val-CAC, or tRF-GG, or with synthetic tRF-GG oligos either bearing appropriate modified nucleotides (modified) or lacking these modifications (unmodified). U7 levels were quantitated by Northern blot (n=4), and normalized relative to 5S rRNA levels. Change in U7 levels is expressed relative to tRF-Ser-GCT inhibition, revealing a significant (p=0.03) decrease in U7 levels in response to tRF-GG inhibition, as well as modestly increased U7 levels in tRF-GGsupplemented cells. See also **Figure 3.6**. **D)** Effects of tRF-GG KD on histone 3' UTR reporters are suppressed by supplementation with additional U7 snRNA. ES cells were transfected with the LNA antisense to tRF-GG, with or without additional in vitro-synthesized U7 RNA. Effects of tRF-GG KD were significant (p = 0.0039 and 0.00013 for H3 and H4 reporters, respectively), while tRF-GG KD + U7 was statistically indistinguishable from control (p=0.24 and 0.48, respectively).

Proper biogenesis of the histone mRNA involves a complex ribonucleoprotein assembly of 3' UTR-associated proteins, as well as the noncoding U7 RNA which directs UTR processing via base pairing to the HDE of the histone 3' UTR (Marzluff and Koreski, 2017). Intriguingly, in addition to downregulation of histone genes, we noted that the other consequence of tRF-GG KD in both human and mouse ES cells was decreased expression of several major classes of noncoding RNA, including snoRNAs, and scaRNAs (**Figures 3.4 A-B, Figure 3.5**). Notably, all of these RNAs share a common biogenesis pathway with U7, as these RNAs are all processed in Cajal bodies (Wu and Gall, 1993; Gall, 2000; Machyna, Heyn and Neugebauer, 2013). To determine whether tRF-GG also affected levels of U7 RNA (which was not detected in our RNA-Seq data, potentially as a consequence of its secondary structure), we assayed U7

levels in tRF-GG KD and overexpression cells by Northern blotting (**Figure 3.4C**) and q-RT-PCR (**Figure 3.5B,D**). Consistent with the effects of tRF-GG manipulation on other noncoding RNAs, we find that inhibition of tRF-GG led to reduced U7 expression, while transfecting cells with the synthetic tRF-GG oligo supported higher expression of U7. Together, these findings reveal a conserved role for tRF-GG in promoting noncoding RNA production associated with Cajal bodies, and suggest that its effects on the histone 3' UTR result from altered U7 levels.

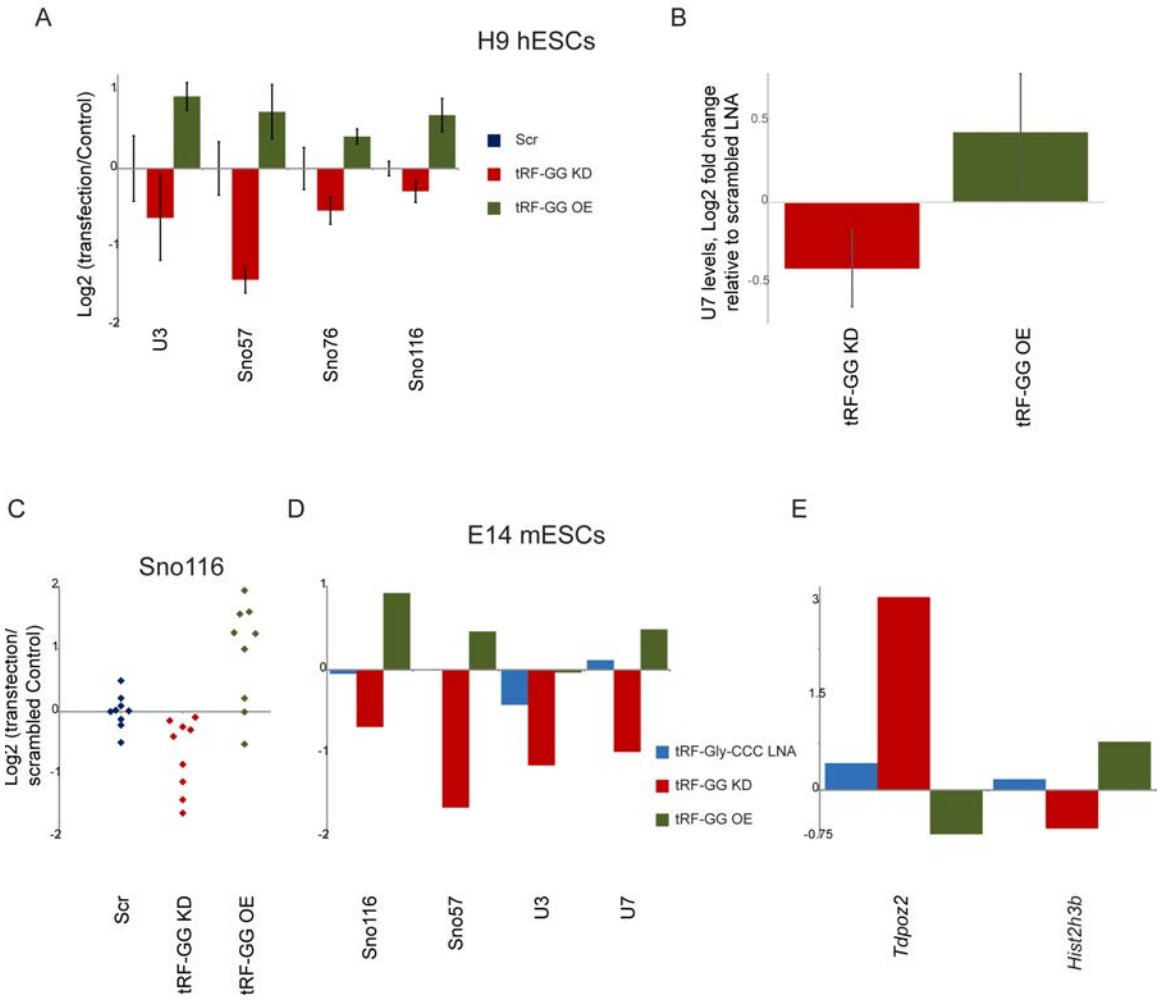


Figure 3.5. tRF-GG inhibition affects a variety of noncoding RNAs. A-B)

qRT-PCR for five noncoding RNAs, normalized relative to Actb. Average and standard deviation are shown for three replicate experiments in H9 human ESCs, with each experiment comparing transfections with the anti-tRF-GG LNA antisense, the modified synthetic tRF-GG oligo, or a scrambled oligo control (Methods). Y axis shows log2 fold change relative to the scrambled oligo control. The four ncRNAs in (A) were assayed in one experiment, while the U7 q-RT-PCRs represent an independent set of transfections and so are shown separately. **C-E)** Data for E14 murine ESCs. (C) shows individual replicates for 9 separate experiments comparing effects of scrambled oligo, tRF-GG antisense, and tRF-GG transfection, on Sno116 (normalized to Actb). (D) shows data for the four indicated noncoding RNAs (Actb-normalized), comparing tRF-Gly-CCC antisense, tRF-GG antisense, and tRF-GG to the scrambled oligo control. Data for Sno57, U3, and U7 were gathered in duplicates, while Sno116 data included 9 replicates for scramble, tRF-GG KD, and tRF-GG OE, and 2 replicates for tRF-Gly-CCC KD. (E) shows data for Tdpoz2 and Hist2h3b in the replicate pairs shown in (D), with the overexpression of these genes in tRF-GG KD cells providing additional replication of the findings documented in Fig.3.1 and the previous chapter.

The hypothesis that tRF-GG control of U7 levels is responsible for changes in histone and MERVL expression makes the prediction that manipulating U7 levels should suppress tRF-GG effects on histone expression and MERVL targets. We therefore transfected our histone 3’ UTR mESC line with the anti-tRF-GG LNA – which results in decreased U7 levels – with or without supplementation of additional U7 RNA, and assayed luciferase activity and MERVL target gene expression by qRT-PCR. Restoring U7 levels in tRF-GG KD cells reversed the inhibition of histone expression in these cells as assayed by both luciferase reporters (**Figure 3.4D**) and by qRT-PCR (**Figure 3.6A**). The converse also held true – antisense oligonucleotides directed against U7 were able to reverse the increase in histone levels in ES cells transfected with excess tRF-GG (**Figure 3.6A**). Importantly, restoring histone mRNA levels via U7 replenishment in tRF-GG-inhibited cells was able to partially suppress the transcriptional derepression of MERVL-linked genes as assayed both by qRT-

PCR and using a MERVL-driven fluorescent reporter cell line (**Figure 3.6B-D**). This supports a pathway in which MERVL repression is downstream of tRF-GG-mediated histone expression, rather than being secondary to the tRF's effects on snoRNA or other noncoding RNA production.

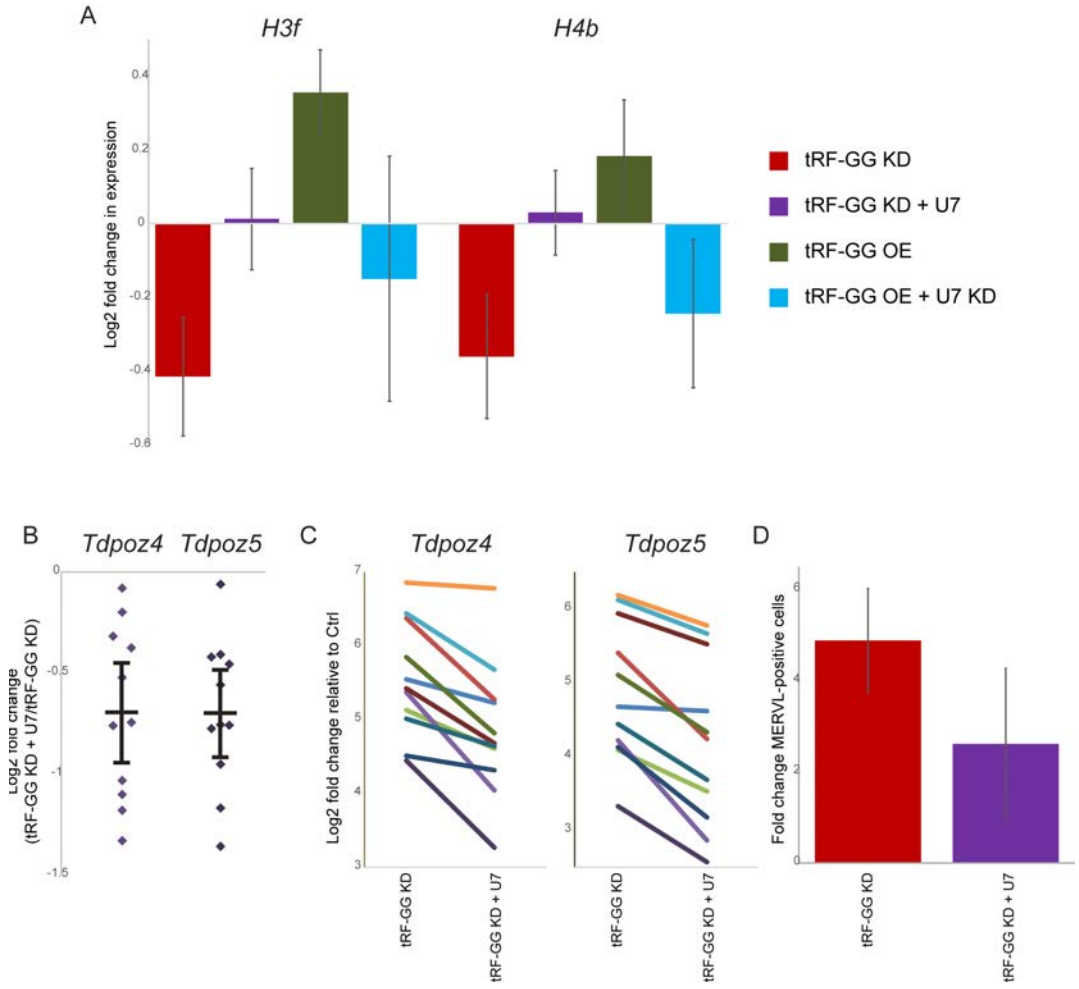


Figure 3.6. U7 suppresses effects of tRF-GG on histone mRNAs and MERVL targets. A) qRT-PCR for HIST3F and HIST4B (normalized to ACTB) in H9 hESCs transfected with scrambled LNA, LNA against tRF-GG with or without synthetic U7 RNA, or synthetic tRF-GG with or without antisense oligonucleotides targeting U7. Data show mean and standard deviation for three biological replicates. These data confirm the ability of tRF-GG to support histone mRNA expression, and show that the effects of tRF-GG – both positive and negative – on histone expression can be suppressed by appropriate manipulation of U7 ncRNA levels. **B-C)** U7 supplementation partially rescues the derepression

of MERVL targets in response to tRF-GG inhibition in murine ESCs. qRT-PCR data for Tdpoz4 and Tdpoz5, normalized to Actb, are shown for 11 biological replicates. **(B)** shows the relative effect of U7 on the background of tRF-GG inhibition (p=0.00015 and 4.9e-5 for Tdpoz4 and Tdpoz5, respectively), with dots showing individual replicates and line+whiskers showing mean +/- s.e.m. **(C)** shows the effects of tRF-GG KD, with or without U7 supplementation, relative to Control ES transfections – lines connect paired transfections of the anti-tRF LNA with or without U7 supplementation. Note that y-axis does not begin at zero. In all cases, we observe a modest but consistent rescue (average log2 FC of -0.7) of Tdpoz repression upon U7 supplementation. The partial rescue here could result from inefficient utilization of exogenously supplied U7 ncRNA, either due to the absence of modified nucleotides in the synthetic RNA, inadequate levels of U7, or exogenous U7 not being produced at the appropriate subcellular locus (the Cajal body). **D)** Fold change in tdTomato-positive MERVL LTR reporter ESCs following tRF-GG knockdown with or without U7 supplementation.

hnRNPF/H are tRF-GG binding proteins and are robust repressors of the 2C-like state

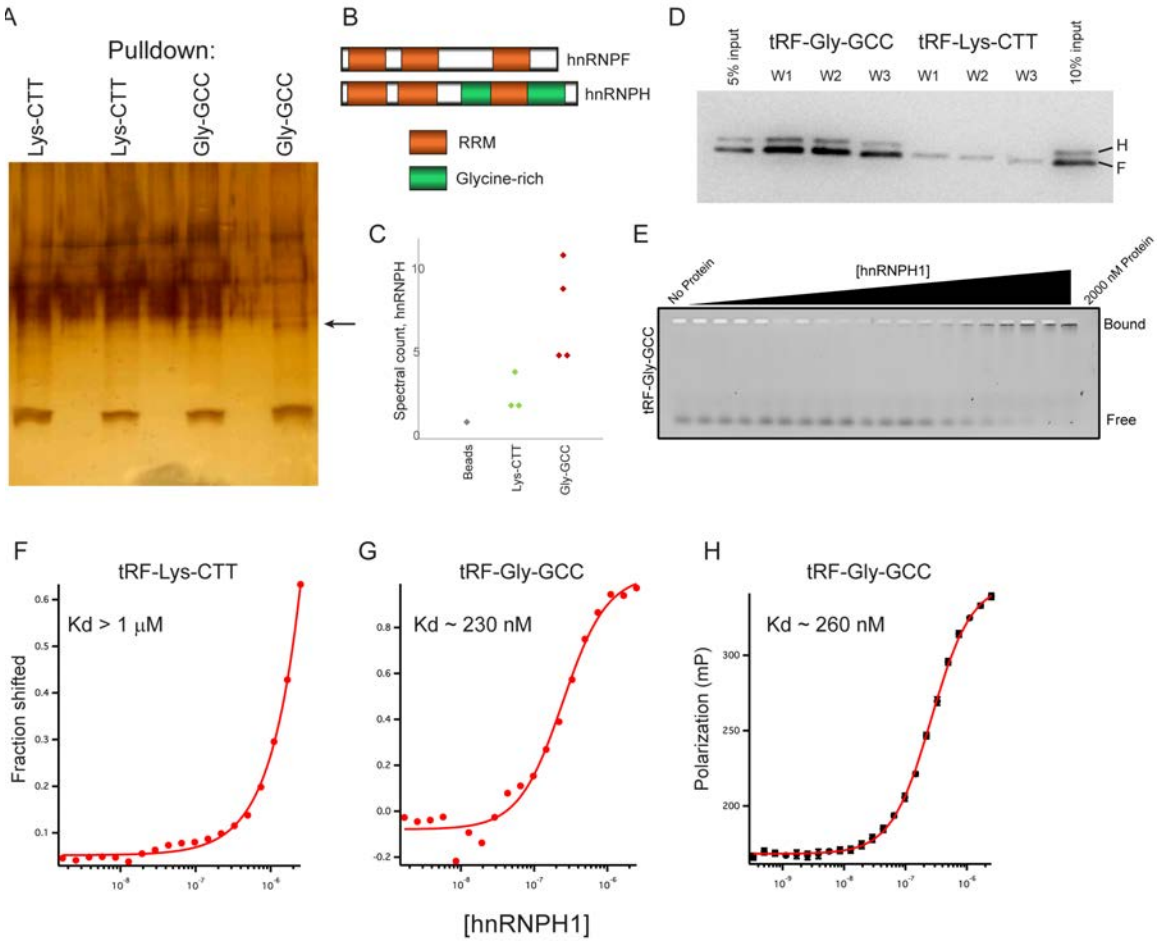


Figure 3.7. tRF-Gly-GCC binds to hnRNPF/H. A) Biotin-oligo pulldowns from murine ES cell extracts. Silver stained gel shows two replicates each for pulldowns using biotin-tRF-GG or biotin-tRF-Lys-CTT, as indicated. Arrow indicates ~50 kD band enriched in tRF-GG pulldowns. **B)** Domain architecture of hnRNPF and hnRNPH. **C)** Peptide counts for hnRNPH in control, tRF-Lys-CTT, or tRF-GG pulldowns. **D)** Western blots show hnRNPF/H recovery following tRF-GG or tRF-Lys-CTT pulldown. Pulldowns were washed 4 X 3 minutes with 50 mM Tris pH 8.0 supplemented with 100 mM (W1), 250 mM (W2), or 500 mM (W3) NaCl. **E)** Gel shift analysis of hnRNPH1 binding to tRF-GG. A synthetic oligonucleotide corresponding to tRF-Gly-GCC (GCAJULGUGGUUCAGUGGDAGAAUUCUCGC) was labeled at the 3' end using fluorescein 5-thiosemicarbazide, then incubated at 3 nM in equilibration buffer (0.01% Igepal, 0.01 mg/ml carrier tRNA, 50 mM Tris, pH 8.0, 100 mM NaCl, 2 mM DTT) for 3 hours along with increasing concentrations of purified hnRNPH1 protein from 1.35 nM to 2000 nM. See also **Figure 3.8**. **F-G)** Fit of gel shift binding data for tRF-Lys-CTT and tRF-Gly-GCC. Fitting the binding data yields an estimated Kd of hnRNPH1 of ~230 nM for tRF-Gly-GCC, and >1 mM for tRF-Lys-CTT. **H)** Fluorescence polarization data for hnRNPH1 incubations with labeled tRF-Gly-GCC. Polarization values against the protein concentrations are fit to the Hill equation using the Igor Pro software.

Finally, we turn to the question of how tRF-GG alters noncoding RNA production. To identify direct binding partners of tRF-GG, we used a 3'-biotinylated tRF-GG to isolate candidate tRF-binding proteins from mESC whole cell extracts. tRF-GG, but not the unrelated tRF-Lys-CTT oligo, pulled down a protein of ~50 kD (**Figure 3.7A**). We carried out four replicate mass spectrometry analyses, in three cases analyzing bulk tRF-GG or tRF-Lys-CTT pulldowns, and in one case first gel-purifying the tRF-GG-enriched ~50 kD band. Together, these efforts identified several potential binding partners enriched in tRF-GG pulldowns relative to the control tRF-Lys-CTT pulldown. Based on follow-up functional studies of several top candidates (Appendix II), we focus on the highly homologous pair of heterogeneous nuclear ribonucleoproteins (hnRNP) hnRNPF and hnRNPH (**Figures 3.7B-C**). These RNA-binding proteins function redundantly *in vivo*, and will therefore be referred to below as "hnRNPF/H". Using

an antibody that detects both hnRNPF and hnRNPH, we confirmed the interaction between tRF-GG and hnRNPF/H by Western blotting – hnRNPF/H was robustly detected in tRF-GG pulldowns, with only modest hnRNPF/H levels detected following tRF-Lys-CTT pulldown (**Figure 3.7D**). Furthermore, we expressed and purified hnRNPH1 protein, and carried out quantitative gel shift and fluorescence polarization analyses of hnRNPH1 binding to a fluorescently-labeled tRF-GG oligonucleotide. Both assays revealed specific binding between hnRNPH1 and tRF-GG, in contrast to the nearly undetectable binding observed for tRF-Lys-CTT (**Figure 3.7E-H, Figure 3.8**). We note that the apparent K_d for hnRNPH1 binding to tRF-GG is roughly 5-fold weaker ($K_{d,\ app}$ ~230 nM) than that for a positive control – a previously-described hnRNPF/H binding site identified in the SV40 pre-mRNA (Alkan, Martincic and Milcarek, 2006) (K_d ~50 nM). Given that hnRNPF/H are abundant enough to regulate targets bound with 50 nM affinity, and the potential for tRF-GG to occur at much higher abundance than typical pre-mRNA targets, the apparent K_d of hnRNPH1 for tRF-GG is well within a plausible physiologically-functional range.

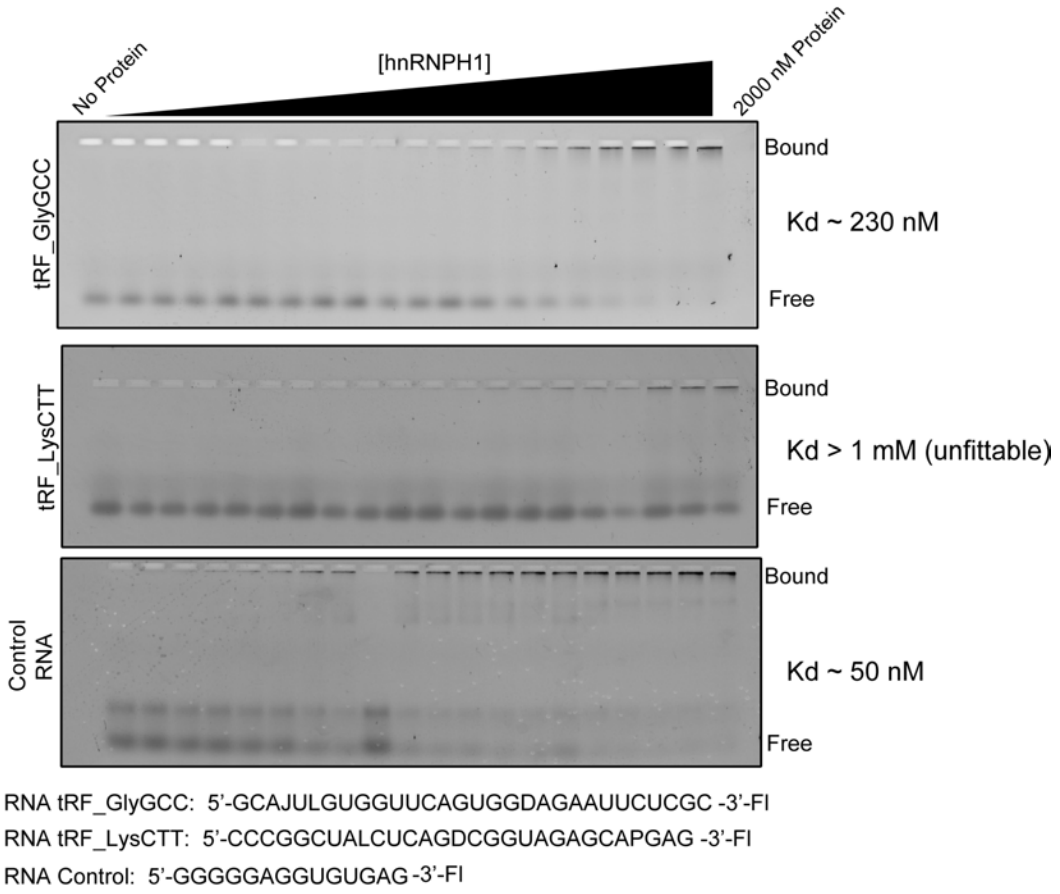


Figure 3.8. Direct binding of hnRNPH1 to tRF-Gly-GCC. Quantitative gel shift for hnRNPH1 binding to tRF-Gly-GCC is reproduced from Figure 3.7E, along with binding data for tRF-Lys-CTT and a positive control oligo (Alkan et al., 2006). Although hnRNPH1 exhibits moderately (~5-fold) weaker binding to tRF-GG than to the positive control, two considerations suggest that this binding could potentially be physiologically relevant. First, the positive control oligo was identified in a 3' UTR shown to be regulated by hnRNPF/H. Depending on cellular growth conditions, a given tRNA fragment could easily exceed the concentration of any given 3' UTR, so the cellular concentration of hnRNPF/H is clearly high enough for this affinity to fall into the physiologically-relevant range. Second, we note that our in vitro binding studies were performed with a synthetic 28-nt oligonucleotide bearing a FITC-modified 3' end, while we recently found that the majority of 5' fragments of tRNA-Gly-GCC in murine epididymis and sperm are 31-nt species bearing a cyclic 2'-3' phosphate at the 3' end (Sharma *et al.*, 2018). It is thus plausible (but difficult to test given the reliance of our binding assay on 3'-modification) that hnRNPF/H binding to the naturally produced tRFGG is of greater affinity than that documented here.

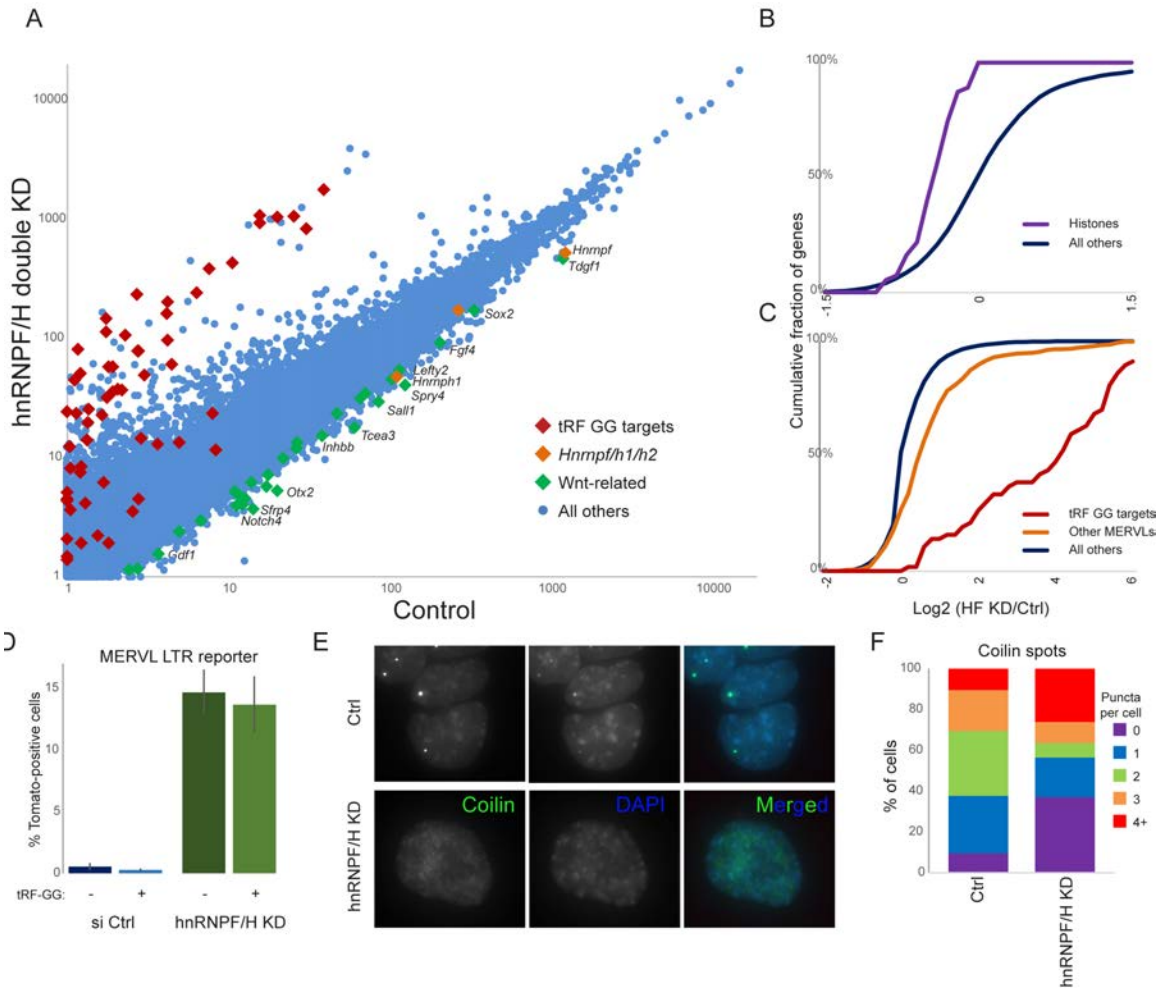


Figure 3.9. hnRNP F/H represses the MERVL program. A) Changes in the ES cell transcriptome following hnRNPF/H knockdown. Scatterplot shows mRNA abundance compared between control KD cells (x axis) and hnRNPF/H KD cells (y axis). Data are shown as median mRNA abundance for three biological replicates. **B)** hnRNPF/H KD results in histone mRNA downregulation. Cumulative distribution plot shows log2 fold change (hnRNPF/H KD/Ctrl) for histone genes, and all other genes, as indicated. **C)** Effects of hnRNPF/H KD on expression of the MERVL program. Cumulative distribution plots show log2 fold change (hnRNPF/H KD/Ctrl) for tRF-GG target genes (Sharma *et al.*, 2016), other genes previously associated with the MERVL program (Macfarlan *et al.*, 2012b), and all other genes, as indicated. **D)** hnRNPF/H suppresses ES cell entry into the 2C-like state. ES lines carrying a MERVL LTR-driven tdTomato (Macfarlan *et al.*, 2012b)were subject to control or hnRNPF/H KD, with bars showing mean +/- standard deviation (n=5 replicates) of the percentage of Tomato-positive cells. **E)** hnRNPF/H is required for normal Cajal body morphology and gross chromatin architecture. Panels show typical images for the Cajal body marker coilin (green), the MERVL LTR-driven Tomato reporter (red), or DAPI (blue), in control or hnRNPF/H KD ES cells. See also **Figure 3.10**.

F) Quantitation of Cajal body number per cell. Stacked bars show the percentage of cells exhibiting 0 through 4+ coilin-positive puncta, as indicated. Data represent the average of two replicate transfections, with 85 total control and 98 hnRNPF/H KD cells quantitated across the two experiments.

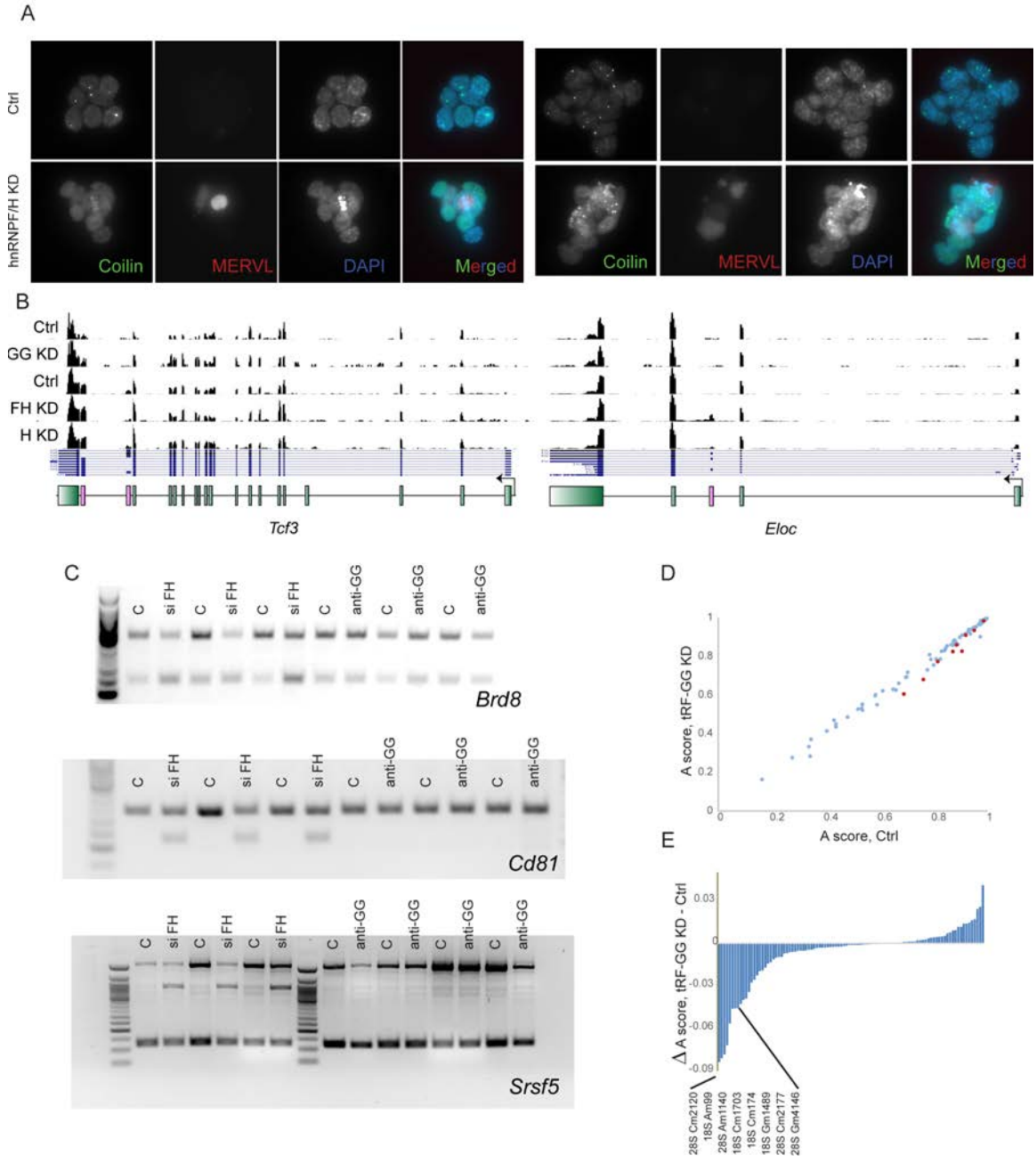


Figure 3.10. Effects of hnRNPF/H on Cajal bodies and MERVL expression.
A) Immunofluorescence images of MERVL LTR:tdTomato ES cells transfected with siRNAs against either GFP or against hnRNPF and hnRNPH. Images show anti-Coilin IF, tdTomato fluorescence, DAPI counterstaining, and all three merged, as indicated. Interestingly, in preliminary immunofluorescence experiments using our MERVL LTR reporter cell line, we observed that

hnRNPF/H is undetectable in spontaneously arising tdTomato-positive cells (data not shown), suggesting that hnRNPF/H degradation could potentially be involved in spontaneous ESC transitions to the 2C-like state. **B)** Browser shot of Tcf3 and Eloc mRNA abundance in ES cells subjected to the indicated knockdowns. Alternative exon inclusion events are shown in red. **C)** Validation of hnRNPF/H-regulated alternative splicing events by semiquantitative PCR. Three alternative splicing events – hnRNPF/H-dependent skipping of cassette exons in Brd8 and Cd81, or an hnRNPF/H-dependent detained intron in Srsf5 – identified in our hnRNPF/H RNA-Seq data are validated here by PCR. None of these alternative splicing events was recapitulated in the tRF-GG KD samples, confirming that tRF-GG only affects a subset of hnRNPF/H functions. **D-E)** Altered rRNA 2'-O-methylation in tRF-GG KD cells. We carried out Ribometh-seq and calculated methylation "A scores" as previously described (Birkedal *et al.*, 2014; Marchand *et al.*, 2016). **(D)** shows a scatterplot of methylation scores for known 18S and 28S rRNA methylation sites, with significant changes in methylation highlighted in red dots. **(E)** shows the change in methylation level for the known rRNA nucleotides.

Do hnRNPF/H share any of the *in vivo* functions we identified for tRF-Gly-GCC? To determine potential roles for hnRNPF/H in histone gene regulation and MERVL repression, we carried out RNA-Seq in ES cells subject to double knockdown of both hnRNPF/H – we note that hnRNPF and H can compensate for one another; for example, both hnRNPH RNA and protein are upregulated in hnRNPF knockdown cells (not shown). Double knockdown of hnRNPF/H resulted in dramatic alterations in expression of several hundred genes (**Figure 3.9A**), including significant downregulation of developmentally-relevant genes (*Sfrp4*, *Otx2*, *Dact2*, *Spry1/4*, *Gbx2*, *Notum*, *Notch4*, *Sall1, Fgf15*, *Tdgf1*, *Inhbb*, *Ltbp3*, *Fgf4*, *Pou4f2*, *Prdm14*, *Lefty2*, *Bmp4*, etc.). The misregulation of this wide array of differentiation-associated genes presumably results from the recently discovered role for hnRNPF/H in alternative splicing of Tcf3, a key transcription factor involved in stem cell maintenance and differentiation (Yamazaki *et al.*, 2018).

Consistent with the possibility that tRF-GG functions through hnRNPF/H binding, we find that hnRNPF/H also affected the same groups of genes regulated by tRF-GG, with hnRNPF/H knockdown resulting in downregulation of histone genes and a dramatic derepression of the MERVL program (**Figures 3.9A-C**). Using a MERVL reporter ES cell line (Macfarlan *et al.*, 2012a), we find that hnRNPF/H KD led to a ~30-fold increase in Tomato-positive cells (**Figure 3.9D, Figure 3.10A**), confirming the derepression of the MERVL program observed in the RNA-Seq dataset. Importantly, transfection of synthetic tRF-GG had no effect on MERVL repression in the absence of hnRNPF/H (**Figure 3.9D**), consistent with the hypothesis that tRF-GG acts by binding these proteins. Next, given the general role for tRF-GG in supporting the output of multiple Cajal body products, we examined Cajal body morphology in hnRNPF/H KD cells, using the well-known Cajal body marker coilin (Gall, 2000). In contrast to control ES cells, which exhibit one or two bright Cajal bodies per nucleus, we find that knockdown of hnRNPF/H leads to more diffuse coilin staining (**Figures 3.9E-F, Figure 3.10A**). Moreover, DAPI staining was clearly distinctive in hnRNPF/H KD cells, with the typical discrete chromocenters being replaced by more diffuse “lumpy” DAPI staining (**Figure 3.9E**), potentially secondary to the dramatic changes in histone expression in these cells. Taken together, our data reveal a novel role for hnRNPF/H in supporting normal Cajal body function in mouse embryonic stem cells, with downstream consequences for histone expression and chromatin-mediated repression of MERVL elements.

Discussion

Here, we investigated the mechanism by which tRF-Gly-GCC represses MERVL-associated gene expression in ES cells and preimplantation embryos. Surprisingly, we find that tRF-GG indirectly represses MERVL-driven transcription as a downstream consequence of its effects on noncoding RNA biogenesis. tRF-GG regulates a cascade of events through its role in supporting U7 snRNA levels, which in turn enhances expression of canonical histone genes, ultimately resulting in increased heterochromatin-dependent repression of LTR-associated genes (**Figure 3.11**).

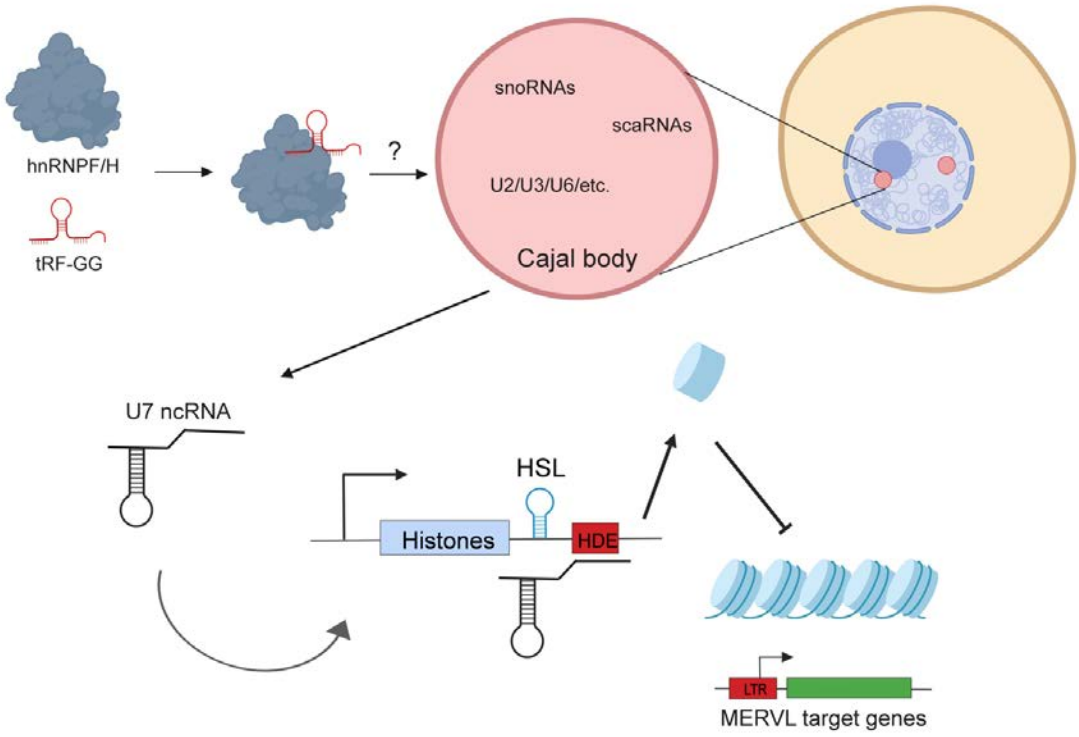


Figure 3.11. Schematic of proposed mechanism for tRF-GG function. Together, our data support a model in which 5' tRF-Gly-GCC binds to Hnrnpf/h and serves to support production of a variety of noncoding RNAs in Cajal bodies. Central to the current study is control of U7 ncRNA production, which plays a central role in processing of the histone 3' UTR via base-pairing to the histone downstream element (HDE). Altered expression of histones then leads to

downstream effects on the expression of MERVL-associated genes in murine embryonic stem cells and preimplantation embryos.

Our data demonstrate the effects of tRF-GG on MERVL target genes are mediated by altered histone gene expression. We document effects of tRF inhibition and tRF “overexpression” on histone expression by q-RT-PCR, by deep sequencing, and by quantitative Western blot, we confirm the expected downstream effects on chromatin compaction by ATAC-Seq, and finally we show that tRF regulation of histone genes can be recapitulated using two distinct histone 3’ UTR reporters. Moreover, the effects of tRF-GG manipulation – both inhibition and overexpression – on histone regulation and on MERVL target repression can be suppressed by appropriate manipulation of the U7 noncoding RNA. Our data argue that, rather than being directly targeted by tRF-GG via sequence homology between tRNAs and the PBS of LTR elements, MERVL instead represents a highly sensitive reporter of global chromatin status in mouse embryonic stem cells. Further supporting this idea is the fact that histone downregulation, rather than ERV derepression, is a conserved consequence of tRF-GG inhibition in both mouse and human ES cells.

tRF-Gly-GCC supports Cajal body output

Upstream of the histone 3’ UTR, we document a surprising and conserved role for tRF-GG as a positive regulator of noncoding RNA production, with effects of tRF manipulation on the histones resulting from altered levels of the U7 noncoding RNA. Indeed, perhaps the most unanticipated discovery described here is the finding that the 5’ tRF-Gly-GCC supports production of noncoding

RNAs that are normally synthesized or processed in the Cajal body. This is demonstrated using both gain and loss of function approaches in both mouse and human ES cells. Given the wide range of functions ascribed to Cajal body ncRNAs, our data also predict that tRF-GG may affect additional downstream pathways beyond the MERVL program previously described. Confirming this prediction, we find subtle effects of tRF-GG inhibition on rRNA 2'-O-methylation (**Figure 3.10 D-E**), which is mediated by snoRNAs, as well as a small global increase (~5%) in intron retention in tRF-inhibited cells, consistent with the observed decrease in spliceosomal RNAs. Thus, fine-tuning of Cajal body output by tRF-GG affects cellular functions ranging from splicing to translation to global chromatin packaging.

hnRNPF/H are novel regulators of the 2C-like state

What is the proximate mechanism by which tRF-GG supports production of Cajal body RNAs? We identify a strong candidate for a relevant effector protein, finding that the closely related hnRNPF and H proteins bind directly to tRF-GG in extracts and *in vitro*. Moreover, functional studies reveal that hnRNPF/H and tRF-GG exhibit heavily overlapping regulatory roles in vivo, identifying a novel and surprising role for hnRNPF/H in control of the MERVL program in mES cells. Indeed, hnRNPF/H represent the strongest repressors of the 2C-like state yet observed, as the ~30-fold derepression of the 2C-like state in response to hnRNPF/H KD is comparable to, and in fact more dramatic than, that previously observed following Chaf-1 or Ubc9 knockdown (Ishiuchi *et al.*, 2015a; Cossec *et al.*, 2018).

hnRNPF/H are RNA-binding proteins with relatively well-characterized roles in mRNA splicing (Wang, Dimova and Cambi, 2007; Wang *et al.*, 2012; Yamazaki *et al.*, 2018), and our RNA-Seq analysis of hnRNPF/H KD cells reveal hundreds of altered splicing events (see **Figure 3.10B-C** for examples), consistent with previous studies (Yamazaki *et al.*, 2018). How do these changes in RNA splicing – or, alternatively, some unrelated activity of hnRNPF/H – ultimately drive repression of the MERVL program? Given the dramatic changes in Cajal body morphology documented in hnRNPF/H KD cells (**Figure 3.9D-E**), we favor the hypothesis that one or more hnRNPF/H-regulated transcripts plays a key role in Cajal body function, with altered U7 RNA production altering histone production and thereby driving changes in the highly heterochromatin-sensitive MERVL program. A number of specific target(s) of hnRNPF/H could be responsible for supporting normal Cajal body biogenesis, as a large number of chromatin and RNA-binding proteins (such as Hnrnpa2b1) with potential roles in Cajal bodies exhibit altered splicing patterns in hnRNPF/H-depleted cells. Given that tRF-GG only affects a small subset of hnRNPF/H functions – tRF-GG inhibition has no effect on *Tcf3* splicing (**Figure 3.10B**), for example – we speculate that the hnRNPF/H targets responsible for Cajal body biogenesis are the most sensitive to subtle changes in hnRNPF/H function. It will be interesting in future studies to determine the mechanism by which tRF-GG enhances the function of hnRNPF/H – whether tRF-GG stabilizes hnRNPF/H for example, or whether hnRNPF/H functions in complex with tRF-GG at a subset of targets.

Implications for tRF-Gly-GCC function in the early embryo

We finally turn to the question of the physiological contexts in which tRF-Gly-GCC is likely to play an important role in cellular function. In typical somatic tissues, tRNAs are cleaved in response to a variety of stress conditions; for example tRNA cleavage occurs in response to arsenite treatment of neurons to produce tRNA fragments which help to direct global translational downregulation (Ivanov, Mohamed M. Emara, *et al.*, 2011). Curiously, beyond the case of stress-dependent tRNA fragment production, it is increasingly clear that tRNA cleavage occurs commonly in the germline of multiple organisms even under apparently stress-free growth conditions (Couvillion, Sachidanandam and Collins, 2010; Peng *et al.*, 2012a), and tRF-GG is one of the most abundant small RNAs present in mammalian sperm and delivered to the zygote upon fertilization. Although we previously showed that manipulating tRF-Gly-GCC levels in the zygote altered expression of ~50 MERVL-associated transcripts, these results were based on single-embryo mRNA-Seq which only reports on transcripts bearing a polyA tail. However, the wide variety of noncoding RNAs affected by tRF-Gly-GCC are predicted to affect an extensive list of additional cellular processes from ribosome biogenesis (snoRNAs) to splicing (scaRNAs and U RNAs) to global chromatin compaction (U7). How modulation of these various functions by tRF-GG delivery to the zygote affects later processes during preimplantation development, and whether any of these alterations have lasting consequences for offspring phenotypes, will be of great interest.

Materials and Methods

Mouse ES cell culture and transfections

All murine ES cell lines were grown in DMEM (Gibco) containing 10% fetal bovine serum and leukemia inhibitory factor (Serum+LIF culture conditions) and all transfections were carried out using OptiMEM and Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions at splitting, unless otherwise specified. Inhibition of tRF-Gly-GCC function in mESCs was performed as described in Sharma et al, 2016. Controls included Lipofectamine 2000 only (Mock), anti-GFP esiRNA transfections and/or a scrambled anti-tRF-GG LNA oligo. Transfection of various 3'tRFs was performed as for the 5'tRF-GG. Concentrations tested were 5 ng and 100 ng.

U7 rescue experiments were performed by supplementing the tRF-Gly-GCC inhibition transfection reaction with 50 ng of in vitro synthesized U7 snRNA per 3X10^5 cells. Mouse U7 small nuclear RNA was generated through in vitro transcription from pGEM-Teasy-U7 plasmid (generous gift from Z.Dominski) using mMessage mMachine T7 kit (Ambion) following plasmid linearization with HindIII restriction enzyme. Human U7 was cloned from hESC cDNA into the pGEM-Teasy plasmid and used as in mouse experiments. Media was changed after 16 hrs and cells were allowed to grow for additional 32 hrs. Cells were processed for various experiments at the end of the 48-hour period. U7 knockdown was achieved by transfecting 10 ng of modified anti-sense oligonucleotides targeting U7 synthesized with phosphorothioate linkages.

Double hnRNPF/H knockdown was performed by transient transfection of 20 pmols of siRNAs against mouse hnRNPF and hnRNPH transcripts, respectively (sc-38273 and sc-35580, Santa Cruz Biotechnologies) in 12-well

format. 40 pmols of siRNA-A (sc-37007, Santa Cruz Biotechnology) was used as knockdown control. Efficiency of knockdown was validated by Western Blotting and RNA-sequencing. Cells were collected 48 hours after transfection for various downstream experiments.

Cell culture (H9)

Line H9 (WA09) human embryonic stem cells were cultured on Matrigel (Corning) in mTESR1 media (Stem Cell Technologies) in 5% CO2 at 37°C. Nucleofection of oligos was done using Human Stem Cell Nucleofector Kit 1 (Lonza Bioscience), according to manufacturer's instructions. 2 ng of LNA was nucleofected by 24-well plate. Nucleofection efficiency was checked using pEGFP plasmid control, with successful experiments having more than 80% GFP+ cells. Cells were harvested 12hours post-nucleofection for RNAseq and qRT-PCR.

RNA-seq

5 µg of total RNA was depleted of ribosomal RNA using Ribo-Zero rRNA Removal Kit (Human, Mouse, Rat, Illumina). Less total RNA was used as input from metabolic labeling experiments for both 4sU labeled and unlabeled fractions. Illumina deep sequencing compatible libraries were constructed from rRNA-depleted RNA using an optimized version of a protocol described by Heyer et al., adding a purification using the RNA Clean and Concentrator (Zymo Research) in between procedures. Ribosome profiling data was published previously (GSE74537), with libraries constructed using the same procedure.

Libraries were quantified, multiplexed and either single-end or paired-end sequenced on Illumina NextSeq 500 sequencer.

Ribometh-seq

Ribometh-seq was done essentially as previous described (Marchand *et al.*, 2016). Briefly, 500 ng of total RNA was fragmented with 100mM alkaline solution at 95 °C for 10 mins. Fragmented RNA was then 3’ dephosphorylated with T4 PNK for one hour, followed by 5’ phosphorylation using T4 PNK in the presence of ATP for one hour. RNA was concentrated using Zymo RNA Clean and Concentrator, and libraries were constructed from 100 ng of prepared RNAs using NEBNext Small RNA Library Prep Kit following manufacturer’s instructions. Ribometh-seq libraries were single-end sequenced on an Illumina NextSeq 500 sequencer.

Deep sequencing data analysis

RNA-seq libraries were demultiplexed using Novobarcode (v3.02.08). Single end libraries were trimmed of 3’ adapters using Fastx-toolkit (v0.0.14). Quantification was done using RSEM (v1.2.29) to RefSeq GTF annotation, mapped with Bowtie (v1.0.0) to mm10 using default parameters.

For splicing analysis, reads were mapped using STAR aligner (v2.5.3a) using the “two-pass” method. Briefly reads were mapped to mm10 genome using default settings, then reads were remapped to the collective set of junctions predicted from the first pass. Alternative splicing was analyzed using JUM (2.0.2,

(Wang and Rio, 2018)). Alternative splicing events with a p-value less than 0.05 were extracted for further analysis.

ATAC-seq libraries were mapped to mm10 using Bowtie2 (v2.3.2) with the following parameters: -D 15 -R 2 -N 1 -L 20 -i S,1,0.50 --maxins 2000 --no-discordant --no-mixed. Fragment lengths were separated using Python, and coverage of reads in various chromatin states was analyzed in R (v3.4.1) using data from ChromHMM (https://github.com/guifengwei/ChromHMM_mESC_mm10). All coverage data was normalized by global read depth prior to further analysis. Circos plots (v0.69-2) were generated from coverage data calculated by Bedtools (v2.25.0).

Ribometh-seq libraries were mapped using Bowtie2 and the 5' locations of the forward read was quantified using Bedtools.

qRT-PCR and semi-quantitative PCR

For mESCs, RNA was isolated using Trizol (Ambion) according to the standard protocol and the samples were treated with TurboDNase to eliminate genomic DNA contamination. Following TurboDNase treatment, RNA was purified using Zymo RNA clean and concentrator 5 kit, quantified on Nanodrop and RT reaction was performed using SuperScript IV RT kit (Invitrogen) according to the manufacturer's protocol. Obtained cDNA was diluted 2X for MERV-L target gene qPCR) and 10X for histone and β-actin amplification (based on the standard curves obtained for the primers used) and Kapa SYBR Fast Universal Mastermix was used in all q-RT-PCR reactions. The amplification conditions were: 95C for 3 minutes followed by 40 cycles of 95C for 5s and 60C

for 20 s (with a plate-reading step between each cycle) based on the previously reported conditions for histone qPCR. All the reactions were read on the BioRad CFX96 qPCR detection machine. The same protocol, with human-specific primers, was performed for H9 human cells.

For semiquantitative PCRs to confirm candidate alternative splicing events, primers were designed in flanking exons. The number of PCR cycles were empirically determined to avoid saturation of PCR amplicons**.**

Western Blotting

Mock or tRF-GG inhibited cells were grown as before and 48 hours after transfection, cells were trypsinized and counted. A defined number of cells (usually 50000) was spun from each group, washed in PBS, pelleted and lysed directly by boiling at 100 °C for 15 minutes in 2X Laemmli Sample buffer containing β-mercaptoethanol. 2X serial dilutions of the cell lysates were loaded onto a 15% SDS-polyacrylamide gel. Following protein resolution on the gel, proteins were transferred to a nitrocellulose membrane for 1 hour at 100 V through wet transfer. Membrane was blocked in 5% milk in TBSt for at least an hour prior to the addition of primary antibodies overnight at 4 °C. Primary antibodies used were anti-histone H3 (Millipore, #07-690), anti-histone H4 (Millipore, #05-858), anti β-actin (abcam ab8224) and anti-GAPDH (abcam ab9485). Membranes were then washed 3X in TBSt for 15 minutes at room temperature. Membrane was incubated with secondary antibody in 5% milk for 1 hour at room temperature. Secondary antibodies used were anti-mouse IgG, HRP-linked (Cell Signaling, #7076S) and anti-rabbit IgG, HRP-linked (Cell

Signaling 7074S). Following 3 additional TBSt washes, the substrate for HRP was added (Amersham ECL Western Blotting Analysis System, GE) and the blots were exposed on the Amersham Imager 600 machine.

Northern Blotting

DNA probes for snoRNAs were ordered from IDT and 5' end-labeled with [γ-^{32}P] using PNK (NEB). U7 was present in very low abundance and had to be probed using a full-length [γ-^{32}P]C labeled RNA probe synthesized from a pGEM plasmid containing a cloned mouse U7. 2ug of total RNA was separated on a 6% urea PAGE gel. RNA was transferred onto a Hybond-N+ membrane (GE Healthcare) by semi-dry transfer method (BioRad) in SSC buffer. RNA was then crosslinked to the membrane at 254nm. Probes were then hybridized to RNA overnight with agitation at 68C. Membranes were then washed, and exposed to BioMax film (Kodak) for up to three days. Data was quantified using ImageJ.

Histone 3' UTR luciferase reporter assay.

Histone H3b (*Hist2h3b*) and histone H4j (*Hist1h4j*) 3'UTR sequences (~300 nucleotides) were cloned into the PsiCheck 2 vector downstream of the Renilla luciferase coding sequence, between XhoI and NotI restriction sites. As PsiCheck2 vector does not encode for a eukaryotic selectable marker, cells were co-transfected with PsiCheck2-empty or PsiCheck2-Histone3'UTR together with a carrier plasmid pCDNA3.1+ Hygro and stable cell lines were generated as described in (Connelly, Thomas and Deiters, 2012). Following 7 days of Hygromycin selection, individual clones were picked, expanded and tested for luciferase expression. Based on their expression level, one clone from all cell

lines was selected for subsequent experiments. Reporter ES cell lines were transfected as previously. 48 hours after transfection, cells were washed 2X with PBS and cell lysate for the luminescence reading was prepared as directed by the Dual-Luciferase Assay System (Promega). Firefly and Renilla luminescence was measured using GLOMAX 96 Microplate Luminometer, and Renilla luminescence was normalized to the internal control of Firefly luminescence.

Cell cycle analysis

E14 mESCs were transfected with antisense 5'tRF-Gly-GCC LNA-containing oligo or mock transfected as described above. Media was changed 16 hours after transfection. After 24 hours, cells were synchronized in G1/S-phase of the cell cycle by single thymidine block (5 mM thymidine in culture media) for 16 hours. Following treatment, cells were washed 2X with PBS, and fresh culture media without thymidine was added and were allowed to progress through the cell cycle. Cells were collected by trypsinization and 2X PBS washes at time zero and every 2 hours for an 8 hour period after removal of the thymidine block. Cold 70% Ethanol was added to cells dropwise with light vortexing and cells were fixed for 30 minutes at 4 °C. Following fixation, cells were pelleted, washed 2X with PBS and treated with RNase A (final concentration 0.2 mg/ml). Cells were stained with Propidium Iodide solution (final concentration 10 mg/ml) and DNA content was analyzed using FACScan Flow Cytometer (Becton Dickinson).

Oligonucleotide pulldowns and mass spectrometry.

Mouse ES cells were washed 2X with PBS and collected by trypsinization. Around 20 million cells were lysed in NP-40 buffer (50 mM Tris pH 7.5, 0.1% NP-

40, 150 mM NaCl, 1 mM CaCl2, 1 mM DTT, 40 U SuperaseIn, 1X Protease inhibitor cocktail) on ice for 15 minutes. Lysates were cleared by centrifugation for 15 minutes at 4°C and maximum speed and the supernatant was transferred to a new tube. The lysates were then divided in 2 and incubated with 500 pmols of tRF-Gly-GCC-biotin or tRF-Lys-CTT-biotin for 1 hour at room temperature with end-over-end rotation. Following the incubation, 100 µl of C1 Dynabeads were added to the reactions and incubated with end-over-end rotation for an additional hour. Captured RNPs were then washed three times each with low salt buffer (30 mM Tris-HCl pH7.5, 120 mM KCl, 3.5 mM MgCl2, 0.5 mM DTT), medium salt buffer (30 mM Tris-HCl pH7.5, 300 mM KCl, 3.5 mM MgCl2, 0.5 mM DTT) and high salt buffer (30 mM Tris-HCl pH 7.5, 0.5 M KCl, 3.5 mM MgCl2, 0.5 mM DTT) for 5 minutes at room temperature. Following the last wash, elution was performed by disrupting the biotin-streptavidin bond using 95% formamide and 10 mM EDTA at 65°C for 5 minutes. Eluate was boiled in 2X Laemmli sample buffer and loaded into 4-20% gradient polyacrylamide gel. Gels were stained by Coomassie Brilliant Blue solution for 1 hour at room temperature, then destained overnight in the destaining solution (40% MeOH, 10% acetic acid). Bands were cut from the gel and submitted for Mass Spectrometry at the UMass Medical School Mass Spectrometry facility.

Streptavidin RNA pulldown assay

For each assay, 2.5 uM of biotin labelled RNA was incubated with streptavidin beads (Invitrogen) according to the manufacturer's instructions. Beads were then incubated for 2 h with cellular lysate in binding buffer (0.01

mg/mL tRNA, 0.01% NP40, 0.1 mg/mL BSA, 50 mM Tris-Cl (pH 8.0), 100 mM NaCl, and 1/20 SuperaseIn (Ambion). After 2 hours of rotation at room temperature, the beads were washed with 200 µl of wash buffer (100 mM NaCl, 50 mM Tris-Cl (pH 8.0), 0.01% NP40, and 0.01 mg/mL tRNA) for 4 times. Proteins were eluted from beads with sample buffer for 5 min at 95 °C and equal amounts are run on an SDS polyacrylamide gel and analyzed by Western analysis.

hnRNPH1 purification

The sequence encoding amino acids 1–449 of mouse hnRNPH1 was cloned into pMal-ac (New England Biolabs) downstream an N-terminal maltose-binding protein (MBP) tag and the cloned construct was transformed into BL21(DE3) cells. The cells were induced with 1 mM isopropyl 1-thio-β-D-galactopyranoside for 3 h, at 37 °C. to express the protein with an N-terminal MBP tag. The cells were lysed in 200 mM NaCl, 50 mM Tris, pH 8.8, 2 mM DTT, and EDTA-free protease inhibitor tablet. Amylose (New England Biolabs) affinity column was used for the first step of purification of hnRNPH1. Protein fractions were eluted in lysis buffer supplemented with 10 mM maltose. Fractions containing the protein were pooled and dialyzed into an S-column buffer (20 mM NaCl, 50 mM MOPS pH 6.0, 2 mM DTT). Purification was followed by HiTrap S at 4 °C. Elution of the protein fractions was achieved by a salt gradient ranging from a low salt buffer (20 mM NaCl, 50 mM Tris MOPS pH 6.0, 2 mM DTT) to a high salt buffer (1 M NaCl, 50 mM MOPS pH 6.0, 2 mM DTT). Pure fractions were dialyzed in a Q-column buffer (20 mM NaCl, 50 mM Tris pH 8.8, 2 mM

DTT). Final purification was done using a HiTrap Q ion exchange column at 4 °C. Protein fractions were eluted by a salt gradient ranging from a low salt buffer (20 mM NaCl, 50 mM Tris, pH 8.8, 2 mM DTT) to a high salt buffer (1 M NaCl, 50 mM Tris, pH 8.8, 2 mM DTT). Pure fractions were determined by Coomassie-stained SDS-PAGE, and purified hnRNPH1 was dialyzed into storage buffer (25 mM Tris, pH 8.0, 25 mM NaCl, 2 mM DTT) and stored at 4 °C. Pure fractions were concentrated using an Amicon spin concentrator.

Preparation of Fluorescently Labeled RNA

RNA oligonucleotides were 3′-end labeled with fluorescein 5-thiosemicarbazide as previously described (Pagano, Clingman and Ryder, 2011). Briefly, RNA is first oxidized with sodium periodate and then reacted it with fluorescein 5-thiosemicarbazide to form a covalent bond. Labeled RNA is then purified over a Sephadex G25 column.

Electrophoretic Mobility Shift Assay

Electrophoretic mobility shift experiments and data analysis were carried out as previously described with a few modifications (Pagano, Clingman and Ryder, 2011). Briefly, 3 nM of labeled RNA was incubated with a gradient of hnRNPH1 concentration in equilibration buffer (0.01% Igepal, 0.01 mg/ml tRNA, 50 mM Tris, pH 8.0, 100 mM NaCl, 2 mM DTT) for 3 h. After equilibration, polarization readings were taken in a Victor plate reader. The samples were then mixed with bromocrescol green loading dye and loaded on a 5% native, slab polyacrylamide gel in 1× TBE buffer. The gels were run in 1× TBE buffer for 120 min at 120 volts and at 4 °C and then scanned using a fluor imager (Fujifilm FLA-

5000) with a blue laser at 473 nm. The fraction of bound protein against the protein was fit to the Hill equation using Igor Pro software.

Immunofluorescence

E14 mouse ES cells were grown and transfected as described above, and plated onto gelatinized coverslips. For immunofluorescence, cells were washed 2X in PBS and fixed in 4% paraformaldehyde for 20 minutes at room temperature with mild agitation. Fixed cells were permeabilized by 0.5% Triton-X solution for 20 minutes at room temperature followed by 3 washes in PBS containing 0.05% Tween 20 (PBS-Tween). Blocking was performed using 5% milk in TBSt for 1 hour at 37 °C. Cells were incubated in primary antibody in 3% BSA for 1 hour at 37 °C, followed by 3 washed with PBS-Tween. Primary antibodies used were anti-coilin (Abcam, ab210785, 1:50 dilution) and anti-hnRNPF/H (Santa Cruz Biotechnology, sc-32310, 1:500 dilution). Following the washes, cells were incubated in secondary antibody conjugated with fluorophores for 45 minutes to 1 hour at room temperature in the dark. Secondary antibodies used were Alexa Fluor 488 goat anti-mouse (1:1000 dilution), Alexa Flour 488 goat anti-rabbit (1:500 dilution). Following 3 additional washes with PBS-Tween, cells were mounted in Vectashield mounting medium containing DAPI for DNA visualization. Microscopy was performed on AxioObserver.Z1/7 microscope using 63X/1.4 NA oil objective. Images were analyzed using ImageJ software.

ACKNOWLEDGEMENTS

We thank Zbig Dominski for the U7 expression vector. We thank Sean Ryder and members of the Rando lab for critical reading of the manuscript and

insightful discussions. This work was supported by NIH grant R01HD080224. A.B. was supported by a fellowship from the Human Frontier Science Programme (LT000857/2015-L).

Discussion

The mechanistic basis of paternal epigenetic inheritance, from the carrier of epigenetic information, to the signaling pathway that enables epigenetic modification to the germline, to how epigenetic signaling influences offspring development, has remained elusive (Heard and RA Martienssen, 2014; Rando and Simmons, 2015). Here, I have demonstrated the mechanism by which a tRNA fragment (tRF-GG), modified by paternal diet, may be delivered to sperm via epididymosomes, and ultimately transferred to an oocyte to influence embryonic gene expression, in particular to help repress MERV-L associated transcripts (Chapter II). Mechanistically, I show that the tRF-GG supports the production of numerous small nuclear RNAs associated with the Cajal bodies, particularly the levels of U7 snRNA, to ensure an adequate supply of histone proteins. This in turn safeguards heterochromatin-mediated transcriptional repression of MERV-L elements (Chapter III). Our work illustrates a potential mechanism by which paternal diet may influence offspring development, through changes in the levels of a novel small RNA, which acts as to trigger a cascade of events involving small-nuclear RNAs, histone mRNAs, chromatin compaction, and ultimately in the regulation of ERV expression in the early embryo.

Epididymosome epidemiological epigenomics

First, my work suggests that epididymosomes are capable of shaping the sperm epigenome. Taqman assay of tRFs supported previous deep sequencing data showing that epididymosomal small-RNAs are surprisingly concordant with

sperm small-RNAs, and show a gain of tRFs after sperm exit the testis, and further as sperm transit from caput to cauda (Sharma *et al.*, 2016). Recently, this work has been extended using tissue-specific metabolic labeling of RNA *in vivo*, definitively showing that cauda sperm carry small RNAs originating from the caput epididymis (Sharma *et al.*, 2018). Taken together, these data demonstrate that soma-germline RNA transfer occurs in male mammals, most likely via vesicular trafficking from the epididymis to maturing sperm.

Future work needs to focus on how environmental conditions might influence the sperm epigenome. Understanding the molecular basis by which small-RNAs are sorted and packaged into epididymosomes will be crucial in uncovering the mechanism of communication between somatic cells of epididymis and the maturing gametes, and shed light on how environmental information is signaled to sperm. First, what is the endonuclease responsible for cleavage of tRNAs in the epididymis, or does further processing of tRNAs occur in sperm? Whole animal loss of function experiments *in vivo* for the highly expressed endonucleases, such as RNases 4, and 9-13 cluster, would be crucial to revealing the biogenesis pathway of tRFs in the epididymis. Further dissection using various tissue-specific Cre-drivers would then enable definitive identification of the specific tissue (epididymis or sperm) responsible for generating sperm-borne tRFs. Another important avenue of discovery will be the identification of the cell type(s) responsible for epididymosome biogenesis – using single-cell RNA-seq, and subsequently cell-type specific epitope tagging and FACS sorting to enable more detailed cellular capture follow-up studies.

In terms of further understanding the functions of other small-RNAs carried by sperm, a particularly interesting avenue is the modification of levels of miRNAs in sperm. Specific clusters of miRNAs appear to be lost as sperm exit the testis, and then gained as sperm travels through the epididymis, which are essential for embryonic development (Conine *et al.*, 2018; Sharma *et al.*, 2018). What is the mechanistic basis of this loss then gain? One possibility is *in situ* cleavage, as sperm are loaded with Dicer (Yuan *et al.*, 2016), certain miRNAs could be continuously processed as the sperm matures. This possibility, while parsimonious, cannot explain how certain miRNAs are lost, then gained. An alternative possibility is that specific miRNAs are somehow gained via epididymosome fusion. Indeed, specific miRNAs are enriched in cauda epididymosomes, can be delivered to caput sperm, and can rescue gene expression defects associated with developmental deficiencies in embryos fertilized by caput sperm (Sharma *et al.*, 2016, 2018; Conine *et al.*, 2018). Loss of function studies *in vivo* of Dicer and DGCR8 with tissue specific Cre-drivers would help identify the biogenesis pathway for these miRNAs in the cauda epididymis. Furthermore, the determination of molecular targets of these miRNAs in the embryo is critical. This could be done using CLASH in mouse embryonic stem cells (Helwak *et al.*, 2013), and/or more low-input amenable technologies such as TRIBE fused to AGO2 in embryos (McMahon *et al.*, 2016), which together would further elucidate the underpinnings of sperm small-RNA regulation of embryonic gene expression.

In terms of other sperm-delivered small-RNAs, future work needs to tease

out the potential roles that other tRFs that are gained during caput-cauda transit may play in embryonic development, using embryonic stem cells as a springboard for functional discovery in embryos. For example, 5’ tiRNAs derived from Val-CAC are gained more than 10 fold from caput to cauda in mature sperm. In addition, since sperm contributes such a small-fraction of small-RNAs to the oocyte, how these small-RNAs play such important roles in embryonic development needs to be further analyzed. Perhaps sperm-borne small-RNAs carry specific RNA modifications or come preloaded with effector proteins. Given technological improvements in modification-based immunoprecipitation of RNAs and ultra-low input mass-spectrometry quantifications of proteins and RNA modifications, the testing of modification-dependent effects of tRFs is possible. Finally, within the sperm small-RNA pool, the prevalence, composition, and potential functions of tRF-3’s remain unexplored. Our efforts to clone tRF-3’s have mixed results so far (Appendix I) (Sharma *et al.*, 2018), although we have been able to detect tRF-3’s derived from Val-CAC and Gly-GCC using northern blotting. These tRFs therefore represent a significant blind spot in our understanding of sperm-borne small-RNA function and tRF function in mammalian cells in general. It is also of critical importance as work from others has shown that they might play a central role in silencing of young ERVs in mammalian cells (Andrea J. Schorn *et al.*, 2017), and unveiling their roles in early embryos would be of utmost interest.

One challenge to understanding the general applicability of various tRFs is the inconsistencies of groups in identification and naming of sometimes the same

tRNA fragments. Going forward, studies of tRNA fragments need to become more focused and unified in terms of use of terminology and technology. The diverse use of different terms for the same tRNA fragments also precludes an easy synthesis of the field of research in general, although some efforts have been made in this direction, with mixed reception (Kumar, Kuscu and Dutta, 2016).

tRF-GG regulates a subset of the MERVL totipotency program

Our work shows that one particular tRF, derived from Gly-GCC (tRF-GG), functions to repress the MERVL transcriptional program in early preimplantation embryos. This novel tRF function has the potential to influence the developmental rate of the embryo, as low protein sperm derived embryos (which contain higher levels of tRF-GG, and lower levels of MERVL activation at the 2/4 cell stage) experience a delay to blastocyst formation. This result was surprising given that MERVL is associated with totipotency, and we expected perhaps a differential distribution of the number of trophectoderm vs inner cell mass cells in the blastocyst, possibly leading to differences in placentation. Since placentation is a key determinant of offspring metabolic health (Rando and Simmons, 2015; Woods, Perez-Garcia and Hemberger, 2018), this hypothesis would directly link regulation of MERVL expression via tRF-GG and the metabolic status of offspring.

Nevertheless, available data supports our findings in terms of developmental phenotype resulting from MERVL misregulation in the early

embryo. Deletion of DUX, the transcription factor responsible for driving the MERVL transcriptional program, and knockdown of MERVL itself, lead to disruption of the pace of cell divisions and early embryonic developmental progression (Huang *et al.*, 2017; Iaco *et al.*, 2019). Furthermore, loss of Dppa2/4, potential inducers of DUX expression, leads to partial early embryonic lethality (Madan *et al.*, 2009). In addition, loss of maternal G9a, which leads to derepression of MERVL and other ERVs, also leads to a developmental delay phenotype (Yeung *et al.*, 2019). Taken together, the regulated expression of MERVL at the 2/4 cell stage appears to synchronize the developmental rate of the pre-implantation embryo, which sets it up for proper post-implantation development. Indeed, delay in blastocyst formation in humans is associated with lower implantation rates and poor pregnancy outcomes (Levens *et al.*, 2008; Irani *et al.*, 2018). Given that high-fat diet sperm-borne small-RNAs can recapitulate whole sperm induced metabolic consequences in the offspring (Chen, Menghong Yan, *et al.*, 2016; Zhang *et al.*, 2018), future work needs to investigate the molecular underpinnings of small-RNA induced changes in developmental rate, and directly test the hypothesis of whether changes in developmental rate resulting from paternal dietary perturbations can lead to changes in metabolic outcome of the offspring.

tRF-GG regulates histone mRNA levels via control of U7 snRNA

However MERVL misregulation could influence embryonic development, the latter part of my dissertation was focused on how tRF-GG regulates MERVL

expression (Chapter III). Surprisingly, our work suggests that tRF-GG represses the expression of small-nuclear RNAs, including the levels of snRNA U7. Therefore, influencing the levels of tRF-GG in sperm can potentially fine-tune small-nuclear levels in the embryo. Importantly, tRF-GG effects on histone mRNA levels, activity of a histone 3'UTR reporter, and expression of MERV-L associated transcripts can all be suppressed by appropriate manipulation of U7 RNA levels. Of note, the observed partial rescue of U7 on tRF-GG effects on MERVL could result from two reasons that are not mutually exclusive: potential inefficiencies of U7 levels in all cells (which could result from transfection issues or U7snRP assembly), or that tRF-GG acts through other pathways to repress MERVL expression as well.

Proximal experiments involve understanding the impact of perturbing U7 levels in embryos, and in general, perturbations of small-nuclear RNA levels. Microinjections of anti-sense oligos targeting these small-nuclear RNAs, and subsequent single-RNA RNAseq would begin to elucidate the molecular functions of these small-nuclear RNAs in embryos. These experiments would be greatly aided however, first by a better understanding of the regulation of small-nuclear RNA dynamics and histone expression in mammalian embryos. Investigations in this line of thinking promise to be interesting, as histone dynamics in mammalian embryos are unique in many ways, and many facets of this intricate regulation are not understood, including when embryonic histone genes become expressed to replace the maternal deposited pool, and whether small-nuclear RNAs play any role in regulation of embryonic development. Work

on these topics have so far focused on *Drosophila* and Xenopus embryos (Nizami, Deryusheva and Gall, 2010; Marzluff and Koreski, 2017), but signs point to the possibility that the story in mammalian embryos will be unique: preliminary staining of the Cajal bodies and histone locus bodies using Coilin as a marker gene has revealed interesting dynamics in the early embryo (Ana Boskovic unpublished data). In addition to immunostaining and imaging of specific marker genes for HLB and Cajal bodies, which would be greatly aided by live imaging techniques becoming possible in mammalian embryos, understanding the unique characteristics of the formation and dynamics of these bodies, which have the appearance of phase-separating nuclear bodies such as nuclear speckles (Trinkle-Mulcahy and Sleeman, 2017), would also illuminate the biophysical properties of the early mammalian embryo nucleus. Furthermore, deep sequencing of small-nuclear RNA pools and histone genes would also provide helpful supportive data. This is currently limited by deep sequencing technology used for single-embryo RNA-seq, which uses poly-A bead enrichment of mRNAs. However, this hump will certainly be overcome, as novel deep sequencing approaches are developed to target these currently ignored RNA species. Already, improved technology for ultra-low input small-RNA sequencing has helped us gain a better understanding of small-RNA dynamics in the early mammalian embryo (Yang *et al.*, 2016).

hnRNPF/H bind tRF-GG and repress the 2C-like state

How does tRF-GG regulate small-nuclear RNA levels? Our current model

invokes the potential tRF-GG effector protein hnRNPF/H (Chapter III **Figure 3.11**). Our work demonstrates that hnRNPF and H bind directly to tRF-GG, and display a stark overlap of *in vivo* functions to tRF-GG: repression of MERVL target gene expression and the 2-cell like state in mESCs, regulation of histone mRNA levels, and stabilization of the Cajal body. The identification of the exact mechanism by which hnRNPF/H regulates Cajal body formation/function, and by extension, what role tRF-GG plays in this process is paramount. First, does hnRNPF/H regulate small-nuclear RNA levels and/or their modifications? Addressing these questions directly using deep-sequencing, northern blotting, and qRT-PCR in mESCs depleted of hnRNPF/H would be of primary importance. LC-MS/MS, primer-extension analysis and/or novel deep sequencing methods mapping 2'-O-methylation and pseudo-uridinylation would help probe quantitative changes in modification levels of various small-nuclear RNAs.

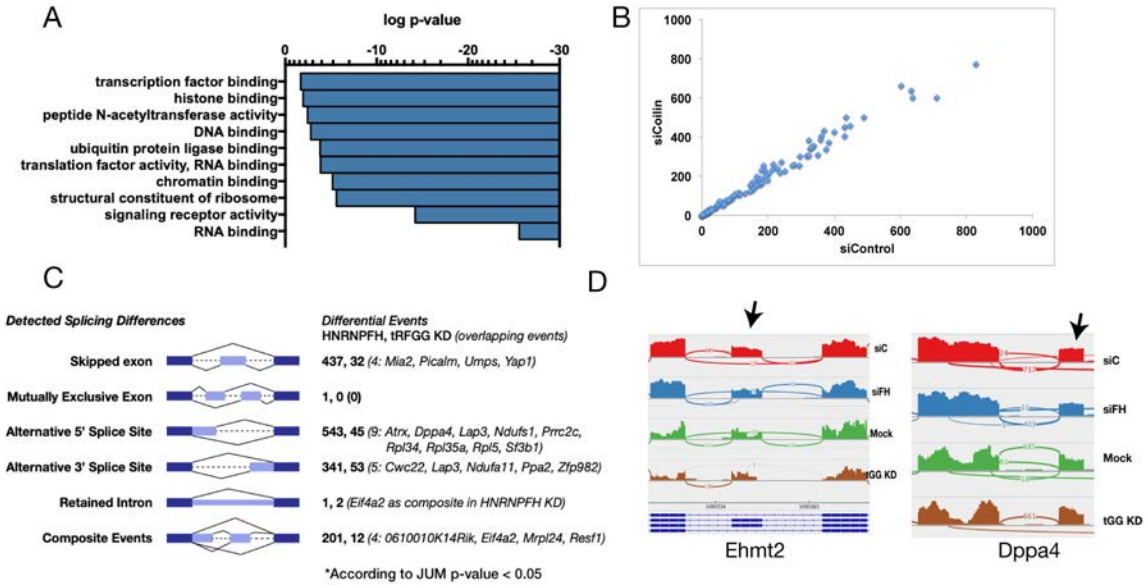


Figure 4.1. Molecular dissection of hnRNPF/H function. A) GO term enrichment for alternatively spliced targets of hnRNPF/H identified by JUM (Wang and Rio, 2018). **B)** Scatterplot of expression levels quantified by RNAseq

for all snoRNAs and scaRNAs in siCoilin vs siControl mESCs. **C)** Overlap of alternative spliced targets between hnRNPF/H and tRF-GG. A quantitative analysis of percent spliced index (PSI) values may provide a clearer picture of tRF-GG effect on hnRNPF/H splicing targets. **D)** Example browser tracks of splicing changes in two MERVL target genes in hnRNPF/H depleted mESCs (siFH). Note no change in tRF-GG inhibited cells.

Next, if hnRNPF/H regulates the levels of small-nuclear RNAs, how does it do so, and what role does tRF-GG play in this regulation? This could be possibly accomplished through its involvement in the stabilization of the Cajal body, and its functions within. As discussed in Chapter III, a number of targets of hnRNPF/H could be responsible for supporting normal Cajal body biogenesis, as a large number of chromatin and RNA-binding proteins (such as Hnrnpa2b1) with potential roles in Cajal bodies exhibit altered splicing patterns in hnRNPF/H-depleted cells (**Fig. 4.1A**). In an attempt to interrogate these targets, and to directly perturb Cajal bodies *in vivo,* we knocked down Coilin in mESCs and performed RNAseq. However, the dissociation of Cajal bodies by depleting Coilin in mESCs did not lead to changes in small-nuclear RNA levels (**Fig. 4.1B**), although depletion of Coilin in mouse and in *Drosophila* also do not show this phenotype nor changes in small-nuclear RNA modification levels, despite dissociation of Cajal bodies into residual bodies (Tucker *et al.*, 2001; Jády *et al.*, 2003; Liu *et al.*, 2009). Perhaps more precise experiments could be conducted if we have a better understanding of the direct mechanism of hnRNPF/H regulation of Cajal body function. Currently, our limited immunostaining experiments suggest that hnRNPF/H are depleted from Cajal bodies (Ana Boskovic personal communication.). Then, does hnRNPF/H impart its Cajal body function through its interaction with other RNAs (possibly lnc-RNAs or pre-mRNAs), which are

somehow involved in Cajal body formation/stabilization? This could also be the case if it becomes clear that hnRNPF/H does not regulate small-nuclear RNA levels or modifications. Since hnRNPF/H is known to regulate splicing of pre-mRNAs, CLIP-seq and alternative splicing analysis of cells depleted of hnRNPF/H and tRF-GG, and a combination of the two, would help illuminate the interplay of these two factors in the regulation of pre-mRNAs and lnc-RNAs, and the possible nuclear compartments tRF-GG may influence hnRNPF/H function. Preliminary analysis reveals some overlap between tRF-GG and hnRNPF/H regulated alternative splicing targets (**Fig. 4.1C**). To identify specific hnRNPF/H regulated targets, we are conducting CLIP-seq in tRF-GG depleted mESCs. Future analysis of overlapping targets of hnRNPF/H and tRF-GG might help reveal how tRF-GG may enhance hnRNPF/H function *in vivo*. Together, these ongoing experiments will reveal how tRF-GG may reinforce hnRNPF/H binding at its target sites.

What would be a possible mechanism by which tRF-GG reinforces hnRNPF/H binding to its pre-mRNA targets? *In vitro* studies using hnRNPF/H cognate binding oligos, purified hnRNPF/H and tRF-GG will help resolve the exact mechanism by which this occurs. In addition, as hnRNPF/H probably functions in a complex, some variation of protein-focused IP technique would help elucidate the proteins interacting with hnRNPF/H *in vivo*, including APEX or HRP proximity labeling followed by quantitative MS (Chen and Perrimon, 2017), which would help improve our understanding of the subcellular regions and proteins involved in hnRNPF/H function in conjunction with tRF-GG.

Finally, how does hnRNPF/H repress MERVL transcription and entry of mESCs into the 2-cell like state? There are two main possibilities: direct regulation of splicing and expression of a regulator of MERVL expression, or a more attractive model of some sort of RNA-mass action regulation of chromatin compaction. First, splicing analysis reveals that hnRNPF/H in fact regulates expression of several known MERVL regulators, such as *Ehmt2* and *Dppa4*, and most relevant to our model, histone mRNA levels (**Figure 4.1D, Figure 3.9B**). Can alternative isoforms of MERVL regulators play important roles in repression of MERVL expression? This would add an interesting wrinkle in the story of MERVL repression and the regulation of early embryonic gene expression in general. CRISPR/Cas9 deletion of hnRNPF/H regulated alternatively spliced exons, or perhaps using anti-sense oligonucleotides to block splice sites (Havens and Hastings, 2016) applied to MERVL regulators would help expand our understanding of the role that protein isoforms may play in early developmental programming. This possibility is not however mutually exclusive from the second of RNA-mass action, since it has been shown that LINE-1 elements regulate MERVL both via direct action through protein binding (Percharde *et al.*, 2018) and modulation of chromatin compaction in general (Jachowicz *et al.*, 2017). Perhaps proper binding of lnc-RNAs and/or pre-mRNAs by hnRNPs acts as some sort of chromatin conformation buffer. Systematic loss of function studies of hnRNPs in mESCs and hESCs, and in mouse embryos, coupled with RNAseq, ATAC-seq, and Hi-C/Micro-C approaches would help directly test this possibility.

Big picture questions

Whether tRF-GG ultimately functions in conjunction with or independently of hnRNPF/H to repress MERVL, the discovery that a novel small-RNA and a pre-mRNA binding splicing factor regulates MERV-L transcription reveals the intricate pathways involved in silencing of ERVs in mammalian cells. These pathways contribute to the already discovered transcriptional and post-transcriptional mechanisms that help suppress the expression of ERVs. However, none of the previously discovered factors function from the start of fertilization to fine-tune MERVL expression, which is the case for tRF-GG which appears to be delivered by sperm. How much of this sperm-borne pool influences the oocyte derived pool remains to be addressed, which will require careful spike-in mediated UMI-derived small-RNA libraries, and possibly SNPs or metabolic labeling to separate the maternal from the paternal pool definitively.

Regardless of the outcome of these particular experiments, the ability of sperm-borne RNAs to regulate embryonic gene expression presents a major new arc to the argument of genetic conflict hypothesis. At least three different species of RNAs, miRNAs (Rassoulzadegan *et al.*, 2006; Grandjean *et al.*, 2016; Conine *et al.*, 2018), tRFs (Chen, M Yan, *et al.*, 2016; Sharma *et al.*, 2016; Zhang *et al.*, 2018), and lnc-RNAs (Gapp *et al.*, 2018) have now be shown to influence offspring phenotype. These add to paternally imprinted alleles as an additional layer of influence which fathers contribute to the offspring, which were originally proposed to be in apparent conflict with maternal imprinting (Haig, 2004). A few aspects of the sperm-borne RNA hypothesis are worth considering long term.

First, since RNAs pools are capable of being modified, this confers some level of adaptability to the downstream-induced changes in gene expression. More detailed exploration of the breadth of sperm-borne RNAs capable of inducing gene expression changes in the embryo would be of extreme interest (discussed above). Second, what is the bandwidth of sperm? In other words, do all environmental changes lead to the alterations of the same small-RNAs, or do different stresses alter different RNAs, or possibly other epigenetic carriers of information? Currently, there is evidence that high-fat diet and low protein diet lead to changes in the same tRFs (Sharma *et al.*, 2016; Zhang *et al.*, 2018), although acute stress appears to modify long RNAs (Gapp *et al.*, 2018). A more comprehensive analysis of changes in sperm-borne RNAs in response to various paternal paradigms would help address this question. It is also important to consider the downstream effects of these changes in sperm-borne RNAs or sperm-borne chromatin for that matter. What is the scope of gene expression changes, and consequently, phenotypic changes associated with paternal stress paradigms? We revealed a conserved tRF-dependent modulation of ERV expression through small-nuclear RNA regulation in our low-protein paradigm. Do all dietary stresses lead to small-RNA changes that modulate ERV expression? Does this encompass stress exposures and toxin exposures? Detailed embryonic gene expression analysis of changes associated with all paternal paradigms explored so far would reveal the extent or possibility of a unifying theory of paternal epigenetic inheritance. At least for the final phenotype, it appears that most paternal interventions lead to some sort of metabolic

outcome (Rando, 2012; Rando and Simmons, 2015). The exploration of how early embryonic gene expression lead to metabolic consequences later in life will also be critical to our ultimate understanding of the mechanistic underpinnings of paternal epigenetic inheritance.

Finally, from an evolutionary perspective, are there certain conditions that are more amenable for the emergence of heritable forms of epigenetic variation, and how might this information become fixed genetically? This is perhaps an existential question for the field of transgenerational epigenetic inheritance. At least for our model, we have not extended studies into the F2 and 3 generations, although mammalian environmental epigenetic paradigms in general have not reported effects lasting more than 2 generations after exposure in male or female lineages. However, it is clear that epigenetic signatures are heritable indefinitely, as is the case of genomic imprinting and epialleles. A major intrigue in our study was the finding linking paternal diet and the expression of ERVs in the offspring, particularly a well-adapted ERV such as MERVL that has now assumed presumably important roles in embryonic development. More detailed studies in the potential developmental roles that ERVs such as MERVL play in mammalian embryos will help illuminate the possible adaptive avenues that could lead to triggering of lasting epigenetic memory. The apparent widespread nature of epialleles in plants (Heard and R Martienssen, 2014), and the rapid rate by which genomic imprinting is evolving in placental mammals (Spencer and Clark, 2014), suggest that robust opportunities in the evolutionary history of a species exist for adaptive epigenetic memory to have been, and are being evolved.

APPENDIX I

Diverse cleavage of tRNA fragments

Introduction

There are three well-understood classes of small non-coding regulatory RNAs (sRNAs) – microRNAs (miRNAs), Piwi-interacting RNAs (piRNAs), and small-interfering RNAs (siRNAs) (Ghildiyal and Zamore, 2009). These sRNAs associate with Argonaute proteins to regulate gene expression and development in Eukaryotes (Stefani and Slack, 2008). In recent decades however, deep sequencing of sRNAs, usually of RNAs less than 40nt, in lieu of miRNAs and piRNAs between 20-30nt in length, have revealed a diverse arsenal of small non-coding RNAs derived from tRNAs, snoRNAs, rRNAs, and mRNAs (Kawaji *et al.*, 2008; Kumar, Kuscu and Dutta, 2016).

Of these novel sRNAs, fragments of tRNAs, or tRFs, have gained particular attention due to their high copy number, evolutionary sequence conservation, and potential functional diversity (Keam and Hutvagner, 2015). Since tRNAs are post-transcriptionally modified in many ways, including splicing, cleavage of leader and trailer sequences, addition of non-templated ‘CCA’ sequence at the 3’ end, and chemical modifications (Torres, Batlle and Ribas de Pouplana, 2014), a kaleidoscope of tRNA fragments may therefore be generated from tRNAs at various stages of their life cycle (Fig. A1.1). Adding to this complexity various nucleases have been implicated in cleavage of tRNAs, including RNase III endonuclease Dicer (Babiarz *et al.*, 2008; Cole *et al.*, 2009), RNase Z (Haussecker *et al.*, 2010), and RNase A superfamily related

endonucleases such as Angiogenin (Yamasaki, Ivanov, G. F. Hu, *et al.*, 2009; Emara *et al.*, 2010; Ivanov, Mohamed M. Emara, *et al.*, 2011a; Li *et al.*, 2012), with various associated lengths and ends (discussed in Introduction). In addition, tRNA modifications can hinder or promote cleavage by various nucleases (Tuorto *et al.*, 2012; Guzzi *et al.*, 2018; Lyons, Fay and Ivanov, 2018), influence biological function (Zhang *et al.*, 2018), and impact whether the tRFs are able to be cloned, and therefore, assumed to be “present” in the sample by standard sRNA deep sequencing methods (Cozen *et al.*, 2015).

This study was aimed to better understand the last point – whether the method used to clone sRNAs is an important contributor to at least some of the diversity in the tRFs cloned by different groups. Here, there are two separate important factors to consider – whether tRNA modifications can impede cloning by generic methods, and whether certain tRFs are preferentially cloned due to differential biogenesis pathways. Since it has been shown that treatment of tRFs with dealkylating enzyme *E. coli* AlkB, which demethylates numerous commonly found tRNA methylation marks, can increase cloning efficiency of specific tRFs (Cozen *et al.*, 2015), I will focus on cloning of tRFs resulting from different biogenesis pathways.

Specifically, it has been known for some time that various species of tRFs result from cleavage of diverse endonucleases, and therefore have distinct biochemical differences at the 5’ and 3’ ends (Couvillion, Sachidanandam and Collins, 2010; Haussecker *et al.*, 2010). Cleavage by Dicer and RNase III family endonuclease, leaves a characteristic 3’ hydroxyl group on one 3’ end, and a 5’

monophosphate on the 5’ end of the other. The same ends result from exonuclease digestion. These ends allow efficient ligation by the T4 RNA ligase used in most sRNA cloning protocols. However, cleavage by RNase A superfamily endonucleases, such as Angiogenin (ANG), the nuclease implicated in cleavage of 5’ and 3’ tRNA-halves in certain mammalian cell lines (Yamasaki, Ivanov, G.-F. Hu, *et al.*, 2009; Ivanov, Mohamed M. Emara, *et al.*, 2011), forms a 3’-monophosphate via a 2’3’-cyclic phosphate intermediate on the 3’ end of the 5’ tRNA-halves (or 5tiRs), and a 5’ hydroxyl on the 5’ end of the 3’ tRNA-halves (or 3tiRs) (Kumar, Kuscu and Dutta, 2016). These ends do not allow for efficient ligation by traditional sRNA cloning methods, and attempts have been made to clone these tRFs or RNAs, with some success (German *et al.*, 2009; Schutz, Hesselberth and Fields, 2010; Schifano *et al.*, 2014; Peach, York and Hesselberth, 2015). For example, Peach *et al*. utilized *E. coli* RtcB RNA ligase to ligate cloning adapters to RNAs with 5’ hydroxyl termini, which allowed for specific cloning of 3tiRs, while Schutz *et al*. utilized the *A. thaliana* tRNA ligase to ligate adapters to RNAs with 2’3’-cyclic phosphates, which increased cloning efficiency for 5tiRs. However, these methods require expression and isolation of non-commercial enzymes or custom adapters, which has limited widespread utility.

Results and Discussion

We aimed to develop a sRNA cloning method that was adaptable to popular sRNA cloning methods, utilize enzymes that were commercially available, and could inform us of the diversity of sRNAs in sperm. T4

Polynucleotide Kinase (T4 PNK) is a widely used molecular enzyme that has both 5'-kinase (ATP-dependent) and 3' phosphatase activity (ATP-independent) (Richardson, 1965; Cameron and Uhlenbeck, 1977; Wang, Lima and Shuman, 2002). T4 PNK also has a modified version called 3'phosphatase minus (3'-minus) which does not have 3' phosphatase activity and exclusively phosphorylates 5' hydroxyl groups, which we also used to treat sRNAs prior to cloning with Truseq Small RNA Library Preparation kit. The Truseq Small RNA kit utilizes a standard sRNA cloning method by ligating two short RNA adapters to either end of the sRNA, with the 3' adapter being pre-adenylated to facilitate specific ligation (Viollet *et al.*, 2011), which is a method also used in most custom RNA ligation protocols and the NEBnext Small RNA Library Prep kit. The NEBnext kit is also a popular and cheaper alternative for sRNA cloning, but has given us different results for tRFs as compared to the Truseq kit, which in our hands gives more comparable results to a previously used custom protocol (data not shown). This discrepancy between kits could potentially result from the primer hybridization steps in the NEBnext protocol that is suggested to minimize adapter-dimers, or differences in reagents used during ligation steps (for example PEG decreases ligation bias). Another potential area of discrepancy could result from cloning from total RNAs or gel-purified sRNAs, as this has also led to changes in the propensity to clone different types of sRNAs in our hands.

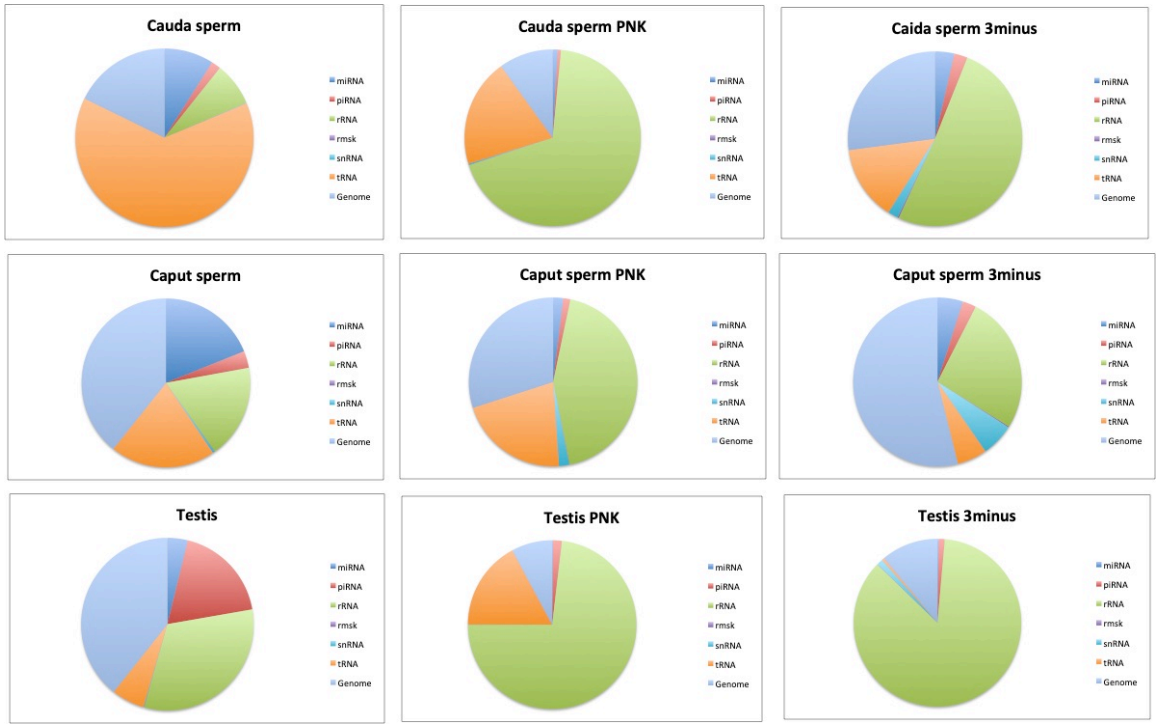
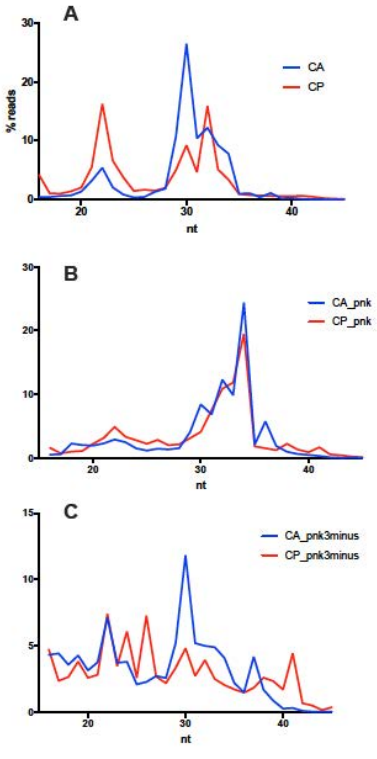


Fig. A1.1. Effect of T4 PNK (PNK) and T4 PNK 3'phosphatase minus (3minus) on cloning of sRNAs in cauda sperm, caput sperm, and testis. Reads were sequentially mapped as previously described (Sharma *et al.*, 2016).

The introduction of 30 mins of T4 PNK or 3'-minus treatment to cauda sperm sRNAs prior to cloning led to striking changes to the different types of sRNAs represented in the total library (Fig. A1.1). While the majority of cloned cauda sperm sRNAs were tRFs using the original Truseq protocol (63.56%), agreeing with previous results (Sharma *et al.*, 2016), PNK or 3'-minus treatment led to a large reduction in tRFs cloned (19.83% and 13.89%, respectively). Instead, tRFs were replaced in the libraries by fragments derived from rRNA for all samples (Fig. A1.1). It is important to note that caput sperm and testis loss of tRFs were much more modest than that of cauda sperm (~2 fold vs 8 fold, respectively). Tellingly, PNK treatment of caput sperm sRNAs did not lead to an increase in cloned tRFs, while 3'-minus actually led to a reduction from 20.13%

to 5.72% (Fig. A1.1). Together, these data suggest that the majority of tRFs gained during epididymal transit are either Dicer generated, or further exonuclease digested products of RNase A family endonuclease cleaved tRNAs. Instead, the majority of sRNAs gained as sperm mature are rRNA fragments, which are actually similar in size to tRFs (~30-35nt, Fig. A1.2) (Fowler *et al.*, 2018). The major caveat in this conclusion is the assumption that cloning of different types of sRNAs are equally efficient, which is not true if certain species of tRFs contain hard-stop modifications (Cozen *et al.*, 2015). Support for either possibility will come from northern blotting of testicular sperm (instead of whole testis, which could explain rRNA contamination in these libraries), caput and cauda sperm, for rRNA and tRNA fragments. When considering probes for northern blot analysis, one should ponder the rRNA sequence to probe in northerns, as different treatments reveal rRNA fragments derived from different parts of rRNAs (Fig. A1.3).

Fig. A1.2. Dynamics of rRNA cleavage during sperm maturation. The majority of sRNAs cloned after PNK treatment are 30-35nt rRNA fragments, suggesting they are cleaved by RNase A superfamily proteins into this length.



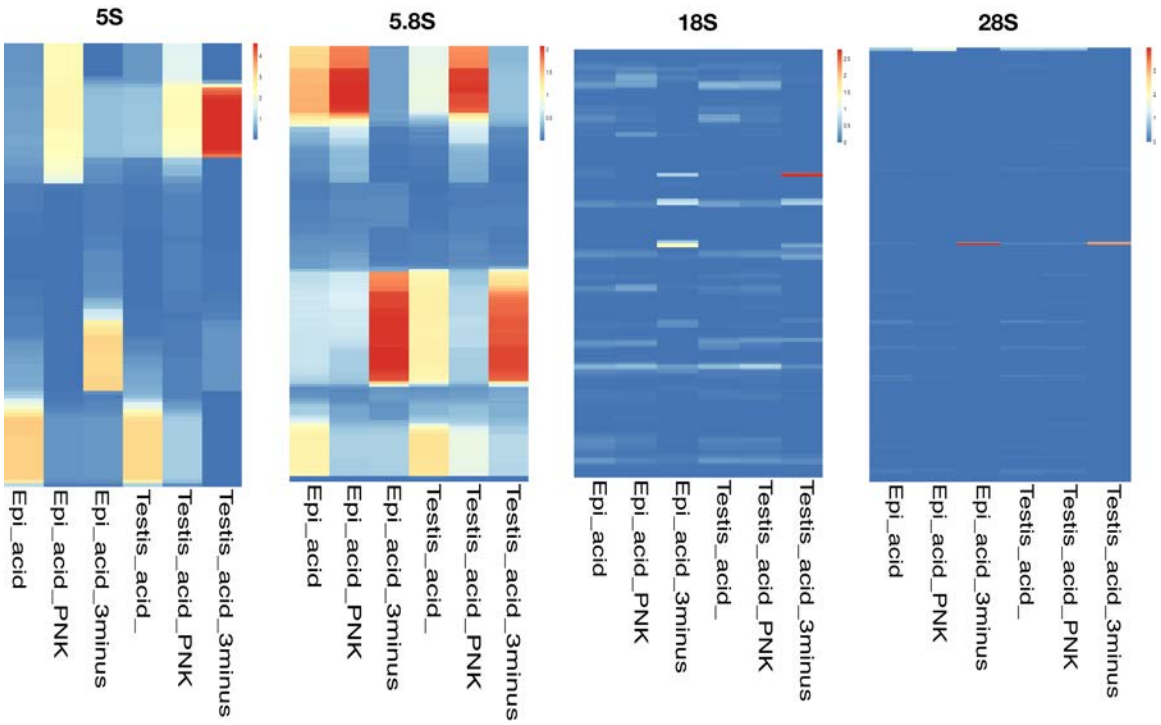


Fig. A1.3. Certain regions of rRNA generate specific types of fragments with distinct ends. Note tissue differences are minor (Epi: epididymis). Potential explanations for these specific cleavage events include *in vivo* structure of rRNA and modifications.

Enzymatic treatment of sRNAs also led to substantial changes in the species of tRFs preferentially cloned (Fig. A1.4, Fig. A1.5). PNK treatment revealed more tRF-5a and tRF-3a species (cleavage in and around the D loop and T loop, respectively) for a variety of tRNA from various anticodons, suggesting that RNase A family enzymes also cleave tRFs in the D and T loops, in addition to the anticodon loop. Importantly, in regards to 5tiRs and 3tiRs, PNK treatment also revealed a substantial amount of full-length RNase-A mediated cleavage products (Fig. A1.4). Interestingly, tRNAs of different anticodons have distinct enrichment of specific tRF species, and other potential tRF species cannot be captured regardless of enzymatic treatments. For example, tRFs derived from GlyGCC are mostly 5tiRs (Fig. A1.5B), while tRFs derived from

IleAAT are mostly 3tiRs (Fig. A1.5D). It is also important to note that the majority of 3’tRFs are derived from mature tRNAs, as mapping to tRNA sequences with an addition of CCA increased tRNA mapping reads approximately 2 fold in testis sRNAs and 50% in epididymis sRNAs.

Taken together, our data support a model whereby tRFs are cleaved by specific RNases from mostly mature tRNAs, and specific species of tRFs are relatively stable as compared to the other segments derived from the same tRNA. The stability of these tRFs may derive from their structure after cleavage, the various modifications they carry, whether they become bound by specific proteins and therefore become shielded from further degradation, and the concentration of specific RNases in a given tissue. More work is needed to elucidate the specific mechanism by which various tRF species are generated and preferentially retained in specific cell types, including whether these tRFs support different biological functions through their interactions with different proteins.

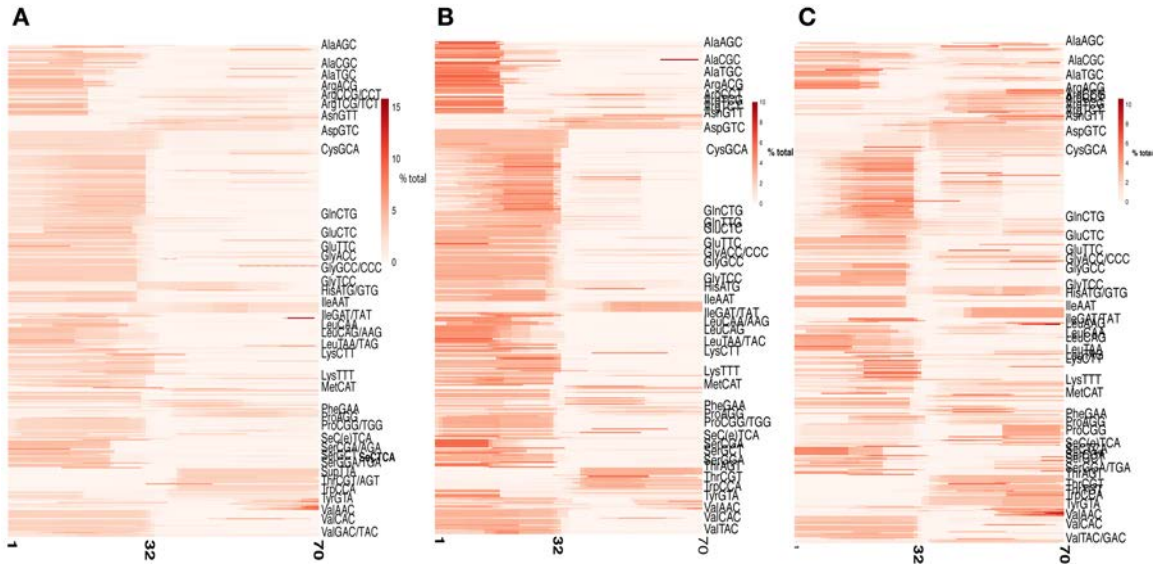


Fig. A1.4. Effect of enzymatic treatment on tRF coverage. Cauda sperm was either not treated (A), treated with T4 PNK (B), or with PNK 3minus (C). Coverage of tRNAs with the same anticodon sequence was averaged across the RNA.

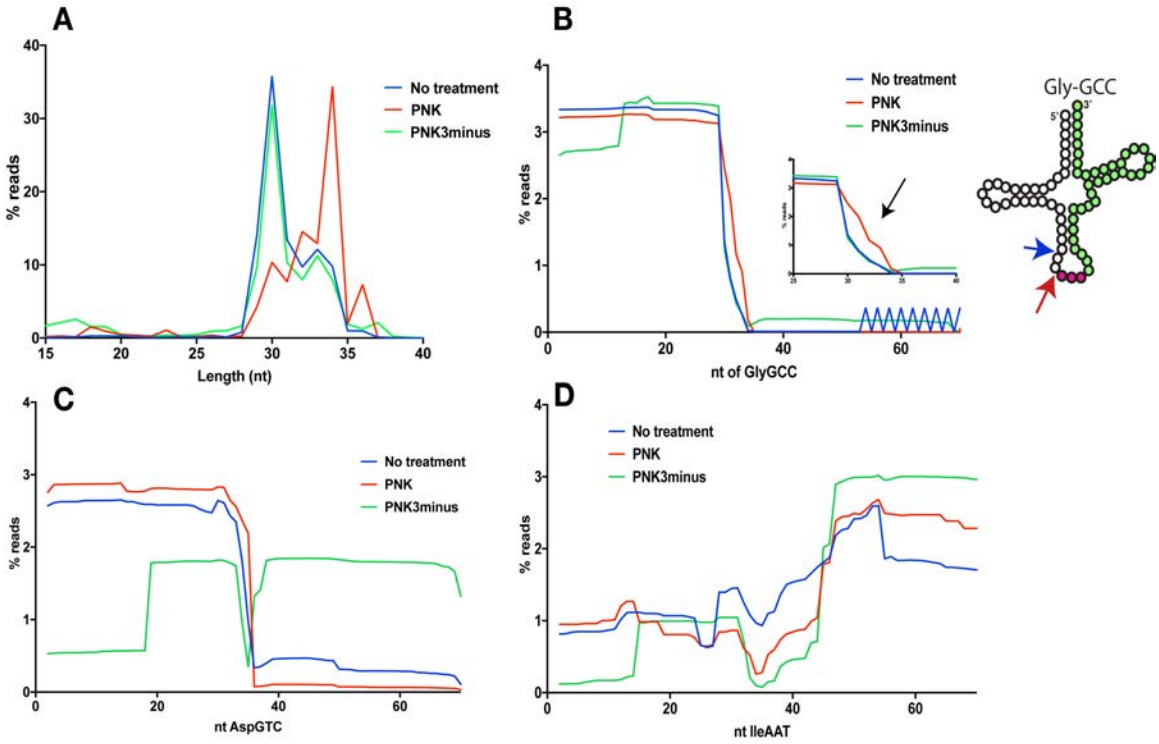


Fig. A1.5. PNK treatment leads to preferential cloning of full length cleavage products (5tiRs) at the anticodon loop, whereas 3'minus reveals 3tiRs derived from a subset of tRNAs.

Materials and Methods

Small RNA cloning

Small RNAs 18-40nt were size selected from denaturing polyacrylamide gel, and subjected to 30mins of enzymatic treatment by T4 PNK or T4 PNK 3'-phosphatase minus (NEB), or directly subjected to cloning using the TruSeq Small RNA Library Preparation Kit (Illumina). Libraries were pooled and deep sequenced on the Nextseq 500 (Illumina) with single-end 75nt reads to cover the extent of all sRNAs cloned.

Data analysis

Data was analyzed using the "seqmapping" method as previously described (Sharma *et al.*, 2016) to avoid repeated mapping of reads to various repetitive elements. Briefly, adapters are removed from the 3' end using the FastX toolkit (Hannon lab). Reads are then sequentially mapped to different repetitive classes using Bowtie2 (2.1.0), with mapped reads removed each round and a new Fastq file generated to remap to the new category. The number of mapped reads for each sample were counted using Samtools idxstats (v1.3). sRNA sizes of various classes was acquired by first converting bam files into Fastq files, and then sizes were compiled using the following command (`awk 'NR%4 == 2 {lengths[length($0)]++} END {for (l in lengths) {print l, lengths[l]}}' mapped_reads.fastq > mapped_reads_sizes.txt`). Coverage of reads on various categories of repeat elements was compiled using the command (`bedtools coverage -a rRNA18S.bed -b mapped_reads.bam -d > mapped_18S_cov.txt`). Most tRNA anticodons contain numerous copies throughout the genome, and all reads mapping to the same anticodon were combined in these analysis.

APPENDIX II

tRF-proteome dynamics in mESCs

Introduction

Well-characterized small RNAs (sRNAs) such as miRNAs, siRNAs, and piRNAs have very defined biogenesis pathways and lengths, and all associate with Argonaute family proteins (Ghildiyal and Zamore, 2009). Recently, it has been found that sRNAs cleaved from tRNAs, known as tRNA fragments or tRFs, are found in various organisms and have been shown to support diverse functions (Keam and Hutvagner, 2015; Kumar, Kuscu and Dutta, 2016). The wide range of proposed functions could be potentially attributed to the heterogeneous nature of tRFs, in length, sequence, and modifications, depending on which part of the tRNA they are derived from (Kumar, Kuscu and Dutta, 2016). Consequently, a variety of different biogenesis pathways have been identified, and a myriad of protein effectors or binders have been identified for different types of tRFs.

Recently, a technique called gradient profiling by sequencing (Grad-seq) was described which allowed for analysis of the full ensemble of *E. coli* cellular RNAs in a functional context, based on their biochemical profiles (Smirnov *et al.*, 2016). Here, I describe my efforts to adapt this technique for unbiased tRF-bound RNP discovery in mouse embryonic stem cells (mESCs). I find that tRFs derived from different tRNAs interact with different protein complexes, supporting the notion that tRFs can facilitate a multitude of biological functions. I found that tRF-Gly-GCC (tRF-GG), a tRF of interest in the lab, does not overlap with Mir34c and

associated RISC complex, arguing against Argonaute-based function for tRF-GG. I also discuss potential ways in which the current protocol can be improved to gain a better handle on specific tRF-bound proteins.

Results and Discussion

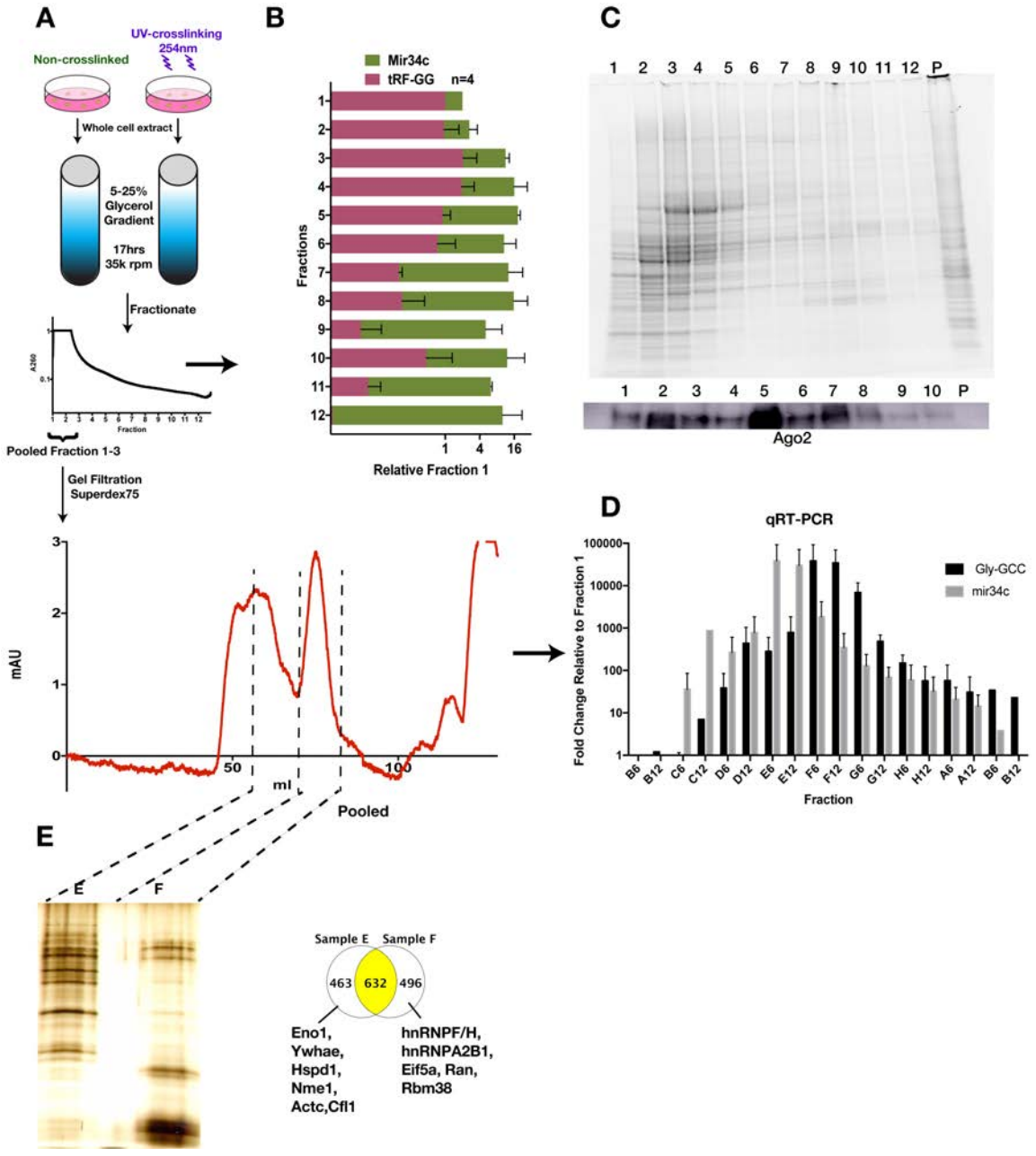


Fig. A2.1. Grad-seq reveals potential binders to tRF-Gly-GCC. **A.** Grad-seq workflow. **B.** RNA was extracted from fractions and Taqman qRT-PCR was conducted for indicated sRNAs. **C.** PAGE analysis of proteins in various

fractions. Ago2 levels within each fraction was analyzed using western blot. **D.** qRT-PCR analysis of tRF-GG and Mir34c from fractions collected after gel filtration. **E.** Silver staining of proteins pooled from E and F fractions after gel filtration, and the associated proteomic analysis of each fraction.

Identifying tRF-GG associated RNPs

The first aim to adapting Grad-seq in mESC extract was to identify the protein(s) bound to tRF-Gly-GCC (tRF-GG). Grad-seq utilizes a 10-40% (wt/vol) glycerol gradient, which was not able to separate sRNAs well, with most sRNAs (including miRNAs and tRFs) in the first one or two fractions (which translates to about 1/12 total volume, data not shown). Therefore, I adjusted the gradient to 5-25% wt/vol glycerol, and incorporated a longer ultracentrifugal spin. This resulted in better resolution of tRF-GG from miRNA-34c (Mir34c, Fig. A2.1A and B), a sRNA that should be bound to Argonaute proteins, specifically Ago2, in mESCs (Greve, Judson and Blelloch, 2013). To show that protein complexes maintain stability with associated sRNAs during the gradient protocol, I probed for enrichment of Ago2, a protein that binds miRNAs in mESCs. As expected, Ago2 was mostly found in fractions 5-7 of the gradient, co-localizing with Mir34c as analyzed by qRT-PCR (Fig. A2.1B and C). Various RNA-binding protein centric techniques utilize UV crosslinking to more stably associated RNAs with their binding proteins. However, I found that crosslinking of mouse extract with a 254nm UV crosslinker did not lead to noticeable changes in protein profiles through the fractions as assayed by polyacrylamide gel electrophoresis and qRT-PCR of tRF-GG and Mir34c (data not shown), although it did lead to problems detecting epitopes in western blotting. Since tRF-GG appears to be relatively

stable through multiple fractionation techniques, crosslinking may therefore be unnecessary to maintain binding stability of tRF-GG, if not most tRFs.

To further partition tRF-GG interacting protein(s) from irrelevant proteins of a similar size, I added a gel-filtration step, which was able to further separate different types of sRNAs (Fig. A2.1A and D). Proteomic analysis of samples pooled from the fractions where sRNAs of interest were enriched revealed divergent types of proteins within each fraction (Fig. A2.1E), including hnRNPH1/2, which we identified using other techniques as a bona-fide binder to tRF-GG. Indeed, hnRNPH1/2 was enriched in Fraction F vs. E by 22:4 distinct mass spectra, supporting the biochemical relevance of the gradient profiling technique.

Together, this data suggests that the Grad-seq protocol is able to separate RNPs associated with various sRNAs based on size. In particular, the gradient is able to segregate RNPs bound to tRF-GG from that bound to Mir34c, strongly arguing against an Argonaute based mechanism for tRF-GG function. However, a major caveat is that there remain hundreds of distinct proteins in each fraction, many of which are RNA-binding proteins (Fig. A2.1E). One potential way to better enrich for specific tRFs, and therefore tRF-bound RNPs, is by pre-incubating the cell extract with biotinylated oligos, then subject the extract to a glycerol gradient followed by gel filtration if needed, to reduce the potential background normally observed in biotinylation-based purification of RNPs.

Gradient profiling reveals the diversity of tRF-bound RNPs

To characterize the diversity of tRFs associated with various RNPs, I carried out deep sequencing of sRNAs isolated from select fractions of the gradient (*i.e.* Grad-seq). This revealed a wide distribution of different types of tRFs in various fractions of the gradient (Fig. A2.2A). tRFs derived from Gly-GCC were enriched in fraction 1 and 3, confirming previous qRT-PCR analysis (Fig. A2.1B, A2.2C). tRFs derived from Alanine anticodons and Gly-CCC, which have been shown to be associated with YBX1 *in vivo* (Ivanov, Mohamed M. Emara, *et al.*, 2011b; Goodarzi, Liu, Hoang C.B. Nguyen, *et al.*, 2015b), were predominantly enriched in fraction 1 and fraction 6. This could possibly result from binding of tRF-Ala with various protein complexes (Ivanov, Mohamed M. Emara, *et al.*, 2011b). The majority of tRF species were enriched in fraction 6, where Mir34c and Ago2 is enriched, suggesting these tRFs could potentially be bound to and function with the RISC complex or protein complexes of similar density (Fig. A2.2A).

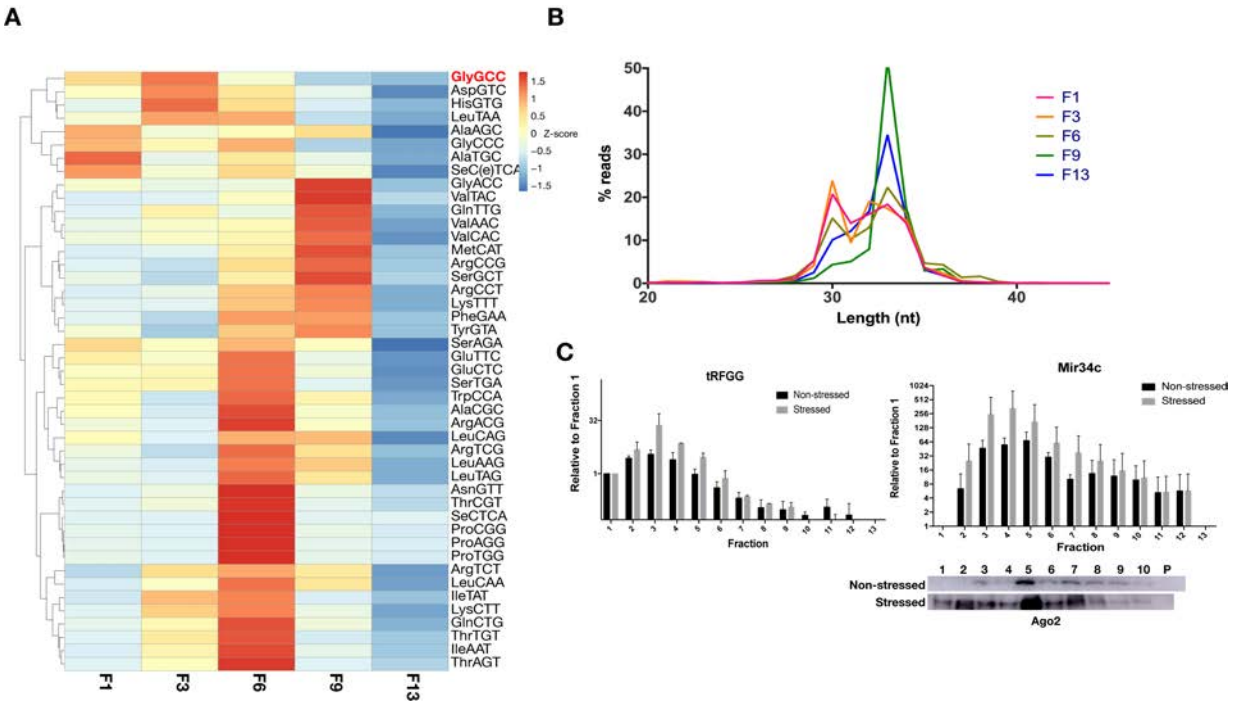


Fig. A2.2. Diversity of tRF-bound RNPs. **A.** Enrichment of tRFs derived from various anticodons throughout the gradient. Note all tRFs are depleted from the pellet (F13), which contains all ribosomes, confirmed via western blotting (data not shown). **B.** Fragment size distribution of tRFs throughout the fractions. **C.** Arsenite stress induces tRF-GG and Mir34c. Note the minimal redistribution of sRNAs across the fractions.

Interestingly, there is also a change in the size distribution of tRFs found in various fractions: while fractions 1-6 had two peaks at 30nt and ~33nt, fractions 9-13 were predominantly composed of tRFs that are 33nt long. This suggests that full length 5' and 3'tiRNAs resulting from RNase A family protein (possibly ANG) cleavage are predominantly associated with very large protein complex more dense than the RISC complex but less than ribosomes. Future work focused on isolating RNPs bound to tRFs enriched in these fractions could elucidate novel mechanisms of tRF function in mESCs.

Lastly, it has been shown that various stresses can induce stress granule production and cleavage of tRNAs (Fu *et al.*, 2009a; Yamasaki, Ivanov, G.-F. Hu,

et al., 2009), and that stress associated tRFs predominantly play a role in translation inhibition (Ivanov, Mohamed M. Emara, *et al.*, 2011b; Goncalves *et al.*, 2016). To interrogate whether stress induces tRF production in mESCs, I briefly treated mESCs with sodium arsenite for 1 hour. One hour treatment was chosen because 2 hour treatments at 500 uM led to extensive cell death, suggesting mESCs are highly sensitive to arsenite stress. I conducted qRT-PCR of gradient separated fractions following the Grad-seq protocol. Interestingly, I observed up-regulation of tRF-GG and Mir34c in arsenite stress conditions (Fig. A2.2C), suggesting that arsenite stress might induce production of not only tRFs, but also miRNAs. Furthermore, I failed to observe enrichment of tRF-GG with ribosomes (F13, or pellet), arguing against tRF-GG playing any role in translation inhibition during stress induced translational repression (see also Chapter II) (Sharma *et al.*, 2016).

Taken together, Grad-seq profiling in mESCs for sRNAs revealed the diversity of tRF-bound RNPs (Fig. A2.2). The preferential enrichment of certain tRF species in specific fractions strongly supports the idea that different tRFs from different species can play a multitude of biological functions, through their interactions with a wide range of RNPs, and is a strong refutation of the argument that tRFs are degradation products with little biological function. The discovery that specific lengths of tRFs are associated with different protein complexes also supports that the idea that tRFs derived from different biogenesis pathways play separate functions *in vivo*. The gradient separation of RNPs also

holds promise as an alternative approach to purification of specific tRFs and their bound RNPs.

Materials and Methods

Grad-seq

UV crosslinking was done at 254nm in ice cold PBS for 8000 µJ/cm$^{3X100.}$ Sodium arsenite stress was done in normal ES cell media at 500uM for one hour.

About 100 million mouse embryonic stem cells (3 X 150mM dish) were scraped and washed in ice cold PBS, and lysed in 1mL of ice cold lysis buffer (10mM Tris-HCl pH7.5, 5mM MgCl2, 100mM KCl, 1% Triton-X, 2mM DTT, and EDTA-free Protease Inhibitor Cocktial (Roche)). After resuspension and lysis by repeated pipetting, the cells were allowed to lyse on ice for 10 mins, then cell debris was removed by 10min 1100g centrifugation. The supernatant can be snap frozen and stored, or further concentrated using Amicon Ultra 5k. The extract is overlay onto a 13ml 5-25% (wt/vol) glycerol gradient (20mM HEPES-KOH pH7.5, 5mM MgCl2, 100mM KCl, 2mM DTT, glycerol wt/vol). The gradient was premade and allowed to chill to 4°C. The gradient was then spun at 35k, 4°C for 17hrs (overnight). Fractions were separated by hand into 960uL fractions, with the last fraction being the pellet (ribosomes).

Each fraction is separated into two for protein and RNA isolation. RNA was extracted by adding SDS to 1% final, Proteinase K to final concentration of 200ug/mL, incubated at 42°C for 45mins with gentle mixing. The RNA was separated from proteins by acid phenol/chloroform, then spun at 12000g for 5mins in a phase lock tube. The supernatant was added to 0.1V of NaAc (3M,

pH5.2), 10mM final of MgCl2, and 1uL GlycoBlue, and mixed with 1V of 100% isopropanol. The whole mixture is placed at -20°C for at least 20mins, then pelleted at Vmax for 35mins, washed with 500uL of 70% ethanol. The RNA pellet was air dried and resuspended in ddH2O.

Proteins can be concentrated on an Amicon Ultra 5K for western blotting, or TCA precipitated. This concentration step was particularly important for gel filtration. Gel filtration was done with 1ml concentrated solution of 3mL pooled fractions 1-3. This concentrate was applied to a Superdex 75 Increase 10/300 GL column with a flow rate of 0.25ml/min with the glycerol gradient buffer at room temperature. Fractions were collected in 1mL aliquots for 130 fractions (total volume 130mL, or 5X bed volume). Proteins or RNA were isolated from fractions as previously described, then subject to qRT-PCR.

Deep Sequencing

sRNAs 15-45nt were isolated using a denaturing PAGE gel and extracted overnight with RNA extraction buffer (300mM NaCl, 1mM EDTA). The sRNA was ethanol precipitated and resuspended in H2O, and followed up with cloning using Truseq Small RNA Library Prep Kit (Illumina). Libraries were pooled and sequenced on the Nextseq 500. Data analysis was done as described in previous sections.

Taqman qRT-PCR

tRF and miRNA quantification was performed using custom designed Taqman MicroRNA Assays according to manufacturer's instructions (Applied Biosystems). 1ng of total RNA from each fraction was reverse transcribed using

the Taqman MicroRNA reverse transcription kit. qRT-PCR was performed in 15uL reactions using Taqman Universal PCR Master Mix, following standard protocol. Serial dilutions of template were previously run to confirm amplification linearity and efficiency for all Taqman probes.

Bibliography

Aebi, M. *et al.* (1990) ‘Isolation of a temperature-sensitive mutant with an altered tRNA nucleotidyltransferase and cloning of the gene encoding tRNA nucleotidyltransferase in the yeast Saccharomyces cerevisiae.’, *The Journal of biological chemistry*, 265(27), pp. 16216–20. Available at: http://www.ncbi.nlm.nih.gov/pubmed/2204621 (Accessed: 7 May 2019).

Agrawal, A. A., Laforsch, C. and Tollrian, R. (1999) ‘Transgenerational induction of defences in animals and plants’, *Nature*. Nature Publishing Group, 401(6748), pp. 60–63. doi: 10.1038/43425.

Aiken, C. E. and Ozanne, S. E. (2014) ‘Transgenerational developmental programming’, *Human Reproduction Update*. Narnia, 20(1), pp. 63–75. doi: 10.1093/humupd/dmt043.

Alabert, C. and Groth, A. (2012) ‘Chromatin replication and epigenome maintenance’, *Nature Reviews Molecular Cell Biology*. Nature Publishing Group, 13(3), pp. 153–167. doi: 10.1038/nrm3288.

Alkan, S. A., Martincic, K. and Milcarek, C. (2006) ‘The hnRNPs F and H2 bind to similar sequences to influence gene expression.’, *The Biochemical journal*. Portland Press Limited, 393(Pt 1), pp. 361–71. doi: 10.1042/BJ20050538.

Almouzni, G. and Wolffe, A. P. (1995) ‘Constraints on transcriptional activator function contribute to transcriptional quiescence during early Xenopus embryogenesis.’, *The EMBO journal*. European Molecular Biology Organization, 14(8), pp. 1752–65. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7737126 (Accessed: 8 May 2019).

Ambrosio, L. *et al.* (2014) ‘Phylogenetic Analyses and Characterization of RNase X25 from Drosophila melanogaster Suggest a Conserved Housekeeping Role and Additional Functions for RNase T2 Enzymes in Protostomes’, *PLoS ONE*. Edited by D. Boudko. Public Library of Science, 9(8), p. e105444. doi: 10.1371/journal.pone.0105444.

Amodeo, A. A. *et al.* (2015) ‘Histone titration against the genome sets the DNA-to-cytoplasm threshold for the *Xenopus* midblastula transition’, *Proceedings of the National Academy of Sciences*, 112(10), pp. E1086–E1095. doi: 10.1073/pnas.1413990112.

Anderson, L. M. *et al.* (2006) ‘Preconceptional fasting of fathers alters serum glucose in offspring of mice’, *Nutrition*. Elsevier, 22(3), pp. 327–331. doi: 10.1016/J.NUT.2005.09.006.

Andrés, F. and Coupland, G. (2012) ‘The genetic basis of flowering responses to seasonal cues’, *Nature Reviews Genetics*. Nature Publishing Group, 13(9), pp. 627–639. doi: 10.1038/nrg3291.

Anway, M. D. *et al.* (2005) ‘Epigenetic transgenerational actions of endocrine disruptors and male fertility.’, *Science (New York, N.Y.)*. American Association for the Advancement of Science, 308(5727), pp. 1466–9. doi: 10.1126/science.1108190.

Aravin, A. *et al.* (2006) ‘A novel class of small RNAs bind to MILI protein in mouse testes’, *Nature*. Nature Publishing Group, 442(7099), pp. 203–207. doi: 10.1038/nature04916.

Aravin, A. A. *et al.* (2007) 'Developmentally Regulated piRNA Clusters Implicate MILI in Transposon Control', *Science*, 316(5825), pp. 744–747. doi: 10.1126/science.1142612.
Aravin, A. A. *et al.* (2008) 'A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice.', *Molecular cell*. NIH Public Access, 31(6), pp. 785–99. doi: 10.1016/j.molcel.2008.09.003.
Arney, K. L. *et al.* (2002) *Histone methylation defines epigenetic asymmetry in the mouse zygote*, *Int. J. Dev. Biol*. Available at: www.ijdb.ehu.es (Accessed: 12 April 2019).
Arteaga-Vazquez, M. A. and Chandler, V. L. (2010) 'Paramutation in maize: RNA mediated trans-generational gene silencing', *Current Opinion in Genetics & Development*. Elsevier Current Trends, 20(2), pp. 156–163. doi: 10.1016/J.GDE.2010.01.008.
Ashe, A. *et al.* (2012) 'piRNAs Can Trigger a Multigenerational Epigenetic Memory in the Germline of C. elegans', *Cell*. Cell Press, 150(1), pp. 88–99. doi: 10.1016/J.CELL.2012.06.018.
Autran, D. *et al.* (2011) 'Maternal Epigenetic Pathways Control Parental Contributions to Arabidopsis Early Embryogenesis', *Cell*. Cell Press, 145(5), pp. 707–719. doi: 10.1016/J.CELL.2011.04.014.
Babiarz, J. E. *et al.* (2008) 'Mouse ES cells express endogenous shRNAs, siRNAs, and other Microprocessor-independent, Dicer-dependent small RNAs.', *Genes & development*, 22(20), pp. 2773–85. doi: 10.1101/gad.1705308.
Baker, C. L. and Pera, M. F. (2018) 'Cell Stem Cell Review Capturing Totipotent Stem Cells'. doi: 10.1016/j.stem.2017.12.011.
Bąkowska-Żywicka, K. *et al.* (2016) 'The widespread occurrence of tRNA-derived fragments in *Saccharomyces cerevisiae*', *FEBS Open Bio*. John Wiley & Sons, Ltd, 6(12), pp. 1186–1200. doi: 10.1002/2211-5463.12127.
Bale, T. L. (2015) 'Epigenetic and transgenerational reprogramming of brain development', *Nature Reviews Neuroscience*, 16(6), pp. 332–344. doi: 10.1038/nrn3818.
Balhorn, R. (2007) 'The protamine family of sperm nuclear proteins', *Genome Biology*. BioMed Central, 8(9), p. 227. doi: 10.1186/gb-2007-8-9-227.
Barlow, D. P. and Bartolomei, M. S. (2014) 'Genomic imprinting in mammals.', *Cold Spring Harbor perspectives in biology*. Cold Spring Harbor Laboratory Press, 6(2), p. a018382. doi: 10.1101/cshperspect.a018382.
Bartolomei, M. S. and Ferguson-Smith, A. C. (2011) 'Mammalian genomic imprinting.', *Cold Spring Harbor perspectives in biology*. Cold Spring Harbor Laboratory Press, 3(7), p. a002592. doi: 10.1101/cshperspect.a002592.
Bateson, P. (2001) 'Fetal experience and good adult design.', *International journal of epidemiology*, 30(5), pp. 928–34. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11689495 (Accessed: 11 April 2019).
Bateson, W. and Pellew, C. (1915) 'On the genetics of "Rogues" among culinary peas (Pisum sativum)', *Journal of Genetics*. Springer India, 5(1), pp. 13–36. doi: 10.1007/BF02982150.
Battle, D. J. *et al.* (2006) 'The SMN Complex: An Assembly Machine for RNPs', *Cold Spring Harbor Symposia on Quantitative Biology*, 71(0), pp. 313–320. doi:

10.1101/sqb.2006.71.001.
Beach, R. S., Gershwin, M. E. and Hurley, L. S. (1982) ‘Gestational zinc deprivation in mice: persistence of immunodeficiency for three generations.’, *Science (New York, N.Y.)*, 218(4571), pp. 469–71. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7123244 (Accessed: 21 May 2019).
Belele, C. L. *et al.* (2013) ‘Specific Tandem Repeats Are Sufficient for Paramutation-Induced Trans-Generational Silencing’, *PLoS Genetics*. Edited by N. M. Springer. Public Library of Science, 9(10), p. e1003773. doi: 10.1371/journal.pgen.1003773.
Belleannée, C. *et al.* (2013) ‘Epididymosomes Convey Different Repertoires of MicroRNAs Throughout the Bovine Epididymis1’, *Biology of Reproduction*, 89(2), p. 30. doi: 10.1095/biolreprod.113.110486.
Bellve, A. *et al.* (1977) ‘Spermatogenic cells of the prepuberal mouse: isolation and morphological characterization’, *The Journal of Cell Biology*, 74(1), pp. 68–85. doi: 10.1083/jcb.74.1.68.
Bénit, L. *et al.* (1999) ‘ERV-L elements: a family of endogenous retrovirus-like elements active throughout the evolution of mammals.’, *Journal of virology*, 73(4), pp. 3301–3308.
Benyshek, D. C. *et al.* (2008) ‘Insulin sensitivity is normalized in the third generation (F3) offspring of developmentally programmed insulin resistant (F2) rats fed an energy-restricted diet’, *Nutrition & Metabolism*, 5(1), p. 26. doi: 10.1186/1743-7075-5-26.
Berger, S. L. *et al.* (2009) ‘An operational definition of epigenetics.’, *Genes & development*. Cold Spring Harbor Laboratory Press, 23(7), pp. 781–3. doi: 10.1101/gad.1787609.
Biémont, C. and Voytas, D. F. (2010) ‘A brief history of the status of transposable elements: from junk DNA to major players in evolution.’, *Genetics*. Genetics, 186(4), pp. 1085–93. doi: 10.1534/genetics.110.124180.
Birkedal, U. *et al.* (2014) ‘Profiling of Ribose Methylations in RNA by High-Throughput Sequencing’, *Angewandte Chemie International Edition*. John Wiley & Sons, Ltd, 54(2), p. n/a-n/a. doi: 10.1002/anie.201408362.
Blanco, S. *et al.* (2014) ‘Aberrant methylation of tRNAs links cellular stress to neuro-developmental disorders.’, *The EMBO journal*, 33(18), pp. 2020–39. doi: 10.15252/embj.201489282.
Blanco, S. *et al.* (2016) ‘Stem cell function and stress response are controlled by protein synthesis’, *Nature*. Nature Publishing Group, 534(7607), pp. 335–340. doi: 10.1038/nature18282.
Bogu, G. K. *et al.* (2015) ‘Chromatin and RNA Maps Reveal Regulatory Long Noncoding RNAs in Mouse.’, *Molecular and cellular biology*. American Society for Microbiology Journals, 36(5), pp. 809–19. doi: 10.1128/MCB.00955-15.
Bond, U. M., Yario, T. A. and Steitz, J. A. (1991) *Multiple processing-defective mutations m a mammahan histone pre-mRNA are suppressed by compensatory changes in U7 RNA both in vivo and in vitro*. Available at: http://genesdev.cshlp.org/content/5/9/1709.full.pdf (Accessed: 16 April 2019).
Bongiorno-Borbone, L. *et al.* (2008) ‘FLASH and NPAT positive but not Coilin positive Cajal bodies correlate with cell ploidy’, *Cell Cycle*. Taylor & Francis,

7(15), pp. 2357–2367. doi: 10.4161/cc.6344.
Bonnefoy, E. *et al.* (2007) ‘The Essential Role of Drosophila HIRA for De Novo Assembly of Paternal Chromatin at Fertilization’, *PLoS Genetics*, 3(10), p. e182. doi: 10.1371/journal.pgen.0030182.
Borges, F. and Martienssen, R. A. (2015) ‘The expanding world of small RNAs in plants’, *Nature Reviews Molecular Cell Biology*. Nature Publishing Group, 16(12), pp. 727–741. doi: 10.1038/nrm4085.
Bošković, A. B. and Rando, O. J. (2018) ‘Transgenerational Epigenetic Inheritance’. doi: 10.1146/annurev-genet-120417.
Boskovic, A. *et al.* (2014) ‘Higher chromatin mobility supports totipotency and precedes pluripotency in vivo’, *Genes & Development*. Cold Spring Harbor Laboratory Press, 28(10), pp. 1042–1047. doi: 10.1101/gad.238881.114.
Bošković, A. and Rando, O. J. (2018) ‘Transgenerational Epigenetic Inheritance’, *Annual Review of Genetics*. Annual Reviews , 52(1), pp. 21–41. doi: 10.1146/annurev-genet-120417-031404.
Bostick, M. *et al.* (2007) ‘UHRF1 Plays a Role in Maintaining DNA Methylation in Mammalian Cells’, *Science*, 317(5845), pp. 1760–1764. doi: 10.1126/science.1147939.
Bourc’his, D. and Bestor, T. H. (2004) ‘Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L’, *Nature*. Nature Publishing Group, 431(7004), pp. 96–99. doi: 10.1038/nature02886.
Bourc’his, D. and Voinnet, O. (2010) ‘A small-RNA perspective on gametogenesis, fertilization, and early zygotic development.’, *Science (New York, N.Y.)*. American Association for the Advancement of Science, 330(6004), pp. 617–22. doi: 10.1126/science.1194776.
Brink, R. A. (1956) ‘A GENETIC CHANGE ASSOCIATED WITH THE R LOCUS IN MAIZE WHICH IS DIRECTED AND POTENTIALLY REVERSIBLE’, *Genetics*, 41(6), pp. 872–889. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1224369/pdf/872.pdf (Accessed: 10 April 2019).
Brink, R. A. (1958) ‘Paramutation at the R locus in maize.’, *Cold Spring Harbor symposia on quantitative biology*. Cold Spring Harbor Laboratory Press, 23, pp. 379–91. doi: 10.1101/SQB.1958.023.01.036.
Brink, R. A., Styles, E. D. and Axtell, J. D. (1968) ‘Paramutation: directed genetic change. Paramutation occurs in somatic cells and heritably alters the functional state of a locus.’, *Science (New York, N.Y.)*, 159(3811), pp. 161–70. Available at: http://www.ncbi.nlm.nih.gov/pubmed/5634904 (Accessed: 18 May 2019).
Brooks, L. *et al.* (2015) ‘A multiprotein occupancy map of the mRNP on the 3’ end of histone mRNAs.’, *RNA (New York, N.Y.)*. Cold Spring Harbor Laboratory Press, 21(11), pp. 1943–65. doi: 10.1261/rna.053389.115.
Brykczynska, U. *et al.* (2010) ‘Repressive and active histone methylation mark distinct promoters in human and mouse spermatozoa’, *Nature Structural & Molecular Biology*, 17(6), pp. 679–687. doi: 10.1038/nsmb.1821.
Buenrostro, J. D. *et al.* (2015) ‘ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide’, in *Current Protocols in Molecular Biology*. Hoboken, NJ, USA: John Wiley & Sons, Inc., pp. 21.29.1-21.29.9. doi:

10.1002/0471142727.mb2129s109.
Bulchand, S. *et al.* (2010) ‘Muscle wasted: a novel component of the Drosophila histone locus body required for muscle integrity’, *Journal of Cell Science*, 123(16), pp. 2697–2707. doi: 10.1242/jcs.063172.
Burch, B. D. *et al.* (2011) ‘Interaction between FLASH and Lsm11 is essential for histone pre-mRNA processing in vivo in Drosophila.’, *RNA (New York, N.Y.)*. Cold Spring Harbor Laboratory Press, 17(6), pp. 1132–47. doi: 10.1261/rna.2566811.
Büssing, I., Slack, F. J. and Grosshans, H. (2008) ‘let-7 microRNAs in development, stem cells and cancer.’, *Trends in molecular medicine*. Elsevier, 14(9), pp. 400–9. doi: 10.1016/j.molmed.2008.07.001.
Cameron, V. and Uhlenbeck, O. C. (1977) ‘3’-Phosphatase activity in T4 polynucleotide kinase.’, *Biochemistry*, 16(23), pp. 5120–6. Available at: http://www.ncbi.nlm.nih.gov/pubmed/199248 (Accessed: 25 March 2019).
Carmo-Fonseca, M. *et al.* (1992) ‘Transcription-dependent colocalization of the U1, U2, U4/U6, and U5 snRNPs in coiled bodies.’, *The Journal of cell biology*. Rockefeller University Press, 117(1), pp. 1–14. doi: 10.1083/JCB.117.1.1.
Carone, B. R. *et al.* (2010) ‘Paternally induced transgenerational environmental reprogramming of metabolic gene expression in mammals’, *Cell*. Elsevier Inc., 143(7), pp. 1084–1096. doi: 10.1016/j.cell.2010.12.008.
Carone, B. R. *et al.* (2014) ‘High-Resolution Mapping of Chromatin Packaging in Mouse Embryonic Stem Cells and Sperm’, *Developmental Cell*. Cell Press, 30(1), pp. 11–22. doi: 10.1016/J.DEVCEL.2014.05.024.
Carvalho, T. *et al.* (1999) ‘The Spinal Muscular Atrophy Disease Gene Product, Smn’, *The Journal of Cell Biology*, 147(4), pp. 715–728. doi: 10.1083/jcb.147.4.715.
Castañeda, J. *et al.* (2014) ‘Reduced pachytene piRNAs and translation underlie spermiogenic arrest in Maelstrom mutant mice.’, *The EMBO journal*. EMBO Press, 33(18), pp. 1999–2019. doi: 10.15252/embj.201386855.
Castel, S. E. and Martienssen, R. A. (2013) ‘RNA interference in the nucleus: roles for small RNAs in transcription, epigenetics and beyond.’, *Nature reviews. Genetics*. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved., 14(2), pp. 100–12. doi: 10.1038/nrg3355.
Castella, S. *et al.* (2004) ‘Train A, an RNase A-like protein without RNase activity, is secreted and reabsorbed by the same epididymal cells under testicular control.’, *Biology of reproduction*, 71(5), pp. 1677–87. doi: 10.1095/biolreprod.104.031666.
Champroux, A. *et al.* (2018) ‘A Decade of Exploring the Mammalian Sperm Epigenome: Paternal Epigenetic and Transgenerational Inheritance’, *Frontiers in Cell and Developmental Biology*. Frontiers Media SA, 6, p. 50. doi: 10.3389/FCELL.2018.00050.
Chan, Y. B. *et al.* (2003) ‘Neuromuscular defects in a Drosophila survival motor neuron gene mutant.’, *Human molecular genetics*, 12(12), pp. 1367–76. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12783845 (Accessed: 15 April 2019).
Chandler, V. L. (2007) ‘Paramutation: From Maize to Mice’, *Cell*. Cell Press,

128(4), pp. 641–645. doi: 10.1016/J.CELL.2007.02.007.
Chandler, V. L. (2010) ‘Paramutation’s Properties and Puzzles’, *Science*, 330(6004), pp. 628–629. doi: 10.1126/science.1191044.
Chandler, V. L., Eggleston, W. B. and Dorweiler, J. E. (2000) ‘Paramutation in maize’, in *Plant Gene Silencing*. Dordrecht: Springer Netherlands, pp. 1–25. doi: 10.1007/978-94-011-4183-3_1.
Chandler, V. L. and Stam, M. (2004) ‘Chromatin conversations: mechanisms and implications of paramutation’, *Nature Reviews Genetics*. Nature Publishing Group, 5(7), pp. 532–544. doi: 10.1038/nrg1378.
Chandler, V. L. and Walbot, V. (1986) *DNA modification of a maize transposable element correlates with loss of activity (Robertson’s Mutator/anthocyanins/unstable bronze 2 mutant alleles)*, *Proc. Nati. Acad. Sci. USA*. Available at: https://www.pnas.org/content/pnas/83/6/1767.full.pdf (Accessed: 11 April 2019).
Cheloufi, S. *et al.* (2015) ‘The histone chaperone CAF-1 safeguards somatic cell identity’, *Nature*. Nature Publishing Group, 528(7581), pp. 218–224. doi: 10.1038/nature15749.
Chen, C.-L. and Perrimon, N. (2017) ‘Proximity-dependent labeling methods for proteomic profiling in living cells’:, *WIREs Dev Biol*, p. 272. doi: 10.1002/wdev.272.
Chen, Q., Yan, Menghong, *et al.* (2016) ‘Sperm tsRNAs contribute to intergenerational inheritance of an acquired metabolic disorder.’, *Science (New York, N.Y.)*. American Association for the Advancement of Science, 351(6271), pp. 397–400. doi: 10.1126/science.aad7977.
Chen, Q., Yan, M, *et al.* (2016) ‘Sperm tsRNAs contribute to intergenerational inheritance of an acquired metabolic disorder’, *Science*, 351(6271), pp. 397–400. doi: 10.1126/science.aad7977.
Chen, Z. *et al.* (2017) ‘Two featured series of rRNA-derived RNA fragments (rRFs) constitute a novel class of small RNAs’, *PLOS ONE*. Edited by Y. Xue. Public Library of Science, 12(4), p. e0176458. doi: 10.1371/journal.pone.0176458.
Cheng, G.-Z. *et al.* (2009) ‘Human ribonuclease 9, a member of ribonuclease A superfamily, specifically expressed in epididymis, is a novel sperm-binding protein’, *Asian Journal of Andrology*, 11(2), pp. 240–251. doi: 10.1038/aja.2008.30.
Cho, S., Beintema, J. J. and Zhang, J. (2005) ‘The ribonuclease A superfamily of mammals and birds: identifying new members and tracing evolutionary histories’, *Genomics*. Academic Press, 85(2), pp. 208–220. doi: 10.1016/J.YGENO.2004.10.008.
Choi, Y. J. *et al.* (2017) ‘Deficiency of microRNA *miR-34a* expands cell fate potential in pluripotent stem cells’, *Science*, 355(6325), p. eaag1927. doi: 10.1126/science.aag1927.
Chou, H.-J. *et al.* (2017) ‘Transcriptome-wide Analysis of Roles for tRNA Modifications in Translational Regulation’, *Molecular Cell*. Cell Press, 68(5), pp. 978-992.e4. doi: 10.1016/J.MOLCEL.2017.11.002.
Cirio, M. C. *et al.* (2008) ‘Preimplantation expression of the somatic form of

Dnmt1 suggests a role in the inheritance of genomic imprints', *BMC Developmental Biology*, 8(1), p. 9. doi: 10.1186/1471-213X-8-9.
Cole, C. *et al.* (2009) 'Filtering of deep sequencing data reveals the existence of abundant Dicer-dependent small RNAs derived from tRNAs.', *RNA (New York, N.Y.)*, 15(12), pp. 2147–60. doi: 10.1261/rna.1738409.
Collier, S. *et al.* (2006) 'A Distant Coilin Homologue Is Required for the Formation of Cajal Bodies in *Arabidopsis*', *Molecular Biology of the Cell*. Edited by A. G. Matera, 17(7), pp. 2942–2951. doi: 10.1091/mbc.e05-12-1157.
Conine, C. C. *et al.* (2018) 'Small RNAs Gained during Epididymal Transit of Sperm Are Essential for Embryonic Development in Mice', *Developmental Cell*. Cell Press, 46(4), pp. 470-480.e3. doi: 10.1016/J.DEVCEL.2018.06.024.
Connelly, C. M., Thomas, M. and Deiters, A. (2012) 'High-throughput luciferase reporter assay for small-molecule inhibitors of MicroRNA function', *Journal of Biomolecular Screening*. doi: 10.1177/1087057112439606.
Consortium, M. G. S. (2002) 'Initial sequencing and comparative analysis of the mouse genome', *Nature*. Nature Publishing Group, 420(6915), pp. 520–562. doi: 10.1038/nature01262.
Cora, E. *et al.* (2014) 'The MID-PIWI module of Piwi proteins specifies nucleotide- and strand-biases of piRNAs', *RNA*, 20(6), pp. 773–781. doi: 10.1261/rna.044701.114.
Cornwall, G. A. (2009) 'New insights into epididymal biology and function.', *Human reproduction update*. Oxford University Press, 15(2), pp. 213–27. doi: 10.1093/humupd/dmn055.
Cossec, J.-C. *et al.* (2018) 'SUMO Safeguards Somatic and Pluripotent Cell Identities by Enforcing Distinct Chromatin States', *Cell Stem Cell*. Cell Press, 23(5), pp. 742-757.e8. doi: 10.1016/J.STEM.2018.10.001.
Cotten, M. *et al.* (1988) 'Specific contacts between mammalian U7 snRNA and histone precursor RNA are indispensable for the in vitro 3' RNA processing reaction.', *The EMBO journal*. European Molecular Biology Organization, 7(3), pp. 801–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/3396543 (Accessed: 18 April 2019).
Couvillion, Mary T *et al.* (2012) 'A Tetrahymena Piwi bound to mature tRNA 3' fragments activates the exonuclease Xrn2 for RNA processing in the nucleus.', *Molecular cell*, 48(4), pp. 509–20. doi: 10.1016/j.molcel.2012.09.010.
Couvillion, Mary T. *et al.* (2012) 'A Tetrahymena Piwi Bound to Mature tRNA 3′ Fragments Activates the Exonuclease Xrn2 for RNA Processing in the Nucleus', *Molecular Cell*. Cell Press, 48(4), pp. 509–520. doi: 10.1016/J.MOLCEL.2012.09.010.
Couvillion, M. T., Sachidanandam, R. and Collins, K. (2010) 'A growth-essential Tetrahymena Piwi protein carries tRNA fragment cargo.', *Genes & development*. Cold Spring Harbor Laboratory Press, 24(24), pp. 2742–7. doi: 10.1101/gad.1996210.
Cozen, A. E. *et al.* (2015) 'ARM-seq: AlkB-facilitated RNA methylation sequencing reveals a complex landscape of modified tRNA fragments', *Nature Methods*. Nature Publishing Group, 12(9), pp. 879–884. doi: 10.1038/nmeth.3508.

Creasey, K. M. *et al.* (2014) ‘miRNAs trigger widespread epigenetically activated siRNAs from transposons in Arabidopsis’, *Nature*. Nature Publishing Group, 508(7496), pp. 411–415. doi: 10.1038/nature13069.
Curley, J. P., Mashoodh, R. and Champagne, F. A. (2011) ‘Epigenetics and the origins of paternal effects’, *Hormones and Behavior*, 59(3), pp. 306–314. doi: 10.1016/j.yhbeh.2010.06.018.
Dacheux, J.-L. and Dacheux, F. (2014) ‘New insights into epididymal function in relation to sperm maturation’, *REPRODUCTION*, 147(2), pp. R27–R42. doi: 10.1530/REP-13-0420.
Dahl, J. A. *et al.* (2016) ‘Broad histone H3K4me3 domains in mouse oocytes modulate maternal-to-zygotic transition’, *Nature*. Nature Publishing Group, 537(7621), pp. 548–552. doi: 10.1038/nature19360.
Dan, J. *et al.* (2017) ‘Zscan4 Inhibits Maintenance DNA Methylation to Facilitate Telomere Elongation in Mouse Embryonic Stem Cells’, *Cell Reports*. Cell Press, 20(8), pp. 1936–1949. doi: 10.1016/J.CELREP.2017.07.070.
Darzacq, X. (2002) ‘Cajal body-specific small nuclear RNAs: a novel class of 2’-O-methylation and pseudouridylation guide RNAs’, *The EMBO Journal*, 21(11), pp. 2746–2756. doi: 10.1093/emboj/21.11.2746.
Daxinger, L. and Whitelaw, E. (2012) ‘Understanding transgenerational epigenetic inheritance via the gametes in mammals’, *Nature Reviews Genetics*. Nature Publishing Group, 13(3), pp. 153–162. doi: 10.1038/nrg3188.
Deng, J. *et al.* (2015) ‘Respiratory Syncytial Virus Utilizes a tRNA Fragment to Suppress Antiviral Responses Through a Novel Targeting Mechanism’, *Molecular Therapy*. Cell Press, 23(10), pp. 1622–1629. doi: 10.1038/MT.2015.124.
Deng, W. and Lin, H. (2002) ‘miwi, a murine homolog of piwi, encodes a cytoplasmic protein essential for spermatogenesis.’, *Developmental cell*, 2(6), pp. 819–30. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12062093 (Accessed: 16 May 2019).
Deryusheva, S. and Gall, J. G. (2009) ‘Small Cajal Body–specific RNAs of *Drosophila* Function in the Absence of Cajal Bodies’, *Molecular Biology of the Cell*. Edited by A. G. Matera, 20(24), pp. 5250–5259. doi: 10.1091/mbc.e09-09-0777.
Di Giacomo, M. *et al.* (2013) ‘Multiple Epigenetic Mechanisms and the piRNA Pathway Enforce LINE1 Silencing during Adult Spermatogenesis’, *Molecular Cell*, 50(4), pp. 601–608. doi: 10.1016/j.molcel.2013.04.026.
Dias, B. G. and Ressler, K. J. (2014) ‘Parental olfactory experience influences behavior and neural structure in subsequent generations’, *Nature Neuroscience*. Nature Publishing Group, 17(1), pp. 89–96. doi: 10.1038/nn.3594.
Dietz, D. M. *et al.* (2011) ‘Paternal Transmission of Stress-Induced Pathologies’, *Biological Psychiatry*, 70(5), pp. 408–414. doi: 10.1016/j.biopsych.2011.05.005.
Ding, G.-L. *et al.* (2012) ‘Transgenerational glucose intolerance with Igf2/H19 epigenetic alterations in mouse islet induced by intrauterine hyperglycemia.’, *Diabetes*, 61(5), pp. 1133–42. doi: 10.2337/db11-1314.
Dion, M. F. *et al.* (2007) ‘Dynamics of Replication-Independent Histone Turnover in Budding Yeast’, *Science*, 315(5817), pp. 1405–1408. doi:

10.1126/science.1134053.
Dominguez-Salas, P. *et al.* (2014) ‘Maternal nutrition at conception modulates DNA methylation of human metastable epialleles’, *Nature Communications*, 5(1), p. 3746. doi: 10.1038/ncomms4746.
Dominski, Z. and Marzluff, W. F. (1999) ‘Formation of the 3′ end of histone mRNA’, *Gene*. Elsevier, 239(1), pp. 1–14. doi: 10.1016/S0378-1119(99)00367-4.
Dorweiler, J. E. *et al.* (2000) ‘mediator of paramutation1 is required for establishment and maintenance of paramutation at multiple maize loci.’, *The Plant cell*, 12(11), pp. 2101–18. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11090212 (Accessed: 18 March 2019).
Drake, A. J. *et al.* (2011) ‘Multigenerational programming in the glucocorticoid programmed rat is associated with generation-specific and parent of origin effects’, *Epigenetics*. Taylor & Francis, 6(11), pp. 1334–1343. doi: 10.4161/epi.6.11.17942.
Dubin, M. J., Mittelsten Scheid, O. and Becker, C. (2018) ‘Transposons: a blessing curse’, *Current Opinion in Plant Biology*. Elsevier Current Trends, 42, pp. 23–29. doi: 10.1016/J.PBI.2018.01.003.
Dunn, G. A. and Bale, T. L. (2009) ‘Maternal High-Fat Diet Promotes Body Length Increases and Insulin Insensitivity in Second-Generation Mice’, *Endocrinology*. Narnia, 150(11), pp. 4999–5009. doi: 10.1210/en.2009-0500.
Dunn, G. A. and Bale, T. L. (2011) ‘Maternal High-Fat Diet Effects on Third-Generation Female Body Size via the Paternal Lineage’, *Endocrinology*. Narnia, 152(6), pp. 2228–2236. doi: 10.1210/en.2010-1461.
Durdevic, Z. *et al.* (2013) ‘The RNA Methyltransferase Dnmt2 Is Required for Efficient Dicer-2-Dependent siRNA Pathway Activity in Drosophila’, *Cell Reports*, 4(5), pp. 931–937. doi: 10.1016/j.celrep.2013.07.046.
Durdevic, Z. and Schaefer, M. (2013) ‘tRNA modifications: necessary for correct tRNA-derived fragments during the recovery from stress?’, *BioEssays : news and reviews in molecular, cellular and developmental biology*, 35(4), pp. 323–7. doi: 10.1002/bies.201200158.
Eckersley-Maslin, M. A. *et al.* (2016) ‘MERVL/Zscan4 Network Activation Results in Transient Genome-wide DNA Demethylation of mESCs.’, *Cell reports*. Elsevier, 17(1), pp. 179–192. doi: 10.1016/j.celrep.2016.08.087.
Eckersley-Maslin, M. A. *et al.* (2018) ‘Dppa2 and Dppa4 directly regulate the Dux driven zygotic transcriptional programme’, *bioRxiv*. Cold Spring Harbor Laboratory, p. 431890. doi: 10.1101/431890.
Elbarbary, R. a. *et al.* (2009) ‘Modulation of gene expression by human cytosolic tRNase ZL through 5???-half-tRNA’, *PLoS ONE*, 4(6). doi: 10.1371/journal.pone.0005908.
Elsässer, S. J. *et al.* (2015) ‘Histone H3.3 is required for endogenous retroviral element silencing in embryonic stem cells’, *Nature*, 522(7555), pp. 240–4. doi: 10.1038/nature14345.
Emara, M. M. *et al.* (2010) ‘Angiogenin-induced tRNA-derived stress-induced RNAs promote stress-induced stress granule assembly.’, *The Journal of biological chemistry*, 285(14), pp. 10959–68. doi: 10.1074/jbc.M109.077560.
Erhard, K. F. *et al.* (2009) ‘RNA polymerase IV functions in paramutation in Zea

mays.', *Science (New York, N.Y.)*. American Association for the Advancement of Science, 323(5918), pp. 1201–5. doi: 10.1126/science.1164508.
Erkek, S. *et al.* (2013) 'Molecular determinants of nucleosome retention at CpG-rich sequences in mouse spermatozoa', *Nature Structural & Molecular Biology*, 20(7), pp. 868–875. doi: 10.1038/nsmb.2599.
Ernst, C. *et al.* (2016) 'Successful transmission and transcriptional deployment of a human chromosome via mouse male meiosis', *eLife*, 5. doi: 10.7554/eLife.20235.
Ernst, C., Odom, D. T. and Kutter, C. (2017) 'The emergence of piRNAs against transposon invasion to preserve mammalian genome integrity', *Nature Communications*. Nature Publishing Group, 8(1), p. 1411. doi: 10.1038/s41467-017-01049-7.
Evdokimova, V. *et al.* (2001) 'The major mRNA-associated protein YB-1 is a potent 5' cap-dependent mRNA stabilizer', *The EMBO Journal*, 20(19), pp. 5491–5502. doi: 10.1093/emboj/20.19.5491.
Fazio, S. De *et al.* (2011) 'The endonuclease activity of Mili fuels piRNA amplification that silences LINE1 elements', *Nature*, 480, pp. 264–267. Available at: https://www.nature.com/articles/nature10547 (Accessed: 16 May 2019).
Feil, R. and Fraga, M. F. (2012) 'Epigenetics and the environment: emerging patterns and implications', *Nature Reviews Genetics*. Nature Publishing Group, 13(2), pp. 97–109. doi: 10.1038/nrg3142.
Fire, A. *et al.* (1998) 'Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans', *Nature*, 391(6669), pp. 806–811. doi: 10.1038/35888.
Fleming, T. P. *et al.* (2012) 'Adaptive responses of the embryo to maternal diet and consequences for post-implantation development', *Reproduction, Fertility and Development*. CSIRO PUBLISHING, 24(1), p. 35. doi: 10.1071/RD11905.
Flemr, M. *et al.* (2013) 'A Retrotransposon-Driven Dicer Isoform Directs Endogenous Small Interfering RNA Production in Mouse Oocytes', *Cell*, 155(4), pp. 807–816. doi: 10.1016/j.cell.2013.10.001.
Fowden, A. L. *et al.* (2008) 'The Placenta and Intrauterine Programming', *Journal of Neuroendocrinology*, 20(4), pp. 439–450. doi: 10.1111/j.1365-2826.2008.01663.x.
Fowler, E. K. *et al.* (2018) 'Small RNA populations revealed by blocking rRNA fragments in Drosophila melanogaster reproductive tissues', *PLOS ONE*. Edited by S. Semsey. Public Library of Science, 13(2), p. e0191966. doi: 10.1371/journal.pone.0191966.
Franke, V. *et al.* (2017) 'Long terminal repeats power evolution of genes and gene expression programs in mammalian oocytes and zygotes.', *Genome research*. Cold Spring Harbor Laboratory Press, 27(8), pp. 1384–1394. doi: 10.1101/gr.216150.116.
Frantz, E. D. C. *et al.* (2011) 'Transgenerational endocrine pancreatic adaptation in mice from maternal protein restriction in utero', *Mechanisms of Ageing and Development*. Elsevier, 132(3), pp. 110–116. doi: 10.1016/J.MAD.2011.01.003.
Freimer, J. W. *et al.* (2018) 'Expression of Alternative Ago2 Isoform Associated with Loss of microRNA-Driven Translational Repression in Mouse Oocytes',

Current Biology, 28(2), pp. 296-302.e3. doi: 10.1016/j.cub.2017.11.067.
Frey, M. R. *et al.* (1995) 'Coiled bodies contain U7 small nuclear RNA and associate with specific DNA sequences in interphase human cells.', *Proceedings of the National Academy of Sciences*. National Academy of Sciences, 92(13), pp. 5915–5919. doi: 10.1073/pnas.92.13.5915.
Fricker, R. *et al.* (2019) 'A tRNA half modulates translation as stress response in Trypanosoma brucei', *Nature Communications*. Nature Publishing Group, 10(1), p. 118. doi: 10.1038/s41467-018-07949-6.
Fu, H. *et al.* (2009a) 'Stress induces tRNA cleavage by angiogenin in mammalian cells.', *FEBS letters*, 583(2), pp. 437–42. doi: 10.1016/j.febslet.2008.12.043.
Fu, H. *et al.* (2009b) 'Stress induces tRNA cleavage by angiogenin in mammalian cells', *FEBS Letters*, 583(2), pp. 437–442. doi: 10.1016/j.febslet.2008.12.043.
Fullston, T. *et al.* (2012) 'Diet-induced paternal obesity in the absence of diabetes diminishes the reproductive health of two subsequent generations of mice', *Human Reproduction*. Narnia, 27(5), pp. 1391–1400. doi: 10.1093/humrep/des030.
Gall, J. G. (2000) 'Cajal Bodies: The First 100 Years', *Annual Review of Cell and Developmental Biology*. Annual Reviews 4139 El Camino Way, P.O. Box 10139, Palo Alto, CA 94303-0139, USA , 16(1), pp. 273–300. doi: 10.1146/annurev.cellbio.16.1.273.
Galloway, L. F. (2005) 'Maternal effects provide phenotypic adaptation to local environmental conditions', *New Phytologist*. John Wiley & Sons, Ltd (10.1111), 166(1), pp. 93–100. doi: 10.1111/j.1469-8137.2004.01314.x.
Gapp, K. *et al.* (2014a) 'Implication of sperm RNAs in transgenerational inheritance of the effects of early trauma in mice.', *Nature neuroscience*. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved., 17(5), pp. 667–9. doi: 10.1038/nn.3695.
Gapp, K. *et al.* (2014b) 'Implication of sperm RNAs in transgenerational inheritance of the effects of early trauma in mice.', *Nature neuroscience*, 17(5), pp. 667–9. doi: 10.1038/nn.3695.
Gapp, K. *et al.* (2018) 'Alterations in sperm long RNA contribute to the epigenetic inheritance of the effects of postnatal trauma', *Molecular Psychiatry*. Nature Publishing Group, p. 1. doi: 10.1038/s41380-018-0271-6.
Garcia-Perez, J., Widmann, T. and Adams, I. (2016) 'The impact of transposable elements on mammalian development.', *Development (Cambridge, England)*. Oxford University Press for The Company of Biologists Limited, 143(22), pp. 4101–4114. doi: 10.1242/dev.132639.
Gebetsberger, J. *et al.* (2012a) 'tRNA-derived fragments target the ribosome and function as regulatory non-coding RNA in Haloferax volcanii.', *Archaea (Vancouver, B.C.)*, 2012, p. 260909. doi: 10.1155/2012/260909.
Gebetsberger, J. *et al.* (2012b) 'tRNA-Derived Fragments Target the Ribosome and Function as Regulatory Non-Coding RNA in *Haloferax volcanii*', *Archaea*, 2012, pp. 1–11. doi: 10.1155/2012/260909.
Gebetsberger, J. *et al.* (2017a) 'A tRNA-derived fragment competes with mRNA for ribosome binding and regulates translation during stress', *RNA Biology*, 14(10), pp. 1364–1373. doi: 10.1080/15476286.2016.1257470.

Gebetsberger, J. *et al.* (2017b) ‘A tRNA-derived fragment competes with mRNA for ribosome binding and regulates translation during stress’, *RNA Biology*. Taylor & Francis, 14(10), pp. 1364–1373. doi: 10.1080/15476286.2016.1257470.
Genenncher, B. *et al.* (2018) ‘Mutations in Cytosine-5 tRNA Methyltransferases Impact Mobile Element Expression and Genome Stability at Specific DNA Repeats In Brief’. doi: 10.1016/j.celrep.2018.01.061.
German, M. A. *et al.* (2009) ‘Construction of Parallel Analysis of RNA Ends (PARE) libraries for the study of cleaved miRNA targets and the RNA degradome’, *Nature Protocols*. Nature Publishing Group, 4(3), pp. 356–362. doi: 10.1038/nprot.2009.8.
Ghildiyal, M. and Zamore, P. D. (2009) ‘Small silencing RNAs: an expanding universe’, *Nature Reviews Genetics*. Nature Publishing Group, 10(2), pp. 94–108. doi: 10.1038/nrg2504.
Ghule, P.N. *et al.* (2008) ‘Staged assembly of histone gene expression machinery at subnuclear foci in the abbreviated cell cycle of human embryonic stem cells’, *Proceedings of the National Academy of Sciences*, 105(44), pp. 16964–16969. doi: 10.1073/pnas.0809273105.
Ghule, Prachi N *et al.* (2008) *Staged assembly of histone gene expression machinery at subnuclear foci in the abbreviated cell cycle of human embryonic stem cells*. Available at: www.pnas.org/cgi/content/full/ (Accessed: 15 April 2019).
Gick, O. *et al.* (1987) *Heat-labile regulatory factor is required for 3’ processing of histone precursor mRNAs (posttranscriptional regulation/cell proliferation/anti-Sm antibodies/U7 RNA)*, *Proc. Nati. Acad. Sci. USA*. Available at: https://www.pnas.org/content/pnas/84/24/8937.full.pdf (Accessed: 16 April 2019).
Girard, A. *et al.* (2006) ‘A germline-specific class of small RNAs binds mammalian Piwi proteins’, *Nature*. Nature Publishing Group, 442(7099), pp. 199–202. doi: 10.1038/nature04917.
Goh, W. S. S. *et al.* (2015) ‘piRNA-directed cleavage of meiotic transcripts regulates spermatogenesis’, *Genes & Development*, 29(10), pp. 1032–1044. doi: 10.1101/gad.260455.115.
Göke, J. *et al.* (2015) ‘Dynamic Transcription of Distinct Classes of Endogenous Retroviral Elements Marks Specific Populations of Early Human Embryonic Cells’, *Cell Stem Cell*, 16(2), pp. 135–141. doi: 10.1016/j.stem.2015.01.005.
Gokhman, D. *et al.* (2013) ‘Multilayered chromatin analysis reveals E2f, Smad and Zfx as transcriptional regulators of histones’, *Nature Structural & Molecular Biology*. Nature Publishing Group, 20(1), pp. 119–126. doi: 10.1038/nsmb.2448.
Goncalves, K. A. *et al.* (2016) ‘Angiogenin Promotes Hematopoietic Regeneration by Dichotomously Regulating Quiescence of Stem and Progenitor Cells’, *Cell*. Elsevier, 166(4), pp. 894–906. doi: 10.1016/j.cell.2016.06.042.
Goodarzi, H., Liu, X., Nguyen, Hoang C B, *et al.* (2015) ‘Endogenous tRNA-Derived Fragments Suppress Breast Cancer Progression via YBX1 Displacement.’, *Cell*. Elsevier, 161(4), pp. 790–802. doi: 10.1016/j.cell.2015.02.053.
Goodarzi, H., Liu, X., Nguyen, Hoang C.B., *et al.* (2015a) ‘Endogenous tRNA-Derived Fragments Suppress Breast Cancer Progression via YBX1

Displacement', *Cell*, 161(4), pp. 790–802. doi: 10.1016/j.cell.2015.02.053.
Goodarzi, H., Liu, X., Nguyen, Hoang C.B., *et al.* (2015b) 'Endogenous tRNA-Derived Fragments Suppress Breast Cancer Progression via YBX1 Displacement', *Cell*. Elsevier, 161(4), pp. 790–802. doi: 10.1016/j.cell.2015.02.053.
Gou, L.-T. *et al.* (2014) 'Pachytene piRNAs instruct massive mRNA elimination during late spermiogenesis', *Cell Research*, 24(6), pp. 680–700. doi: 10.1038/cr.2014.41.
Gowen, J. W. and Gay, E. H. (1933) 'EFFECT OF TEMPERATURE ON EVERSPORTING EYE COLOR IN DROSOPHILA MELANOGASTER.', *Science (New York, N.Y.)*. American Association for the Advancement of Science, 77(1995), p. 312. doi: 10.1126/science.77.1995.312.
Grandjean, V. *et al.* (2016) 'RNA-mediated paternal heredity of diet-induced obesity and metabolic disorders', *Scientific Reports*. Nature Publishing Group, 5(1), p. 18193. doi: 10.1038/srep18193.
Greve, T. S., Judson, R. L. and Blelloch, R. (2013) 'microRNA control of mouse and human pluripotent stem cell behavior.', *Annual review of cell and developmental biology*. NIH Public Access, 29, pp. 213–239. doi: 10.1146/annurev-cellbio-101512-122343.
Grivna, S. T. *et al.* (2006) 'A novel class of small RNAs in mouse spermatogenic cells.', *Genes & development*. Cold Spring Harbor Laboratory Press, 20(13), pp. 1709–14. doi: 10.1101/gad.1434406.
Gruber, J. J. *et al.* (2012) 'Ars2 promotes proper replication-dependent histone mRNA 3' end formation.', *Molecular cell*. Elsevier, 45(1), pp. 87–98. doi: 10.1016/j.molcel.2011.12.020.
Gu, W. *et al.* (2009) 'Distinct Argonaute-Mediated 22G-RNA Pathways Direct Genome Surveillance in the C. elegans Germline', *Molecular Cell*. Cell Press, 36(2), pp. 231–244. doi: 10.1016/J.MOLCEL.2009.09.020.
Gu, W. *et al.* (2012) 'CapSeq and CIP-TAP Identify Pol II Start Sites and Reveal Capped Small RNAs as C. elegans piRNA Precursors', *Cell*. Cell Press, 151(7), pp. 1488–1500. doi: 10.1016/J.CELL.2012.11.023.
Guallar, D. *et al.* (2012) 'Expression of endogenous retroviruses is negatively regulated by the pluripotency marker Rex1/Zfp42', *Nucleic Acids Research*, 40(18), pp. 8993–9007. doi: 10.1093/nar/gks686.
Guo, H. *et al.* (2014) 'The DNA methylation landscape of human early embryos', *nature*, 511, pp. 606–10. Available at: https://www.nature.com/articles/nature13544 (Accessed: 17 May 2019).
Guo, J. U. and Bartel, D. P. (2016) 'RNA G-quadruplexes are globally unfolded in eukaryotic cells and depleted in bacteria.', *Science (New York, N.Y.)*. NIH Public Access, 353(6306). doi: 10.1126/science.aaf5371.
Guo, L. *et al.* (2017) 'Sperm-carried RNAs play critical roles in mouse embryonic development.', *Oncotarget*. Impact Journals, LLC, 8(40), pp. 67394–67405. doi: 10.18632/oncotarget.18672.
Guzzi, N. *et al.* (2018) 'Pseudouridylation of tRNA-Derived Fragments Steers Translational Control in Stem Cells', *Cell*. Cell Press, 173(5), pp. 1204-1216.e26. doi: 10.1016/J.CELL.2018.03.008.

Haig, D. (2004) ‘Genomic Imprinting and Kinship: How Good is the Evidence?’, *Annual Review of Genetics*, 38(1), pp. 553–585. doi: 10.1146/annurev.genet.37.110801.142741.
Hale, C. J. *et al.* (2009) ‘Production and Processing of siRNA Precursor Transcripts from the Highly Repetitive Maize Genome’, *PLoS Genetics*. Edited by G. P. Copenhaver. Public Library of Science, 5(8), p. e1000598. doi: 10.1371/journal.pgen.1000598.
Hales, C. N. and Barker, D. J. (1992) ‘Type 2 (non-insulin-dependent) diabetes mellitus: the thrifty phenotype hypothesis.’, *Diabetologia*, 35(7), pp. 595–601. Available at: http://www.ncbi.nlm.nih.gov/pubmed/1644236 (Accessed: 11 April 2019).
Hales, C. N. and Barker, D. J. P. (2001) ‘The thrifty phenotype hypothesis’, *British Medical Bulletin*, 60(1), pp. 5–20. doi: 10.1093/bmb/60.1.5.
Hammoud, S. S., Nix, David A, *et al.* (2009) ‘Distinctive chromatin in human sperm packages genes for embryo development.’, *Nature*, 460(7254), pp. 473–8. doi: 10.1038/nature08162.
Hammoud, S. S., Nix, David A., *et al.* (2009) ‘Distinctive chromatin in human sperm packages genes for embryo development’, *Nature*, 460(7254), pp. 473–478. doi: 10.1038/nature08162.
Han, B. W. *et al.* (2015) ‘piRNA-guided transposon cleavage initiates Zucchini-dependent, phased piRNA production’, *Science*, 348(6236), pp. 817–821. doi: 10.1126/science.aaa1264.
Han, L. *et al.* (2019) ‘Differential roles of Stella in the modulation of DNA methylation during oocyte and zygotic development’, *Cell Discovery*. Nature Publishing Group, 5(1), p. 9. doi: 10.1038/s41421-019-0081-2.
Hanada, T. *et al.* (2013) ‘CLP1 links tRNA metabolism to progressive motor-neuron loss.’, *Nature*. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved., 495(7442), pp. 474–80. doi: 10.1038/nature11923.
Hartmann-Goldstein, I. J. (1967) ‘On the relationship between heterochromatization and variegation in Drosophila, with special reference to temperature sensitive periods’, *Genetical Research*. Cambridge University Press, 10(02), p. 143. doi: 10.1017/S0016672300010880.
Harvey, Z. H., Chen, Y. and Jarosz, D. F. (2018) ‘Protein-Based Inheritance: Epigenetics beyond the Chromosome.’, *Molecular cell*. Elsevier, 69(2), pp. 195–202. doi: 10.1016/j.molcel.2017.10.030.
Haussecker, D. *et al.* (2010) ‘Human tRNA-derived small RNAs in the global regulation of RNA silencing.’, *RNA (New York, N.Y.)*, 16(4), pp. 673–695. doi: 10.1261/rna.2000810.
Havens, M. A. and Hastings, M. L. (2016) ‘Splice-switching antisense oligonucleotides as therapeutic drugs.’, *Nucleic acids research*. Oxford University Press, 44(14), pp. 6549–63. doi: 10.1093/nar/gkw533.
He, J. *et al.* (2019) ‘Transposable elements are regulated by context-specific patterns of chromatin marks in mouse embryonic stem cells’, *Nature Communications*. Nature Publishing Group, 10(1), p. 34. doi: 10.1038/s41467-018-08006-y.

Heard, E. and Martienssen, R (2014) ‘Leading Edge Review Transgenerational Epigenetic Inheritance: Myths and Mechanisms’, *Cell*, 157, pp. 95–109. doi: 10.1016/j.cell.2014.02.045.
Heard, E. and Martienssen, RA (2014) ‘Transgenerational epigenetic inheritance: Myths and mechanisms’, *Cell*. Elsevier Inc., 157(1), pp. 95–109. doi: 10.1016/j.cell.2014.02.045.
Helwak, A. *et al.* (2013) ‘Mapping the Human miRNA Interactome by CLASH Reveals Frequent Noncanonical Binding’, *Cell*, 153(3), pp. 654–665. doi: 10.1016/j.cell.2013.03.043.
Hendrickson, P. G. *et al.* (2017) ‘Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons’, *Nature Genetics*. Nature Publishing Group, 49(6), pp. 925–934. doi: 10.1038/ng.3844.
Herbst, A. L., Ulfelder, H. and Poskanzer, D. C. (1971) ‘Adenocarcinoma of the Vagina’, *New England Journal of Medicine*. Massachusetts Medical Society , 284(16), pp. 878–881. doi: 10.1056/NEJM197104222841604.
Hirasawa, R. *et al.* (2008) ‘Maternal and zygotic Dnmt1 are necessary and sufficient for the maintenance of DNA methylation imprints during preimplantation development.’, *Genes & development*. Cold Spring Harbor Laboratory Press, 22(12), pp. 1607–16. doi: 10.1101/gad.1667008.
Hochberg, Z. *et al.* (2011) ‘Child Health, Developmental Plasticity, and Epigenetic Programming’, *Endocrine Reviews*, 32(2), pp. 159–224. doi: 10.1210/er.2009-0039.
Hoefig, K. P. *et al.* (2013) ‘Eri1 degrades the stem-loop of oligouridylated histone mRNAs to induce replication-dependent decay’, *Nature Structural & Molecular Biology*. Nature Publishing Group, 20(1), pp. 73–81. doi: 10.1038/nsmb.2450.
Hogan, B. *et al.* (no date) *Manipulating the Mouse Embryo: A Laboratory Manual (2nd edn)*. Available at: https://www.cell.com/immunology/pdf/0167-5699(95)80059-X.pdf (Accessed: 5 April 2019).
Hollick, J. B. (2017) ‘Paramutation and related phenomena in diverse species’, *Nature Reviews Genetics*. Nature Publishing Group, 18(1), pp. 5–23. doi: 10.1038/nrg.2016.115.
Hollick, J. B. and Chandler, V. L. (2001) ‘Genetic factors required to maintain repression of a paramutagenic maize pl1 allele.’, *Genetics*, 157(1), pp. 369–78. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11139517 (Accessed: 19 March 2019).
Holoch, D. and Moazed, D. (2015) ‘Small-RNA loading licenses Argonaute for assembly into a transcriptional silencing complex.’, *Nature structural & molecular biology*, 22(4), pp. 328–35. doi: 10.1038/nsmb.2979.
Homolka, D. *et al.* (2015) ‘PIWI Slicing and RNA Elements in Precursors Instruct Directional Primary piRNA Biogenesis’, *Cell Reports*, 12(3), pp. 418–428. doi: 10.1016/j.celrep.2015.06.030.
Honda, S. *et al.* (2015) ‘Sex hormone-dependent tRNA halves enhance cell proliferation in breast and prostate cancers’, *Proceedings of the National Academy of Sciences*, 112(29), pp. E3816–E3825. doi: 10.1073/pnas.1510077112.

Horsthemke, B. (2018) ‘A critical view on transgenerational epigenetic inheritance in humans’, *Nature Communications*. Nature Publishing Group, 9(1), p. 2973. doi: 10.1038/s41467-018-05445-5.
Horváth, V., Merenciano, M. and González, J. (2017) ‘Revisiting the Relationship between Transposable Elements and the Eukaryotic Stress Response’. doi: 10.1016/j.tig.2017.08.007.
Hövel, I., Pearson, N. A. and Stam, M. (2015) ‘Cis-acting determinants of paramutation’, *Seminars in Cell & Developmental Biology*. Academic Press, 44, pp. 22–32. doi: 10.1016/J.SEMCDB.2015.08.012.
Howell, C. Y. *et al.* (2001) ‘Genomic imprinting disrupted by a maternal effect mutation in the Dnmt1 gene.’, *Cell*, 104(6), pp. 829–38. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11290321 (Accessed: 17 May 2019).
Huang, Y. *et al.* (2017) ‘Stella modulates transcriptional and endogenous retrovirus programs during maternal-to-zygotic transition.’, *eLife*. eLife Sciences Publications, Ltd, 6. doi: 10.7554/eLife.22345.
Hussain, S. *et al.* (2013) ‘NSun2-Mediated Cytosine-5 Methylation of Vault Noncoding RNA Determines Its Processing into Regulatory Small RNAs’, *Cell Reports*. Cell Press, 4(2), pp. 255–261. doi: 10.1016/J.CELREP.2013.06.029.
Huypens, P. *et al.* (2016) ‘Epigenetic germline inheritance of diet-induced obesity and insulin resistance’, *Nature Genetics*. Nature Publishing Group, 48(5), pp. 497–499. doi: 10.1038/ng.3527.
De Iaco, A. *et al.* (2017) ‘DUX-family transcription factors regulate zygotic genome activation in placental mammals’, *Nature Genetics*. Nature Publishing Group, 49(6), pp. 941–945. doi: 10.1038/ng.3858.
Iaco, A. De *et al.* (2018) ‘DPPA2 and DPPA4 are necessary to establish a totipotent state in mouse embryonic stem cells’, *bioRxiv*. Cold Spring Harbor Laboratory, p. 447755. doi: 10.1101/447755.
Iaco, A. De *et al.* (2019) ‘DUX is a non-essential synchronizer of zygotic genome activation’, *bioRxiv*. Cold Spring Harbor Laboratory, p. 569434. doi: 10.1101/569434.
Imbeault, M., Helleboid, P.-Y. and Trono, D. (2017) ‘KRAB zinc-finger proteins contribute to the evolution of gene regulatory networks’, *Nature*, 543(7646), pp. 550–554. doi: 10.1038/nature21683.
Ingolia, N. T., Lareau, L. F. and Weissman, J. S. (2011) ‘Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes’, *Cell*. Elsevier Inc., 147(4), pp. 789–802. doi: 10.1016/j.cell.2011.10.002.
Ipsaro, J. J. *et al.* (2012) ‘The structural biochemistry of Zucchini implicates it as a nuclease in piRNA biogenesis’, *Nature*, 491(7423), pp. 279–283. doi: 10.1038/nature11502.
Irani, M. *et al.* (2018) ‘Blastocyst development rate influences implantation and live birth rates of similarly graded euploid blastocysts’, *Fertility and Sterility*, 110(1), pp. 95-102.e1. doi: 10.1016/j.fertnstert.2018.03.032.
Ishiuchi, T. *et al.* (2015a) ‘Early embryonic-like cells are induced by downregulating replication-dependent chromatin assembly.’, *Nature structural & molecular biology*. Nature Publishing Group, a division of Macmillan Publishers

Limited. All Rights Reserved., 22(9), pp. 662–671. doi: 10.1038/nsmb.3066.
Ishiuchi, T. *et al.* (2015b) 'Early embryonic-like cells are induced by downregulating replication-dependent chromatin assembly', *Nature Structural & Molecular Biology*. Nature Publishing Group, 22(9), pp. 662–671. doi: 10.1038/nsmb.3066.
Ito, H. *et al.* (2011) 'An siRNA pathway prevents transgenerational retrotransposition in plants subjected to stress', *Nature*. Nature Publishing Group, 472(7341), pp. 115–119. doi: 10.1038/nature09861.
Itou, D. *et al.* (2015) 'Induction of DNA Methylation by Artificial piRNA Production in Male Germ Cells', *Current Biology*. Cell Press, 25(7), pp. 901–906. doi: 10.1016/J.CUB.2015.01.060.
Ivanov, P., Emara, Mohamed M., *et al.* (2011) 'Angiogenin-Induced tRNA Fragments Inhibit Translation Initiation', *Molecular Cell*. Elsevier Inc., 43(4), pp. 613–623. doi: 10.1016/j.molcel.2011.06.022.
Ivanov, P., Emara, Mohamed M., *et al.* (2011a) 'Angiogenin-Induced tRNA Fragments Inhibit Translation Initiation', *Molecular Cell*. Cell Press, 43(4), pp. 613–623. doi: 10.1016/J.MOLCEL.2011.06.022.
Ivanov, P., Emara, Mohamed M., *et al.* (2011b) 'Angiogenin-Induced tRNA Fragments Inhibit Translation Initiation', *Molecular Cell*, 43(4), pp. 613–623. doi: 10.1016/j.molcel.2011.06.022.
Ivanov, P. *et al.* (2014) 'G-quadruplex structures contribute to the neuroprotective effects of angiogenin-induced tRNA fragments.', *Proceedings of the National Academy of Sciences of the United States of America*, 111(51), pp. 18201–6. doi: 10.1073/pnas.1407361111.
Jachowicz, J. W. *et al.* (2017) 'LINE-1 activation after fertilization regulates global chromatin accessibility in the early mouse embryo', *Nature Genetics*. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved., 49, p. 1502. Available at: http://dx.doi.org/10.1038/ng.3945.
Jacobs, F. M. J. *et al.* (2014) 'An evolutionary arms race between KRAB zinc-finger genes ZNF91/93 and SVA/L1 retrotransposons', *Nature*. Nature Publishing Group, 516(7530), pp. 242–245. doi: 10.1038/nature13760.
Jady, B. E. *et al.* (2003) 'Modification of Sm small nuclear RNAs occurs in the nucleoplasmic Cajal body following import from the cytoplasm', *The EMBO Journal*, 22(8), pp. 1878–1888. doi: 10.1093/emboj/cdg187.
Jády, B. E. *et al.* (2003) 'Modification of Sm small nuclear RNAs occurs in the nucleoplasmic Cajal body following import from the cytoplasm.', *The EMBO journal*. European Molecular Biology Organization, 22(8), pp. 1878–88. doi: 10.1093/emboj/cdg187.
Jády, B. E., Bertrand, E. and Kiss, T. (2004) 'Human telomerase RNA and box H/ACA scaRNAs share a common Cajal body–specific localization signal', *The Journal of Cell Biology*, 164(5), pp. 647–652. doi: 10.1083/jcb.200310138.
Jevtić, P. and Levy, D. L. (2017) 'Both Nuclear Size and DNA Amount Contribute to Midblastula Transition Timing in Xenopus laevis', *Scientific Reports*. Nature Publishing Group, 7(1), p. 7908. doi: 10.1038/s41598-017-08243-z.
Jimenez-Chillaron, J. C. *et al.* (2009) 'Intergenerational transmission of glucose intolerance and obesity by in utero undernutrition in mice.', *Diabetes*, 58(2), pp.

460–8. doi: 10.2337/db08-0490.
Jimeno-González, S. *et al.* (2015) ‘Defective histone supply causes changes in RNA polymerase II elongation rate and cotranscriptional pre-mRNA splicing.’, *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 112(48), pp. 14840–5. doi: 10.1073/pnas.1506760112.
Jones, P. A. and Takai, D. (2001) ‘The role of DNA methylation in mammalian epigenetics.’, *Science (New York, N.Y.)*. American Association for the Advancement of Science, 293(5532), pp. 1068–70. doi: 10.1126/science.1063852.
Joseph, S. R. *et al.* (2017a) ‘Competition between histone and transcription factor binding regulates the onset of transcription in zebrafish embryos’, *eLife*, 6. doi: 10.7554/eLife.23326.
Joseph, S. R. *et al.* (2017b) ‘Competition between histone and transcription factor binding regulates the onset of transcription in zebrafish embryos’, *eLife*, 6. doi: 10.7554/eLife.23326.
Kaati, G., Bygren, L. and Edvinsson, S. (2002) ‘Cardiovascular and diabetes mortality determined by nutrition during parents’ and grandparents’ slow growth period’, *European Journal of Human Genetics*, 10(11), pp. 682–688. doi: 10.1038/sj.ejhg.5200859.
Kabayama, Y. *et al.* (2017) ‘Roles of MIWI, MILI and PLD6 in small RNA regulation in mouse growing oocytes’, *Nucleic Acids Research*, 45(9), p. gkx027. doi: 10.1093/nar/gkx027.
Kapitonov, V. V. and Jurka, J. (2008) ‘A universal classification of eukaryotic transposable elements implemented in Repbase’, *Nature Reviews Genetics*. Nature Publishing Group, 9(5), pp. 411–412. doi: 10.1038/nrg2165-c1.
Karaca, E. *et al.* (2014) ‘Human CLP1 mutations alter tRNA biogenesis, affecting both peripheral and central nervous system function.’, *Cell*, 157(3), pp. 636–50. doi: 10.1016/j.cell.2014.02.058.
Kato, Y. *et al.* (2007) ‘Role of the Dnmt3 family in de novo methylation of imprinted and repetitive sequences during male germ cell development in the mouse’, *Human Molecular Genetics*, 16(19), pp. 2272–2280. doi: 10.1093/hmg/ddm179.
Kawaji, H. *et al.* (2008) ‘Hidden layers of human small RNAs.’, *BMC genomics*, 9(1), p. 157. doi: 10.1186/1471-2164-9-157.
Kawaoka, S. *et al.* (2011) ‘3′ End Formation of PIWI-Interacting RNAs In Vitro’, *Molecular Cell*, 43(6), pp. 1015–1022. doi: 10.1016/j.molcel.2011.07.029.
Kazazian, H. H. (2004) ‘Mobile Elements: Drivers of Genome Evolution’, *Science*, 303(5664), pp. 1626–1632. doi: 10.1126/science.1089670.
Keam, S. and Hutvagner, G. (2015) ‘tRNA-Derived Fragments (tRFs): Emerging New Roles for an Ancient RNA in the Regulation of Gene Expression’, *Life*, 5(4), pp. 1638–1651. doi: 10.3390/life5041638.
Kermicle, J. L., Eggleston, W. B. and Alleman, M. (1995) ‘Organization of paramutagenicity in R-stippled maize.’, *Genetics*, 141(1).
Kiani, J. *et al.* (2013) ‘RNA–Mediated Epigenetic Heredity Requires the Cytosine Methyltransferase Dnmt2’, *PLoS Genetics*. Edited by W. A. Bickmore. Public

Library of Science, 9(5), p. e1003498. doi: 10.1371/journal.pgen.1003498.
Kim, D.-H. *et al.* (2009) 'Vernalization: Winter and the Timing of Flowering in Plants', *Annual Review of Cell and Developmental Biology*. Annual Reviews , 25(1), pp. 277–299. doi: 10.1146/annurev.cellbio.042308.113411.
Kim, H. K. *et al.* (2017) 'A transfer-RNA-derived small RNA regulates ribosome biogenesis', *Nature*. Nature Publishing Group, 552(7683), pp. 57–62. doi: 10.1038/nature25005.
Kimmins, S. and Sassone-Corsi, P. (2005) 'Chromatin remodelling and epigenetic features of germ cells', *Nature*. Nature Publishing Group, 434(7033), pp. 583–589. doi: 10.1038/nature03368.
Kinoshita, T., Ikeda, Y. and Ishikawa, R. (2008) 'Genomic imprinting: A balance between antagonistic roles of parental chromosomes', *Seminars in Cell & Developmental Biology*. Academic Press, 19(6), pp. 574–579. doi: 10.1016/J.SEMCDB.2008.07.018.
Kiriyama, M. *et al.* (2009) 'Interaction of FLASH with arsenite resistance protein 2 is involved in cell cycle progression at S phase.', *Molecular and cellular biology*. American Society for Microbiology Journals, 29(17), pp. 4729–41. doi: 10.1128/MCB.00289-09.
Kiss, A. M. *et al.* (2002) 'A Cajal body-specific pseudouridylation guide RNA is composed of two box H/ACA snoRNA-like domains.', *Nucleic acids research*. Oxford University Press, 30(21), pp. 4643–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12409454 (Accessed: 15 April 2019).
Klosin, A. *et al.* (2017) 'Transgenerational transmission of environmental information in *C. elegans*', *Science*, 356(6335), pp. 320–323. doi: 10.1126/science.aah6412.
Knoll, J. H. M. *et al.* (1989) 'Angelman and Prader-Willi syndromes share a common chromosome 15 deletion but differ in parental origin of the deletion', *American Journal of Medical Genetics*. John Wiley & Sons, Ltd, 32(2), pp. 285–290. doi: 10.1002/ajmg.1320320235.
Köhn, M. *et al.* (2015) 'The Y3** ncRNA promotes the 3′ end processing of histone mRNAs', *Genes & Development*. Cold Spring Harbor Laboratory Press, 29(19), pp. 1998–2003. doi: 10.1101/GAD.266486.115.
Kojima-Kita, K. *et al.* (2016) 'MIWI2 as an Effector of DNA Methylation and Gene Silencing in Embryonic Male Germ Cells', *CellReports*, 16, pp. 2819–2828. doi: 10.1016/j.celrep.2016.08.027.
Kolev, N. G. and Steitz, J. A. (2005) 'Symplekin and multiple other polyadenylation factors participate in 3'-end maturation of histone mRNAs.', *Genes & development*. Cold Spring Harbor Laboratory Press, 19(21), pp. 2583–92. doi: 10.1101/gad.1371105.
Komiya, Y. and Habas, R. (2008) 'Wnt signal transduction pathways.', *Organogenesis*. Taylor & Francis, 4(2), pp. 68–75. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19279717 (Accessed: 11 January 2019).
Kondo, Y. *et al.* (2008) 'Downregulation of Histone H3 Lysine 9 Methyltransferase G9a Induces Centrosome Disruption and Chromosome Instability in Cancer Cells', *PLoS ONE*. Edited by A. Imhof. Public Library of Science, 3(4), p. e2037. doi: 10.1371/journal.pone.0002037.

Krapf, D. *et al.* (2012) ‘cSrc is necessary for epididymal development and is incorporated into sperm during epididymal transit’, *Developmental Biology*. Academic Press, 369(1), pp. 43–53. doi: 10.1016/J.YDBIO.2012.06.017.
Kruse, K. *et al.* (2019) ‘Transposable elements drive reorganisation of 3D chromatin during early embryogenesis’, *bioRxiv*. Cold Spring Harbor Laboratory, p. 523712. doi: 10.1101/523712.
Kumar, P. *et al.* (2014a) ‘Meta-analysis of tRNA derived RNA fragments reveals that they are evolutionarily conserved and associate with AGO proteins to recognize specific RNA targets’, *BMC Biology*, 12(1), p. 78. doi: 10.1186/s12915-014-0078-0.
Kumar, P. *et al.* (2014b) ‘Meta-analysis of tRNA derived RNA fragments reveals that they are evolutionarily conserved and associate with AGO proteins to recognize specific RNA targets’, *BMC Biology*, pp. 1–14. doi: 10.1186/s12915-014-0078-0.
Kumar, P., Kuscu, C. and Dutta, A. (2016) ‘Biogenesis and Function of Transfer RNA-Related Fragments (tRFs).’, *Trends in biochemical sciences*. NIH Public Access, 41(8), pp. 679–689. doi: 10.1016/j.tibs.2016.05.004.
Kuramochi-Miyagawa, S. *et al.* (2008) ‘DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes.’, *Genes & development*. Cold Spring Harbor Laboratory Press, 22(7), pp. 908–17. doi: 10.1101/gad.1640708.
Kurihara, Y. *et al.* (2008) ‘Maintenance of genomic methylation patterns during preimplantation development requires the somatic form of DNA methyltransferase 1’, *Developmental Biology*, 313(1), pp. 335–346. doi: 10.1016/j.ydbio.2007.10.033.
Kuscu, C. *et al.* (2018) ‘tRNA fragments (tRFs) guide Ago to regulate gene expression post-transcriptionally in a Dicer-independent manner.’, *RNA (New York, N.Y.)*. Cold Spring Harbor Laboratory Press, 24(8), pp. 1093–1105. doi: 10.1261/rna.066126.118.
Kwong, W. Y. *et al.* (2000) ‘Maternal undernutrition during the preimplantation period of rat development causes blastocyst abnormalities and programming of postnatal hypertension.’, *Development (Cambridge, England)*. Oxford University Press for The Company of Biologists Limited, 127(19), pp. 4195–202. doi: 10.1242/dev.103952.
Lacey, E. P. (1996) ‘PARENTAL EFFECTS IN *PLANTAGO LANCEOLATA* L. I.: A GROWTH CHAMBER EXPERIMENT TO EXAMINE PRE- AND POSTZYGOTIC TEMPERATURE EFFECTS’, *Evolution*. John Wiley & Sons, Ltd (10.1111), 50(2), pp. 865–878. doi: 10.1111/j.1558-5646.1996.tb03895.x.
Lachner, M. *et al.* (2001) ‘Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins.’, *Nature*. Macmillian Magazines Ltd., 410(6824), pp. 116–20. doi: 10.1038/35065132.
Lackey, P. E., Welch, J. D. and Marzluff, W. F. (2016) ‘TUT7 catalyzes the uridylation of the 3’ end for rapid degradation of histone mRNA.’, *RNA (New York, N.Y.)*. Cold Spring Harbor Laboratory Press, 22(11), pp. 1673–1688. doi: 10.1261/rna.058107.116.
Lam, M. K.-P. *et al.* (2000) ‘Second generation effects of maternal alcohol

consumption during pregnancy in rats', *Progress in Neuro-Psychopharmacology and Biological Psychiatry*. Elsevier, 24(4), pp. 619–631. doi: 10.1016/S0278-5846(00)00097-X.
Lane, M., Robker, R. L. and Robertson, S. A. (2014) 'Parenting from before conception.', *Science (New York, N.Y.)*. American Association for the Advancement of Science, 345(6198), pp. 756–60. doi: 10.1126/science.1254400.
Lane, N. *et al.* (2003) 'Resistance of IAPs to methylation reprogramming may provide a mechanism for epigenetic inheritance in the mouse', *genesis*. John Wiley & Sons, Ltd, 35(2), pp. 88–93. doi: 10.1002/gene.10168.
Lau, N. C. *et al.* (2006) 'Characterization of the piRNA Complex from Rat Testes', *Science*, 313(5785), pp. 363–367. doi: 10.1126/science.1130164.
Law, J. A. and Jacobsen, S. E. (2010) 'Establishing, maintaining and modifying DNA methylation patterns in plants and animals', *Nature Reviews Genetics*. Nature Publishing Group, 11(3), pp. 204–220. doi: 10.1038/nrg2719.
Lee, S. R. and Collins, K. (2005) 'Starvation-induced cleavage of the tRNA anticodon loop in Tetrahymena thermophila.', *The Journal of biological chemistry*. American Society for Biochemistry and Molecular Biology, 280(52), pp. 42744–9. doi: 10.1074/jbc.M510356200.
Lee, Y. S. *et al.* (2009) 'A novel class of small RNAs: tRNA-derived RNA fragments (tRFs).', *Genes & development*, 23(22), pp. 2639–49. doi: 10.1101/gad.1837609.
Leese, H. J. *et al.* (2008) 'Female reproductive tract fluids: composition, mechanism of formation and potential role in the developmental origins of health and disease', *Reproduction, Fertility and Development*. CSIRO PUBLISHING, 20(1), p. 1. doi: 10.1071/RD07153.
Lenstra, T. L. *et al.* (2011) 'The Specificity and Topology of Chromatin Interaction Pathways in Yeast', *Molecular Cell*. Cell Press, 42(4), pp. 536–549. doi: 10.1016/J.MOLCEL.2011.03.026.
Levens, E. D. *et al.* (2008) 'Blastocyst development rate impacts outcome in cryopreserved blastocyst transfer cycles', *Fertility and sterility*. NIH Public Access, 90(6), p. 2138. doi: 10.1016/J.FERTNSTERT.2007.10.029.
Li, E. and Zhang, Y. (2014) 'DNA methylation in mammals.', *Cold Spring Harbor perspectives in biology*. Cold Spring Harbor Laboratory Press, 6(5), p. a019133. doi: 10.1101/cshperspect.a019133.
Li, H., Freeling, M. and Lisch, D. (2010) 'Epigenetic reprogramming during vegetative phase change in maize.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 107(51), pp. 22184–9. doi: 10.1073/pnas.1016884108.
Li, J. *et al.* (2010) 'Systematic mapping and functional analysis of a family of human epididymal secretory sperm-located proteins.', *Molecular & cellular proteomics : MCP*. American Society for Biochemistry and Molecular Biology, 9(11), pp. 2517–28. doi: 10.1074/mcp.M110.001719.
Li, X. *et al.* (2008) 'A Maternal-Zygotic Effect Gene, Zfp57, Maintains Both Maternal and Paternal Imprints', *Developmental Cell*. Cell Press, 15(4), pp. 547–557. doi: 10.1016/J.DEVCEL.2008.08.014.
Li, X. Z. *et al.* (2013) 'An Ancient Transcription Factor Initiates the Burst of piRNA

Production during Early Meiosis in Mouse Testes', *Molecular Cell*. Cell Press, 50(1), pp. 67–81. doi: 10.1016/J.MOLCEL.2013.02.016.
Li, Y. *et al.* (2018) 'Stella safeguards the oocyte methylome by preventing de novo methylation mediated by DNMT1', *Nature*. Nature Publishing Group, 564(7734), pp. 136–140. doi: 10.1038/s41586-018-0751-5.
Li, Z. *et al.* (2012) 'Extensive terminal and asymmetric processing of small RNAs from rRNAs, snoRNAs, snRNAs, and tRNAs.', *Nucleic acids research*, 40(14), pp. 6787–99. doi: 10.1093/nar/gks307.
Liang, G. *et al.* (2002) 'Cooperativity between DNA methyltransferases in the maintenance methylation of repetitive elements.', *Molecular and cellular biology*, 22(2), pp. 480–91. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11756544 (Accessed: 8 May 2019).
Lim, S. L. *et al.* (2015) 'HENMT1 and piRNA Stability Are Required for Adult Male Germ Cell Transposon Repression and to Define the Spermatogenic Program in the Mouse', *PLOS Genetics*. Edited by M. Frye, 11(10), p. e1005620. doi: 10.1371/journal.pgen.1005620.
Lisch, D. (2012) 'Regulation of transposable elements in maize', *Current Opinion in Plant Biology*. Elsevier Current Trends, 15(5), pp. 511–516. doi: 10.1016/J.PBI.2012.07.001.
Liu, G. *et al.* (2018) 'Inherited DNA methylation primes the establishment of accessible chromatin during genome activation.', *Genome research*. Cold Spring Harbor Laboratory Press, 28(7), pp. 998–1007. doi: 10.1101/gr.228833.117.
Liu, J.-L. *et al.* (2006) 'The *Drosophila melanogaster* Cajal body', *The Journal of Cell Biology*, 172(6), pp. 875–884. doi: 10.1083/jcb.200511038.
Liu, J.-L. *et al.* (2009) 'Coilin Is Essential for Cajal Body Organization in *Drosophila melanogaster*', *Molecular Biology of the Cell*. Edited by M. Wickens, 20(6), pp. 1661–1670. doi: 10.1091/mbc.e08-05-0525.
Liu, S. *et al.* (2014) 'Setdb1 is required for germline development and silencing of H3K9me3-marked endogenous retroviruses in primordial germ cells.', *Genes & development*, 28(18), pp. 2041–55. doi: 10.1101/gad.244848.114.
Liu, X. *et al.* (2016) 'Distinct features of H3K4me3 and H3K27me3 chromatin domains in pre-implantation embryos', *Nature*. Nature Publishing Group, 537(7621), pp. 558–562. doi: 10.1038/nature19362.
Lorenz, C., Lünse, C. E. and Mörl, M. (2017) 'tRNA Modifications: Impact on Structure and Thermal Adaptation.', *Biomolecules*. Multidisciplinary Digital Publishing Institute (MDPI), 7(2). doi: 10.3390/biom7020035.
Lumey, L. *et al.* (2007) 'Cohort Profile: The Dutch Hunger Winter Families Study', *International Journal of Epidemiology*. Narnia, 36(6), pp. 1196–1204. doi: 10.1093/ije/dym126.
Luo, S. *et al.* (2018) 'Drosophila tsRNAs preferentially suppress general translation machinery via antisense pairing and participate in cellular starvation response', *Nucleic Acids Research*. Oxford University Press, 46(10), pp. 5250–5268. doi: 10.1093/nar/gky189.
Luteijn, M. J. and Ketting, R. F. (2013) 'PIWI-interacting RNAs: from generation to transgenerational epigenetics.', *Nature reviews. Genetics*. Nature Publishing Group, 14(8), pp. 523–34. doi: 10.1038/nrg3495.

Lyons, S. M., Fay, M. M. and Ivanov, P. (2018) ‘The role of RNA modifications in the regulation of tRNA cleavage’, *FEBS Letters*. John Wiley & Sons, Ltd, 592(17), pp. 2828–2844. doi: 10.1002/1873-3468.13205.
Ma, J. *et al.* (2010) ‘MicroRNA Activity Is Suppressed in Mouse Oocytes’, *Current Biology*. Cell Press, 20(3), pp. 265–270. doi: 10.1016/J.CUB.2009.12.042.
Ma, T. *et al.* (2000) ‘Cell cycle-regulated phosphorylation of p220(NPAT) by cyclin E/Cdk2 in Cajal bodies promotes histone gene transcription.’, *Genes & development*. Cold Spring Harbor Laboratory Press, 14(18), pp. 2298–313. doi: 10.1101/GAD.829500.
Ben Maamar, M. *et al.* (2018a) ‘Epigenetic Transgenerational Inheritance of Altered Sperm Histone Retention Sites’, *Scientific Reports*. Nature Publishing Group, 8(1), p. 5308. doi: 10.1038/s41598-018-23612-y.
Ben Maamar, M. *et al.* (2018b) ‘Epigenetic Transgenerational Inheritance of Altered Sperm Histone Retention Sites’, *Scientific Reports*. Nature Publishing Group, 8(1), p. 5308. doi: 10.1038/s41598-018-23612-y.
Macfarlan, T. S. *et al.* (2011) ‘Endogenous retroviruses and neighboring genes are coordinately repressed by LSD1/KDM1A’, *Genes and Development*, 25(6), pp. 594–607. doi: 10.1101/gad.2008511.
Macfarlan, T. S. *et al.* (2012a) ‘Embryonic stem cell potency fluctuates with endogenous retrovirus activity’, *Nature*. Nature Publishing Group, 487(7405), pp. 57–63. doi: 10.1038/nature11244.
Macfarlan, T. S. *et al.* (2012b) ‘Embryonic stem cell potency fluctuates with endogenous retrovirus activity’, *Nature*. Nature Publishing Group, 487(7405), pp. 57–63. doi: 10.1038/nature11244.
Machyna, M., Heyn, P. and Neugebauer, K. M. (2013) ‘Cajal bodies: where form meets function’, *Wiley Interdisciplinary Reviews: RNA*. John Wiley & Sons, Ltd, 4(1), pp. 17–34. doi: 10.1002/wrna.1139.
Mackay, D. J. G. *et al.* (2008) ‘Hypomethylation of multiple imprinted loci in individuals with transient neonatal diabetes is associated with mutations in ZFP57’, *Nature Genetics*, 40(8), pp. 949–951. doi: 10.1038/ng.187.
Madan, B. *et al.* (2009) ‘The pluripotency-associated gene Dppa4 is dispensable for embryonic stem cell identity and germ cell development but essential for embryogenesis.’, *Molecular and cellular biology*. American Society for Microbiology Journals, 29(11), pp. 3186–203. doi: 10.1128/MCB.01970-08.
Maksakova, I. a *et al.* (2013) ‘Distinct roles of KAP1, HP1 and G9a/GLP in silencing of the two-cell-specific retrotransposon MERVL in mouse ES cells.’, *Epigenetics & chromatin*. Epigenetics & Chromatin, 6(1), p. 15. doi: 10.1186/1756-8935-6-15.
Manakov, S. A. *et al.* (2015) ‘MIWI2 and MILI Have Differential Effects on piRNA Biogenesis and DNA Methylation’, *Cell Reports*, 12(8), pp. 1234–1243. doi: 10.1016/j.celrep.2015.07.036.
Manikkam, M. *et al.* (2013) ‘Plastics derived endocrine disruptors (BPA, DEHP and DBP) induce epigenetic transgenerational inheritance of obesity, reproductive disease and sperm epimutations.’, *PloS one*, 8(1), p. e55387. doi: 10.1371/journal.pone.0055387.
Marchand, V. *et al.* (2016) ‘Illumina-based RiboMethSeq approach for mapping

of 2′-O-Me residues in RNA', *Nucleic Acids Research*. Oxford University Press, 44(16), pp. e135–e135. doi: 10.1093/nar/gkw547.
Marquet, R. *et al.* (1995) 'tRNAs as primer of reverse transcriptases', *Biochimie*, 77(1–2), pp. 113–124. doi: 10.1016/0300-9084(96)88114-4.
Marti, A. *et al.* (1999) 'Interaction between ubiquitin–protein ligase SCFSKP2 and E2F-1 underlies the regulation of E2F-1 degradation', *Nature Cell Biology*. Nature Publishing Group, 1(1), pp. 14–19. doi: 10.1038/8984.
Martínez, D. *et al.* (2014) 'In utero undernutrition in male mice programs liver lipid metabolism in the second-generation offspring involving altered Lxra DNA methylation', *Cell Metabolism*, 19(6), pp. 941–951. doi: 10.1016/j.cmet.2014.03.026.
Martinez, G., Choudury, S. G. and Slotkin, R. K. (2017) 'tRNA-derived small RNAs target transposable element transcripts', *Nucleic Acids Research*. Oxford University Press, 45(9), pp. 5142–5152. doi: 10.1093/nar/gkx103.
Marzluff, W. F. and Koreski, K. P. (2017) 'Birth and Death of Histone mRNAs', *Trends in Genetics*. Elsevier Current Trends, 33(10), pp. 745–759. doi: 10.1016/J.TIG.2017.07.014.
Matera, A. G. and Frey, M. R. (1998) 'Coiled Bodies and Gems: Janus or Gemini?', *The American Journal of Human Genetics*, 63(2), pp. 317–321. doi: 10.1086/301992.
Matera, A. G. and Ward, D. C. (1993) 'Nucleoplasmic organization of small nuclear ribonucleoproteins in cultured human cells', *The Journal of Cell Biology*, 121(4), pp. 715–727. doi: 10.1083/jcb.121.4.715.
Mathieu, O. *et al.* (2007) 'Transgenerational Stability of the Arabidopsis Epigenome Is Coordinated by CG Methylation', *Cell*. Cell Press, 130(5), pp. 851–862. doi: 10.1016/J.CELL.2007.07.007.
Matsui, T. *et al.* (2010) 'Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET', *Nature*. Nature Publishing Group, 464(7290), pp. 927–931. doi: 10.1038/nature08858.
Matzke, M. A., Kanno, T. and Matzke, A. J. M. (2015) 'RNA-Directed DNA Methylation: The Evolution of a Complex Epigenetic Pathway in Flowering Plants', *Annual Review of Plant Biology*. Annual Reviews , 66(1), pp. 243–267. doi: 10.1146/annurev-arplant-043014-114633.
Maute, R. L. *et al.* (2013) 'tRNA-derived microRNA modulates proliferation and the DNA damage response and is down-regulated in B cell lymphoma.', *Proceedings of the National Academy of Sciences of the United States of America*, 110(4), pp. 1404–9. doi: 10.1073/pnas.1206761110.
Mayer, W. *et al.* (2000) 'Demethylation of the zygotic paternal genome.', *Nature*, 403(6769), pp. 501–2. doi: 10.1038/35000654.
McClintock, B. (1951) 'CHROMOSOME ORGANIZATION AND GENIC EXPRESSION', *Cold Spring Harbor Symposia on Quantitative Biology*, 16(0), pp. 13–47. doi: 10.1101/SQB.1951.016.01.004.
McClintock, B. (1961) 'Some Parallels Between Gene Control Systems in Maize and in Bacteria', *The American Naturalist*, 95(884), pp. 265–277. doi: 10.1086/282188.
McClintock, B. (1983) 'The significance of responses of the genome to

challenge'. Available at: https://books.google.com/books?hl=en&lr=&id=3ULyAgAAQBAJ&oi=fnd&pg=PA393&ots=j_e6DXfBug&sig=jFUyzUMmFVBA7GgdEz-xj6tmMUk (Accessed: 20 May 2019).
McGrath, J. and Solter, D. (1984) 'Completion of mouse embryogenesis requires both the maternal and paternal genomes', *Cell*. Cell Press, 37(1), pp. 179–183. doi: 10.1016/0092-8674(84)90313-1.
McMahon, A. C. *et al.* (2016) 'TRIBE: Hijacking an RNA-Editing Enzyme to Identify Cell-Specific Targets of RNA-Binding Proteins', *Cell*, 165(3), pp. 742–753. doi: 10.1016/j.cell.2016.03.007.
McPherson, N. O. *et al.* (2013) 'Improving Metabolic Health in Obese Male Mice via Diet and Exercise Restores Embryo Development and Fetal Growth', *PLoS ONE*. Edited by M. B. Aguila. Public Library of Science, 8(8), p. e71459. doi: 10.1371/journal.pone.0071459.
Messerschmidt, D. M. *et al.* (2012) 'Trim28 Is Required for Epigenetic Stability During Mouse Oocyte to Embryo Transition', *Science*, 335(6075), pp. 1499–1502. doi: 10.1126/science.1216154.
Messerschmidt, D. M., Knowles, B. B. and Solter, D. (2014) 'DNA methylation dynamics during epigenetic reprogramming in the germline and preimplantation embryos.', *Genes & development*. Cold Spring Harbor Laboratory Press, 28(8), pp. 812–28. doi: 10.1101/gad.234294.113.
Miao, S. L., Bazzaz, F. A. and Primack, R. B. (1991) 'Persistence of Maternal Nutrient Effects in Plantago Major: The Third Generation', *Ecology*. John Wiley & Sons, Ltd, 72(5), pp. 1634–1642. doi: 10.2307/1940963.
Migicovsky, Z., Yao, Y. and Kovalchuk, I. (2014) 'Transgenerational phenotypic and epigenetic changes in response to heat stress in *Arabidopsis thaliana*', *Plant Signaling & Behavior*, 9(2), p. e27971. doi: 10.4161/psb.27971.
Miguel-Aliaga, I. *et al.* (2000) 'Disruption of SMN function by ectopic expression of the human *SMN* gene in *Drosophila*', *FEBS Letters*, 486(2), pp. 99–102. doi: 10.1016/S0014-5793(00)02243-2.
Mirouze, M. *et al.* (2009) 'Selective epigenetic control of retrotransposition in Arabidopsis', *Nature*. Nature Publishing Group, 461(7262), pp. 427–430. doi: 10.1038/nature08328.
Miska, E. A. and Ferguson-Smith, A. C. (2016) 'Transgenerational inheritance: Models and mechanisms of non-DNA sequence-based inheritance.', *Science (New York, N.Y.)*. American Association for the Advancement of Science, 354(6308), pp. 59–63. doi: 10.1126/science.aaf4945.
Mitchell, M., Bakos, H. W. and Lane, M. (2011) 'Paternal diet-induced obesity impairs embryo development and implantation in the mouse', *Fertility and Sterility*. Elsevier, 95(4), pp. 1349–1353. doi: 10.1016/J.FERTNSTERT.2010.09.038.
Mohn, F. *et al.* (2014) 'The Rhino-Deadlock-Cutoff Complex Licenses Noncanonical Transcription of Dual-Strand piRNA Clusters in Drosophila', *Cell*. Cell Press, 157(6), pp. 1364–1379. doi: 10.1016/J.CELL.2014.04.031.
Molaro, A. *et al.* (2014) 'Two waves of de novo methylation during mouse germ cell development', *Genes & Development*. Cold Spring Harbor Laboratory Press,

28(14), pp. 1544–1549. doi: 10.1101/GAD.244350.114.
Molla-Herman, A. *et al.* (2015) ‘tRNA processing defects induce replication stress and Chk2-dependent disruption of piRNA transcription.’, *The EMBO journal*. EMBO Press, 34(24), pp. 3009–27. doi: 10.15252/embj.201591006.
Monneron, A. and Bernhard, W. (1969) ‘Fine structural organization of the interphase nucleus in some mammalian cells’, *Journal of Ultrastructure Research*, 27(3–4), pp. 266–288. doi: 10.1016/S0022-5320(69)80017-1.
Montellier, E. *et al.* (2013) ‘Chromatin-to-nucleoprotamine transition is controlled by the histone H2B variant TH2B.’, *Genes & development*. Cold Spring Harbor Laboratory Press, 27(15), pp. 1680–92. doi: 10.1101/gad.220095.113.
Morgan, H. D. *et al.* (1999) ‘Epigenetic inheritance at the agouti locus in the mouse’, *Nature Genetics*. Nature Publishing Group, 23(3), pp. 314–318. doi: 10.1038/15490.
Morgan, H. D. *et al.* (2005) ‘Epigenetic reprogramming in mammals.’, *Human molecular genetics*, 14 Spec No(suppl_1), pp. R47-58. doi: 10.1093/hmg/ddi114.
Morgan, T. (1915) *The mechanism of Mendelian heredity*. Available at: https://books.google.com/books?hl=en&lr=&id=734ZAAAAYAAJ&oi=fnd&pg=PA1&ots=G0YldWI8ZQ&sig=NyqR1tVHDfBf2QMEjvn_r--k3Mo (Accessed: 10 April 2019).
Morison, I. and Reeve, A. E. (1998) ‘A catalogue of imprinted genes and parent-of-origin effects in humans and animals’, *Human Molecular Genetics*. Narnia, 7(10), pp. 1599–1609. doi: 10.1093/hmg/7.10.1599.
Murchison, E. P. *et al.* (2007) ‘Critical roles for Dicer in the female germline’, *Genes & Development*, 21(6), pp. 682–693. doi: 10.1101/gad.1521307.
Murphy, P. J. *et al.* (2018) ‘Placeholder Nucleosomes Underlie Germline-to-Embryo DNA Methylation Reprogramming’, *Cell*. Cell Press, 172(5), pp. 993-1006.e13. doi: 10.1016/J.CELL.2018.01.022.
Nagamori, I. *et al.* (2015) ‘Comprehensive DNA Methylation Analysis of Retrotransposons in Male Germ Cells’, *Cell Reports*, 12(10), pp. 1541–1547. doi: 10.1016/j.celrep.2015.07.060.
Nakamura, T. *et al.* (2007) ‘PGC7/Stella protects against DNA demethylation in early embryogenesis’, *Nature Cell Biology*. Nature Publishing Group, 9(1), pp. 64–71. doi: 10.1038/ncb1519.
Narita, T. *et al.* (2007) ‘NELF Interacts with CBC and Participates in 3′ End Processing of Replication-Dependent Histone mRNAs’, *Molecular Cell*, 26(3), pp. 349–365. doi: 10.1016/j.molcel.2007.04.011.
Nekrasov, M. P. *et al.* (2003) ‘The mRNA-binding Protein YB-1 (p50) Prevents Association of the Eukaryotic Initiation Factor eIF4G with mRNA and Inhibits Protein Synthesis at the Initiation Stage’, *Journal of Biological Chemistry*, 278(16), pp. 13936–13943. doi: 10.1074/jbc.M209145200.
Nellåker, C. *et al.* (2012) ‘The genomic landscape shaped by selection on transposable elements across 18 mouse strains.’, *Genome biology*. BioMed Central, 13(6), p. R45. doi: 10.1186/gb-2012-13-6-r45.
Ng, S.-F. *et al.* (2010a) ‘Chronic high-fat diet in fathers programs β-cell dysfunction in female rat offspring.’, *Nature*. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved., 467(7318), pp. 963–6. doi:

10.1038/nature09491.
Ng, S.-F. *et al.* (2010b) ‘Chronic high-fat diet in fathers programs β-cell dysfunction in female rat offspring.’, *Nature*. Nature Publishing Group, 467(7318), pp. 963–6. doi: 10.1038/nature09491.
Nguyen, C. T. *et al.* (2002) ‘Histone H3-lysine 9 methylation is associated with aberrant gene silencing in cancer cells and is rapidly reversed by 5-aza-2’-deoxycytidine.’, *Cancer research*. American Association for Cancer Research, 62(22), pp. 6456–61. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12438235 (Accessed: 24 April 2019).
van Niel, G., D’Angelo, G. and Raposo, G. (2018a) ‘Shedding light on the cell biology of extracellular vesicles’, *Nature Reviews Molecular Cell Biology*. Nature Publishing Group, 19(4), pp. 213–228. doi: 10.1038/nrm.2017.125.
van Niel, G., D’Angelo, G. and Raposo, G. (2018b) ‘Shedding light on the cell biology of extracellular vesicles’, *Nature Reviews Molecular Cell Biology*. Nature Publishing Group, 19(4), pp. 213–228. doi: 10.1038/nrm.2017.125.
Nizami, Z., Deryusheva, S. and Gall, J. G. (2010) ‘The Cajal body and histone locus body.’, *Cold Spring Harbor perspectives in biology*. Cold Spring Harbor Laboratory Press, 2(7), p. a000653. doi: 10.1101/cshperspect.a000653.
Nobuta, K. *et al.* (2008) ‘Distinct size distribution of endogeneous siRNAs in maize: Evidence from deep sequencing in the mop1-1 mutant.’, *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 105(39), pp. 14958–63. doi: 10.1073/pnas.0808066105.
Nora, E. P. *et al.* (2017) ‘Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization’, *Cell*. Cell Press, 169(5), pp. 930-944.e22. doi: 10.1016/J.CELL.2017.05.004.
Novotný, I. *et al.* (2011) ‘In vivo kinetics of U4/U6·U5 tri-snRNP formation in Cajal bodies’, *Molecular Biology of the Cell*. Edited by L. Edelstein-Keshet, 22(4), pp. 513–523. doi: 10.1091/mbc.e10-07-0560.
O’Donnell, L. (2014) ‘Mechanisms of spermiogenesis and spermiation and how they are disturbed’, *Spermatogenesis*, 4(2), p. e979623. doi: 10.4161/21565562.2014.979623.
Ohnishi, Y. *et al.* (2010) ‘Small RNA class transition from siRNA/piRNA to miRNA during pre-implantation mouse development.’, *Nucleic acids research*, 38(15), pp. 5141–51. doi: 10.1093/nar/gkq229.
Oswald, J. *et al.* (2000) ‘Active demethylation of the paternal genome in the mouse zygote’, *Current Biology*, 10(8), pp. 475–478. doi: 10.1016/S0960-9822(00)00448-6.
Pagano, J. M., Clingman, C. C. and Ryder, S. P. (2011) ‘Quantitative approaches to monitor protein-nucleic acid interactions using fluorescent probes.’, *RNA (New York, N.Y.)*. Cold Spring Harbor Laboratory Press, 17(1), pp. 14–20. doi: 10.1261/rna.2428111.
Palmer, J. R. *et al.* (2001) ‘Infertility among Women Exposed Prenatally to Diethylstilbestrol’, *American Journal of Epidemiology*. Narnia, 154(4), pp. 316–321. doi: 10.1093/aje/154.4.316.
Peach, S. E., York, K. and Hesselberth, J. R. (2015) ‘Global analysis of RNA cleavage by 5′-hydroxyl RNA sequencing’, *Nucleic Acids Research*. Narnia,

43(17), pp. e108–e108. doi: 10.1093/nar/gkv536.
Peaston, A., Evsikov, A. V, *et al.* (2004) 'Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos.', *Developmental cell*, 7(4), pp. 597–606. doi: 10.1016/j.devcel.2004.09.004.
Peaston, A., Evsikov, A., *et al.* (2004) 'Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos', *Developmental Cell*, 7(4), pp. 597–606. doi: 10.1016/j.devcel.2004.09.004.
Pecinka, A. and Mittelsten Scheid, O. (2012) 'Stress-Induced Chromatin Changes: A Critical View on Their Heritability', *Plant and Cell Physiology*, 53(5), pp. 801–808. doi: 10.1093/pcp/pcs044.
Pellizzoni, L., J, Y. and G., D. (2002) 'Essential Role for the SMN Complex in the Specificity of snRNP Assembly', *Science*, 298(5599), pp. 1775–1779. doi: 10.1126/science.1074962.
Pembrey, M. *et al.* (2014) 'Human transgenerational responses to early-life experience: potential impact on development, health and biomedical research.', *Journal of medical genetics*. BMJ Publishing Group, 51(9), pp. 563–72. doi: 10.1136/jmedgenet-2014-102577.
Pembrey, M. E. *et al.* (2006) 'Sex-specific, male-line transgenerational responses in humans', *European Journal of Human Genetics*, 14(2), pp. 159–166. doi: 10.1038/sj.ejhg.5201538.
Peng, H. *et al.* (2012a) 'A novel class of tRNA-derived small RNAs extremely enriched in mature mouse sperm', *Cell Research*, pp. 1609–1612. doi: 10.1038/cr.2012.141.
Peng, H. *et al.* (2012b) 'A novel class of tRNA-derived small RNAs extremely enriched in mature mouse sperm', *Cell Research*. Nature Publishing Group, 22(11), pp. 1609–1612. doi: 10.1038/cr.2012.141.
Pentinat, T. *et al.* (2010) 'Transgenerational Inheritance of Glucose Intolerance in a Mouse Model of Neonatal Overnutrition', *Endocrinology*. Narnia, 151(12), pp. 5617–5623. doi: 10.1210/en.2010-0684.
Percharde, M. *et al.* (2018) 'A LINE1-Nucleolin Partnership Regulates Early Development and ESC Identity', *Cell*. Elsevier, 174(2), pp. 391-405.e19. doi: 10.1016/J.CELL.2018.05.043.
Perez, M. F. and Lehner, B. (2019) 'Intergenerational and transgenerational epigenetic inheritance in animals', *Nature Cell Biology*. Nature Publishing Group, 21(2), pp. 143–151. doi: 10.1038/s41556-018-0242-9.
Pezic, D. *et al.* (2014) 'piRNA pathway targets active LINE1 elements to establish the repressive H3K9me3 mark in germ cells.', *Genes & development*, 28(13), pp. 1410–28. doi: 10.1101/gad.240895.114.
Pillai, R. S. *et al.* (2001) 'Purified U7 snRNPs lack the Sm proteins D1 and D2 but contain Lsm10, a new 14 kDa Sm D1-like protein', *The EMBO Journal*. EMBO Press, 20(19), pp. 5470–5479. doi: 10.1093/emboj/20.19.5470.
Pizzo, E. *et al.* (2013) 'Ribonuclease/angiogenin inhibitor 1 regulates stress-induced subcellular localization of angiogenin to control growth and survival', *Journal of Cell Science*, 126(18), pp. 4308–4319. doi: 10.1242/jcs.134551.
Popp, C. *et al.* (2010) 'Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency.', *Nature*. Macmillan

Publishers Limited. All rights reserved, 463(7284), pp. 1101–5. doi: 10.1038/nature08829.
Protter, D. S. W. and Parker, R. (2016) ‘Principles and Properties of Stress Granules.’, *Trends in cell biology*. Elsevier, 26(9), pp. 668–679. doi: 10.1016/j.tcb.2016.05.004.
Ptashne, M. (2007) ‘On the use of the word “epigenetic”’, *Current Biology*. Cell Press, 17(7), pp. R233–R236. doi: 10.1016/J.CUB.2007.02.030.
Qian, Z. *et al.* (1998) ‘Yeast Ty1 retrotransposition is stimulated by a synergistic interaction between mutations in chromatin assembly factor I and histone regulatory proteins.’, *Molecular and cellular biology*. American Society for Microbiology (ASM), 18(8), pp. 4783–92. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9671488 (Accessed: 5 April 2019).
Quenneville, S. *et al.* (2012) ‘The KRAB-ZFP/KAP1 System Contributes to the Early Embryonic Establishment of Site-Specific DNA Methylation Patterns Maintained during Development’, *Cell Reports*. Cell Press, 2(4), pp. 766–773. doi: 10.1016/J.CELREP.2012.08.043.
Rabani, M. *et al.* (2011) ‘Metabolic labeling of RNA uncovers principles of RNA production and degradation dynamics in mammalian cells.’, *Nature biotechnology*, 29(5), pp. 436–442. doi: 10.1038/nbt.1861.
Radford, E. J. *et al.* (2012) ‘An Unbiased Assessment of the Role of Imprinted Genes in an Intergenerational Model of Developmental Programming’, *PLoS Genetics*. Edited by M. Bartolomei. Public Library of Science, 8(4), p. e1002605. doi: 10.1371/journal.pgen.1002605.
Radford, E. J. *et al.* (2014) ‘In utero undernourishment perturbs the adult sperm methylome and intergenerational metabolism’, *Science*, 345(6198), p. 1255903. doi: 10.1126/science.1255903.
Rajendra, T. K. *et al.* (2007) ‘A *Drosophila melanogaster* model of spinal muscular atrophy reveals a function for SMN in striated muscle’, *The Journal of Cell Biology*, 176(6), pp. 831–841. doi: 10.1083/jcb.200610053.
Ramsköld, D. *et al.* (2009) ‘An Abundance of Ubiquitously Expressed Genes Revealed by Tissue Transcriptome Sequence Data’, *PLoS Computational Biology*. Edited by L. J. Jensen, 5(12), p. e1000598. doi: 10.1371/journal.pcbi.1000598.
Rando, O. J. (2012) ‘Daddy issues: Paternal effects on phenotype’, *Cell*. Elsevier Inc., 151(4), pp. 702–708. doi: 10.1016/j.cell.2012.10.020.
Rando, O. J. and Simmons, R. A. (2015) ‘I’m Eating for Two: Parental Dietary Effects on Offspring Metabolism’, *Cell*. Cell Press, 161(1), pp. 93–105. doi: 10.1016/J.CELL.2015.02.021.
Rassoulzadegan, M. *et al.* (2006) ‘RNA-mediated non-mendelian inheritance of an epigenetic change in the mouse’, *Nature*. Nature Publishing Group, 441(7092), pp. 469–474. doi: 10.1038/nature04674.
Ratnam, S. *et al.* (2002) ‘Dynamics of Dnmt1 Methyltransferase Expression and Intracellular Localization during Oogenesis and Preimplantation Development’, *Developmental Biology*, 245(2), pp. 304–314. doi: 10.1006/dbio.2002.0628.
Regulski, M. *et al.* (2013) ‘The maize methylome influences mRNA splice sites and reveals widespread paramutation-like switches guided by small RNA.’,

Genome research. Cold Spring Harbor Laboratory Press, 23(10), pp. 1651–62. doi: 10.1101/gr.153510.112.
Reichmann, J. *et al.* (2012) 'Microarray analysis of LTR retrotransposon silencing identifies Hdac1 as a regulator of retrotransposon expression in mouse embryonic stem cells', *PLoS Computational Biology*, 8(4). doi: 10.1371/journal.pcbi.1002486.
Renfree, M., Hore, Timothy A., *et al.* (2009) 'Evolution of Genomic Imprinting: Insights from Marsupials and Monotremes', *Annual Review of Genomics and Human Genetics*. Annual Reviews, 10(1), pp. 241–262. doi: 10.1146/annurev-genom-082908-150026.
Renfree, M., Hore, Timothy A, *et al.* (2009) 'Evolution of Genomic Imprinting: Insights from Marsupials and Monotremes Genomic imprinting: parent-of-origin-specific expression'. doi: 10.1146/annurev-genom-082908-150026.
Reuter, M. *et al.* (2011) 'Miwi catalysis is required for piRNA amplification-independent LINE1 transposon silencing', *Nature*, 480(7376), pp. 264–267. doi: 10.1038/nature10672.
Ribet, D. *et al.* (2008) 'Murine endogenous retrovirus MuERV-L is the progenitor of the "orphan" epsilon viruslike particles of the early mouse embryo.', *Journal of virology*, 82(3), pp. 1622–1625. doi: 10.1128/JVI.02097-07.
Richardson, C. C. (1965) 'Phosphorylation of nucleic acid by an enzyme from T4 bacteriophage-infected Escherichia coli.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 54(1), pp. 158–65. Available at: http://www.ncbi.nlm.nih.gov/pubmed/5323016 (Accessed: 25 March 2019).
Robaire, B., Hinton, B. and Orgebin-Crist, M. (2002) *The Epididymis: From Molecules to Clinical Practice: From Molecules to Clinical Practice: a Comprehensive Survey of the Efferent Ducts, the Epididymis and*. Available at: https://books.google.com/books?hl=en&lr=&id=EuTm7wFpKQQC&oi=fnd&pg=PA1&ots=xuK24wZ8ty&sig=ESHtF5WwCvGYDJ0MgkZrKUONCbI (Accessed: 13 May 2019).
Robine, N. *et al.* (2009) 'A Broadly Conserved Pathway Generates 3′UTR-Directed Primary piRNAs', *Current Biology*, 19(24), pp. 2066–2076. doi: 10.1016/j.cub.2009.11.064.
Rodgers, A. B. *et al.* (2015) 'Transgenerational epigenetic programming via sperm microRNA recapitulates effects of paternal stress.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 112(44), pp. 13699–704. doi: 10.1073/pnas.1508347112.
Rodrigues, J. A. and Zilberman, D. (2015) 'Evolution and function of genomic imprinting in plants.', *Genes & development*. Cold Spring Harbor Laboratory Press, 29(24), pp. 2517–31. doi: 10.1101/gad.269902.115.
Rodriguez-Terrones, D. and Torres-Padilla, M.-E. (2018) 'Nimble and Ready to Mingle: Transposon Outbursts of Early Development', *Trends in Genetics*. Elsevier Current Trends, 34(10), pp. 806–820. doi: 10.1016/J.TIG.2018.06.006.
Rollins, R. A. *et al.* (2006) 'Large-scale structure of genomic methylation patterns', *Genome Research*. Cold Spring Harbor Laboratory Press, 16(2), pp. 157–163. doi: 10.1101/GR.4362006.

Romeo, V., Griesbach, E. and Schümperli, D. (2014) ‘CstF64: cell cycle regulation and functional role in 3’ end processing of replication-dependent histone mRNAs.’, *Molecular and cellular biology*. American Society for Microbiology Journals, 34(23), pp. 4272–84. doi: 10.1128/MCB.00791-14.
Rowe, H. M. *et al.* (2010) ‘KAP1 controls endogenous retroviruses in embryonic stem cells.’, *Nature*, 463(7278), pp. 237–240. doi: 10.1038/nature08674.
Saitou, M. and Kurimoto, K. (2014) ‘Paternal nucleosomes: are they retained in developmental promoters or gene deserts?’, *Developmental cell*. Elsevier, 30(1), pp. 6–8. doi: 10.1016/j.devcel.2014.06.025.
Santenard, A. *et al.* (2010) ‘Heterochromatin formation in the mouse embryo requires critical residues of the histone variant H3.3’, *Nature Cell Biology*. Nature Publishing Group, 12(9), pp. 853–862. doi: 10.1038/ncb2089.
Santos, F. *et al.* (2005) ‘Dynamic chromatin modifications characterise the first cell cycle in mouse embryos’, *Developmental Biology*, 280(1), pp. 225–236. doi: 10.1016/j.ydbio.2005.01.025.
Sasson, I. E. *et al.* (2015) ‘Pre-gestational vs gestational exposure to maternal obesity differentially programs the offspring in mice’, *Diabetologia*. Springer Berlin Heidelberg, 58(3), pp. 615–624. doi: 10.1007/s00125-014-3466-7.
Sassone-Corsi, P. (2002) ‘Unique Chromatin Remodeling and Transcriptional Regulation in Spermatogenesis’, *Science*. American Association for the Advancement of Science, 296(5576), pp. 2176–2178. doi: 10.1126/SCIENCE.1070963.
Saxe, J. P. *et al.* (2013) ‘Tdrkh is essential for spermatogenesis and participates in primary piRNA biogenesis in the germline’, *The EMBO Journal*, 32(13), pp. 1869–1885. doi: 10.1038/emboj.2013.121.
Schaefer, M. *et al.* (2010) ‘RNA methylation by Dnmt2 protects transfer RNAs against stress-induced cleavage.’, *Genes & development*, 24(15), pp. 1590–5. doi: 10.1101/gad.586710.
Schaffer, A. E. *et al.* (2014) ‘CLP1 founder mutation links tRNA splicing and maturation to cerebellar development and neurodegeneration.’, *Cell*, 157(3), pp. 651–63. doi: 10.1016/j.cell.2014.03.049.
Schaffert, N. *et al.* (2004) ‘RNAi knockdown of hPrp31 leads to an accumulation of U4/U6 di-snRNPs in Cajal bodies’, *The EMBO Journal*, 23(15), pp. 3000–3009. doi: 10.1038/sj.emboj.7600296.
Schaufele, F. *et al.* (1986) ‘Compensatory mutations suggest that base-pairing with a small nuclear RNA is required to form the 3′ end of H3 messenger RNA’, *Nature*. Nature Publishing Group, 323(6091), pp. 777–781. doi: 10.1038/323777a0.
Schifano, J. M. *et al.* (2014) ‘An RNA-seq method for defining endoribonuclease cleavage specificity identifies dual rRNA substrates for toxin MazF-mt3’, *Nature Communications*. Nature Publishing Group, 5(1), p. 3538. doi: 10.1038/ncomms4538.
Schmitt, J., Niles, J. and Wulff, R. D. (1992) ‘Norms of Reaction of Seed Traits to Maternal Environments in Plantago lanceolata’, *The American Naturalist*. University of Chicago Press , 139(3), pp. 451–466. doi: 10.1086/285338.
Schmitz, R. J. *et al.* (2011) ‘Transgenerational Epigenetic Instability Is a Source

of Novel Methylation Variants', *Science*, 334(6054), pp. 369–373. doi: 10.1126/science.1212959.
Schneider, S. *et al.* (2008) 'Vinclozolin—The lack of a transgenerational effect after oral maternal exposure during organogenesis', *Reproductive Toxicology*, 25(3), pp. 352–360. doi: 10.1016/j.reprotox.2008.04.001.
Schoorlemmer, J. *et al.* (2014) 'Regulation of Mouse Retroelement MuERV-L/MERVL Expression by REX1 and Epigenetic Control of Stem Cell Potency.', *Frontiers in oncology*, 4(February), p. 14. doi: 10.3389/fonc.2014.00014.
Schorn, Andrea J. *et al.* (2017) 'LTR-Retrotransposon Control by tRNA-Derived Small RNAs', *Cell*, 170(1), pp. 61-71.e11. doi: 10.1016/j.cell.2017.06.013.
Schorn, Andrea J *et al.* (2017) 'LTR-Retrotransposon Control by tRNA-Derived Small RNAs'. doi: 10.1016/j.cell.2017.06.013.
Schultz, D. C. *et al.* (2002) 'SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins', *Genes & Development*, 16(8), pp. 919–932. doi: 10.1101/gad.973302.
Schulz, K. N. and Harrison, M. M. (2019) 'Mechanisms regulating zygotic genome activation', *Nature Reviews Genetics*. Nature Publishing Group, 20(4), pp. 221–234. doi: 10.1038/s41576-018-0087-x.
Schutz, K., Hesselberth, J. R. and Fields, S. (2010) 'Capture and sequence analysis of RNAs with terminal 2',3'-cyclic phosphates.', *RNA (New York, N.Y.)*. Cold Spring Harbor Laboratory Press, 16(3), pp. 621–31. doi: 10.1261/rna.1934910.
Secco, D. *et al.* (2015) 'Stress induced gene expression drives transient DNA methylation changes at adjacent repetitive 1 elements 2 3 Arabidopsis', *Elife*. Available at: https://cdn.elifesciences.org/articles/09343/elife-09343-v1.pdf (Accessed: 18 May 2019).
Seki, Y. *et al.* (2007) 'Cellular dynamics associated with the genome-wide epigenetic reprogramming in migrating primordial germ cells in mice', *Development*, 134(14), pp. 2627–2638. doi: 10.1242/dev.005611.
Sela, N. *et al.* (2007) 'Comparative analysis of transposed element insertion within human and mouse genomes reveals Alu's unique role in shaping the human transcriptome', *Genome Biology*, 8(6), p. R127. doi: 10.1186/gb-2007-8-6-r127.
Shalek, A. K. *et al.* (2013) 'Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells', *Nature*. Nature Publishing Group, 498(7453), pp. 236–240. doi: 10.1038/nature12172.
Sharma, U. *et al.* (2016) 'Biogenesis and function of tRNA fragments during sperm maturation and fertilization in mammals', *Science*, 351(6271), pp. 391–396. doi: 10.1126/science.aad6780.
Sharma, U. *et al.* (2018) 'Small RNAs Are Trafficked from the Epididymis to Developing Mammalian Sperm', *Developmental Cell*. Cell Press, 46(4), pp. 481-494.e6. doi: 10.1016/J.DEVCEL.2018.06.023.
Sharma, U. and Rando, O. J. (2017) 'Metabolic Inputs into the Epigenome.', *Cell metabolism*. Elsevier, 25(3), pp. 544–558. doi: 10.1016/j.cmet.2017.02.003.
Shea, J. M. *et al.* (2015) 'Genetic and Epigenetic Variation, but Not Diet, Shape

the Sperm Methylome', *Developmental Cell*. Cell Press, 35(6), pp. 750–758. doi: 10.1016/J.DEVCEL.2015.11.024.
Sheldon, C. C. *et al.* (2008) 'Resetting of FLOWERING LOCUS C expression after epigenetic repression by vernalization.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 105(6), pp. 2214–9. doi: 10.1073/pnas.0711453105.
Shirayama, M. *et al.* (2012) 'piRNAs initiate an epigenetic memory of nonself RNA in the C. elegans germline.', *Cell*. Elsevier, 150(1), pp. 65–77. doi: 10.1016/j.cell.2012.06.015.
Shpargel, K. B. and Matera, A. G. (2005) 'Gemin proteins are required for efficient assembly of Sm-class ribonucleoproteins', *Proceedings of the National Academy of Sciences*, 102(48), pp. 17372–17377. doi: 10.1073/pnas.0508947102.
Sienski, G., Dönertas, D. and Brennecke, J. (2012) 'Transcriptional silencing of transposons by Piwi and maelstrom and its impact on chromatin state and gene expression.', *Cell*, 151(5), pp. 964–80. doi: 10.1016/j.cell.2012.10.040.
Siklenka, K. *et al.* (2015a) 'Disruption of histone methylation in developing sperm impairs offspring health transgenerationally.', *Science (New York, N.Y.)*, 350(6261), pp. aab2006-. doi: 10.1126/science.aab2006.
Siklenka, K. *et al.* (2015b) 'Disruption of histone methylation in developing sperm impairs offspring health transgenerationally.', *Science (New York, N.Y.)*, 350(6261). doi: 10.1126/science.aab2006.
Skinner, M. K., Guerrero-Bosagna, C. and Haque, M. M. (2015) 'Environmentally induced epigenetic transgenerational inheritance of sperm epimutations promote genetic mutations', *Epigenetics*, 10(8), pp. 762–771. doi: 10.1080/15592294.2015.1062207.
Slamberová, R., Riley, M. A. and Vathy, I. (2005) 'Cross-generational effect of prenatal morphine exposure on neurobehavioral development of rat pups.', *Physiological research*, 54(6), pp. 655–60. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15717852 (Accessed: 21 May 2019).
Slaughter, A. *et al.* (2012) 'Descendants of primed Arabidopsis plants exhibit resistance to biotic stress.', *Plant physiology*. American Society of Plant Biologists, 158(2), pp. 835–43. doi: 10.1104/pp.111.191593.
Slevin, M. K. *et al.* (2014) 'Deep Sequencing Shows Multiple Oligouridylations Are Required for 3′ to 5′ Degradation of Histone mRNAs on Polyribosomes', *Molecular Cell*. Cell Press, 53(6), pp. 1020–1030. doi: 10.1016/J.MOLCEL.2014.02.027.
Slotkin, R. K. *et al.* (2009) 'Epigenetic Reprogramming and Small RNA Silencing of Transposable Elements in Pollen', *Cell*, 136(3), pp. 461–472. doi: 10.1016/j.cell.2008.12.038.
Slotkin, R. K., Freeling, M. and Lisch, D. (2005) 'Heritable transposon silencing initiated by a naturally occurring transposon inverted duplication', *Nature Genetics*. Nature Publishing Group, 37(6), pp. 641–644. doi: 10.1038/ng1576.
Slotkin, R. and Martienssen, R. (2007a) 'Transposable elements and the epigenetic regulation of the genome.', *Nature reviews. Genetics*, 8(4), pp. 272–85. doi: 10.1038/nrg2072.

Slotkin, R. and Martienssen, R. (2007b) 'Transposable elements and the epigenetic regulation of the genome', *Nature Reviews Genetics*. Nature Publishing Group, 8(4), pp. 272–285. doi: 10.1038/nrg2072.
Smirnov, A. *et al.* (2016) 'Grad-seq guides the discovery of ProQ as a major small RNA-binding protein', *Proceedings of the National Academy of Sciences*. National Academy of Sciences, 113(41), pp. 11591–11596. doi: 10.1073/PNAS.1609981113.
Smith, Z. *et al.* (2012) 'A unique regulatory phase of DNA methylation in the early mammalian embryo', *nature*, 484, pp. 339–44. Available at: https://www.nature.com/articles/nature10960 (Accessed: 17 May 2019).
Sobala, A. and Hutvagner, G. (2013) 'Small RNAs derived from the 5' end of tRNA can inhibit protein translation in human cells.', *RNA biology*, 10(4), pp. 553–63. doi: 10.4161/rna.24285.
Song, J. *et al.* (2012) 'Vernalization – a cold-induced epigenetic switch', *Journal of Cell Science*. The Company of Biologists Ltd, 125(16), pp. 3723–3731. doi: 10.1242/JCS.084764.
Soubry, A. (2015) 'Epigenetic inheritance and evolution: A paternal perspective on dietary influences.', *Progress in biophysics and molecular biology*. doi: 10.1016/j.pbiomolbio.2015.02.008.
Soumillon, M. *et al.* (2013) 'Cellular Source and Mechanisms of High Transcriptome Complexity in the Mammalian Testis', *Cell Reports*. Cell Press, 3(6), pp. 2179–2190. doi: 10.1016/J.CELREP.2013.05.031.
Spencer, H. G. and Clark, A. G. (2014) 'Non-conflict theories for the evolution of genomic imprinting', *Heredity*. Nature Publishing Group, 113(2), pp. 112–118. doi: 10.1038/hdy.2013.129.
Squires, J. E. *et al.* (2012) 'Widespread occurrence of 5-methylcytosine in human coding and non-coding RNA.', *Nucleic acids research*, 40(11), pp. 5023–33. doi: 10.1093/nar/gks144.
Stam, M. *et al.* (2002) 'Differential chromatin structure within a tandem array 100 kb upstream of the maize b1 locus is associated with paramutation.', *Genes & development*. Cold Spring Harbor Laboratory Press, 16(15), pp. 1906–18. doi: 10.1101/gad.1006702.
Staněk, D. *et al.* (2003) 'Targeting of U4/U6 small nuclear RNP assembly factor SART3/p110 to Cajal bodies', *The Journal of Cell Biology*, 160(4), pp. 505–516. doi: 10.1083/jcb.200210087.
Staněk, D. *et al.* (2008) 'Spliceosomal Small Nuclear Ribonucleoprotein Particles Repeatedly Cycle through Cajal Bodies', *Molecular Biology of the Cell*. Edited by W. Bickmore, 19(6), pp. 2534–2543. doi: 10.1091/mbc.e07-12-1259.
Stefani, G. and Slack, F. J. (2008) 'Small non-coding RNAs in animal development', *Nature Reviews Molecular Cell Biology*. Nature Publishing Group, 9(3), pp. 219–230. doi: 10.1038/nrm2347.
Stein, P. *et al.* (2003) 'RNAi: mammalian oocytes do it without RNA-dependent RNA polymerase.', *RNA (New York, N.Y.)*. Cold Spring Harbor Laboratory Press, 9(2), pp. 187–92. doi: 10.1261/RNA.2860603.
Strub, K., Galli, G. and Bimstiel, M. L. (1984) *The cDNA sequences of the sea urchin U7 small nuclear RNA suggest specific contacts between histone mRNA*

precursor and U7 RNA during RNA processing.; The cDNA sequences of the sea urchin U7 small nuclear RNA suggest specific contacts between histone mRNA precursor and U7 RNA during RNA processing., *The EMBO Journal*. doi: 10.1002/j.1460-2075.1984.tb02212.x.
Suh, N. *et al.* (2010) 'MicroRNA Function Is Globally Suppressed in Mouse Oocytes and Early Embryos', *Current Biology*, 20(3), pp. 271–277. doi: 10.1016/j.cub.2009.12.044.
Sullivan, K. D., Steiniger, M. and Marzluff, W. F. (2009) 'A Core Complex of CPSF73, CPSF100, and Symplekin May Form Two Different Cleavage Factors for Processing of Poly(A) and Histone mRNAs', *Molecular Cell*, 34(3), pp. 322–332. doi: 10.1016/j.molcel.2009.04.024.
Sullivan, R. (2015) 'Epididymosomes: a heterogeneous population of microvesicles with multiple functions in sperm maturation and storage', *Asian Journal of Andrology*. Medknow Publications and Media Pvt. Ltd., 0(0), p. 0. doi: 10.4103/1008-682X.155255.
Sullivan, R., Frenette, G. and Girouard, J. (2007) 'Epididymosomes are involved in the acquisition of new sperm proteins during epididymal transit', *Asian Journal of Andrology*, 9(4), pp. 483–491. doi: 10.1111/j.1745-7262.2007.00281.x.
Sullivan, R. and Saez, F. (2013) 'Epididymosomes, prostasomes, and liposomes: their roles in mammalian male reproductive physiology', *REPRODUCTION*, 146(1), pp. R21–R35. doi: 10.1530/REP-13-0058.
Sun, C. *et al.* (2014) 'Mouse early extra-embryonic lineages activate compensatory endocytosis in response to poor maternal nutrition', *Development*, 141(5), pp. 1140–1150. doi: 10.1242/dev.103952.
Surani, M. A. H., Barton, S. C. and Norris, M. L. (1984) 'Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis', *Nature*. Nature Publishing Group, 308(5959), pp. 548–550. doi: 10.1038/308548a0.
Suter, L. and Widmer, A. (2013) 'Environmental Heat and Salt Stress Induce Transgenerational Phenotypic Changes in Arabidopsis thaliana', *PLoS ONE*. Edited by M. A. Blazquez. Public Library of Science, 8(4), p. e60364. doi: 10.1371/journal.pone.0060364.
Svoboda, P. *et al.* (2004) 'RNAi and expression of retrotransposons MuERV-L and IAP in preimplantation mouse embryos', *Developmental Biology*, 269(1), pp. 276–285. doi: 10.1016/j.ydbio.2004.01.028.
Takahashi, K. and Yamanaka, S. (2006) 'Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors', *Cell*. Cell Press, 126(4), pp. 663–676. doi: 10.1016/J.CELL.2006.07.024.
Takahashi, N. *et al.* (2019a) 'ZNF445 is a primary regulator of genomic imprinting'. doi: 10.1101/gad.320069.
Takahashi, N. *et al.* (2019b) 'ZNF445 is a primary regulator of genomic imprinting'. doi: 10.1101/gad.320069.
Tam, O. H. *et al.* (2008) 'Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes.', *Nature*, 453(7194), pp. 534–538. doi: 10.1038/nature06904.
Tam, P. P. L. and Snow, M. H. L. (1981) 'Proliferation and migration of primordial

germ cells during compensatory growth in mouse embryos', *Development*, 64(1).
Tanaka, M. *et al.* (2001) 'A mammalian oocyte-specific linker histone gene H1oo: homology with the genes for the oocyte-specific cleavage stage histone (cs-H1) of sea urchin and the B4/H1M histone of the frog.', *Development (Cambridge, England)*, 128(5), pp. 655–64. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11171391 (Accessed: 12 April 2019).
Tassin, A. *et al.* (2013) 'DUX4 expression in FSHD muscle cells: how could such a rare protein cause a myopathy?', *Journal of Cellular and Molecular Medicine*, 17(1), pp. 76–89. doi: 10.1111/j.1582-4934.2012.01647.x.
Tatomer, D. C. *et al.* (2016) 'Concentrating pre-mRNA processing factors in the histone locus body facilitates efficient histone mRNA biogenesis', *J Cell Biol*. Rockefeller University Press, 213(5), pp. 557–570. doi: 10.1083/JCB.201504043.
Terzo, E. A. *et al.* (2015) 'Distinct self-interaction domains promote Multi Sex Combs accumulation in and formation of the *Drosophila* histone locus body', *Molecular Biology of the Cell*. Edited by K. Weis, 26(8), pp. 1559–1574. doi: 10.1091/mbc.E14-10-1445.
Le Thomas, A. *et al.* (2014) 'Transgenerationally inherited piRNAs trigger piRNA biogenesis by changing the chromatin of piRNA clusters and inducing precursor processing.', *Genes & development*. Cold Spring Harbor Laboratory Press, 28(15), pp. 1667–80. doi: 10.1101/gad.245514.114.
Thompson, D. M. and Parker, R. (2009) 'Stressing Out over tRNA Cleavage', *Cell*. Cell Press, 138(2), pp. 215–219. doi: 10.1016/J.CELL.2009.07.001.
Titus-Ernstoff, L. *et al.* (2010) 'Birth defects in the sons and daughters of women who were exposed in utero to diethylstilbestrol (DES)', *International Journal of Andrology*. John Wiley & Sons, Ltd (10.1111), 33(2), pp. 377–384. doi: 10.1111/j.1365-2605.2009.01010.x.
Tomizawa, S.-I., Nowacka-Woszuk, J. and Kelsey, G. (2012) 'DNA methylation establishment during oocyte growth: mechanisms and significance', *The International Journal of Developmental Biology*. UPV/EHU Press, 56(10-11–12), pp. 867–875. doi: 10.1387/ijdb.120152gk.
Torres-Padilla, M.-E. *et al.* (2006) 'Dynamic distribution of the replacement histone variant H3.3 in the mouse oocyte and preimplantation embryos', *The International Journal of Developmental Biology*. UPV/EHU Press, 50(Next), pp. 455–61. doi: 10.1387/ijdb.052073mt.
Torres, A. G., Batlle, E. and Ribas de Pouplana, L. (2014) 'Role of tRNA modifications in human diseases', *Trends in Molecular Medicine*. Elsevier Current Trends, 20(6), pp. 306–314. doi: 10.1016/J.MOLMED.2014.01.008.
Tosolini, M. *et al.* (2018) 'Contrasting epigenetic states of heterochromatin in the different types of mouse pluripotent stem cells', *Scientific Reports*. Nature Publishing Group, 8(1), p. 5776. doi: 10.1038/s41598-018-23822-4.
Trinkle-Mulcahy, L. and Sleeman, J. E. (2017) 'The Cajal body and the nucleolus: "In a relationship" or "It's complicated"?', *RNA biology*. Taylor & Francis, 14(6), pp. 739–751. doi: 10.1080/15476286.2016.1236169.
Tucker, K. E. *et al.* (2001) 'Residual Cajal bodies in coilin knockout mice fail to recruit Sm snRNPs and SMN, the spinal muscular atrophy gene product.', *The Journal of cell biology*. Rockefeller University Press, 154(2), pp. 293–307. doi:

10.1083/JCB.200104083.
Tuorto, F. *et al.* (2012) 'RNA cytosine methylation by Dnmt2 and NSun2 promotes tRNA stability and protein synthesis.', *Nature structural & molecular biology*. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved., 19(9), pp. 900–5. doi: 10.1038/nsmb.2357.
Turelli, P. *et al.* (2014) 'Interplay of TRIM28 and DNA methylation in controlling human endogenous retroelements.', *Genome research*. Cold Spring Harbor Laboratory Press, 24(8), pp. 1260–70. doi: 10.1101/gr.172833.114.
Valadi, H. *et al.* (2007) 'Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells', *Nature Cell Biology*. Nature Publishing Group, 9(6), pp. 654–659. doi: 10.1038/ncb1596.
Vallaster, M. P. *et al.* (2017) 'Paternal nicotine exposure alters hepatic xenobiotic metabolism in offspring', *eLife*, 6. doi: 10.7554/eLife.24771.
Vasiliauskaitė, L. *et al.* (2017) 'A MILI-independent piRNA biogenesis pathway empowers partial germline reprogramming', *Nature Structural & Molecular Biology*, 24(7), pp. 604–606. doi: 10.1038/nsmb.3413.
Vassoler, F. M. *et al.* (2013) 'Epigenetic inheritance of a cocaine-resistance phenotype', *Nature Neuroscience*, 16(1), pp. 42–47. doi: 10.1038/nn.3280.
Vaughn, M. W. *et al.* (2007) 'Epigenetic Natural Variation in Arabidopsis thaliana', *PLoS Biology*. Edited by J. C. Carrington. Public Library of Science, 5(7), p. e174. doi: 10.1371/journal.pbio.0050174.
Veselovska, L. *et al.* (2015) 'Deep sequencing and de novo assembly of the mouse oocyte transcriptome define the contribution of transcription to the DNA methylation landscape', *Genome Biology*. BioMed Central, 16(1), p. 209. doi: 10.1186/s13059-015-0769-z.
Villarroya-Beltri, C. *et al.* (2013) 'Sumoylated hnRNPA2B1 controls the sorting of miRNAs into exosomes through binding to specific motifs', *Nature Communications*. Nature Publishing Group, 4(1), p. 2980. doi: 10.1038/ncomms3980.
Viollet, S. *et al.* (2011) 'T4 RNA Ligase 2 truncated active site mutants: improved tools for RNA analysis', *BMC Biotechnology*. BioMed Central, 11(1), p. 72. doi: 10.1186/1472-6750-11-72.
Volpe, T. and Martienssen, R. A. (2011) 'RNA interference and heterochromatin assembly.', *Cold Spring Harbor perspectives in biology*, 3(9), p. a003731. doi: 10.1101/cshperspect.a003731.
Vourekas, A. *et al.* (2012) 'Mili and Miwi target RNA repertoire reveals piRNA biogenesis and function of Miwi in spermiogenesis', *Nature Structural & Molecular Biology*, 19(8), pp. 773–781. doi: 10.1038/nsmb.2347.
Vourekas, A. *et al.* (2015) 'The RNA helicase MOV10L1 binds piRNA precursors to initiate piRNA processing.', *Genes & development*. Cold Spring Harbor Laboratory Press, 29(6), pp. 617–29. doi: 10.1101/gad.254631.114.
Walsh, C. P., Chaillet, J. R. and Bestor, T. H. (1998) 'Transcription of IAP endogenous retroviruses is constrained by cytosine methylation', *Nature Genetics*. Nature Publishing Group, 20(2), pp. 116–117. doi: 10.1038/2413.
Walter, M. *et al.* (2016) 'An epigenetic switch ensures transposon repression upon dynamic loss of DNA methylation in embryonic stem cells.', *eLife*. eLife

Sciences Publications, Ltd, 5. doi: 10.7554/eLife.11418.
Wang, E. *et al.* (2012) 'Global Profiling of Alternative Splicing Events and Gene Expression Regulated by hnRNPH/F', *PLoS ONE*. Edited by Y. Xing. Public Library of Science, 7(12), p. e51266. doi: 10.1371/journal.pone.0051266.
Wang, E., Dimova, N. and Cambi, F. (2007) 'PLP/DM20 ratio is regulated by hnRNPH and F and a novel G-rich enhancer in oligodendrocytes', *Nucleic Acids Research*. Oxford University Press, 35(12), pp. 4164–4178. doi: 10.1093/nar/gkm387.
Wang, L. *et al.* (2014) 'Programming and Inheritance of Parental DNA Methylomes in Mammals', *Cell*. Cell Press, 157(4), pp. 979–991. doi: 10.1016/J.CELL.2014.04.017.
Wang, L. *et al.* (2018) 'Hijacking Oogenesis Enables Massive Propagation of LINE and Retroviral Transposons.', *Cell*. Elsevier, 174(5), pp. 1082-1094.e12. doi: 10.1016/j.cell.2018.06.040.
Wang, L. K., Lima, C. D. and Shuman, S. (2002) 'Structure and mechanism of T4 polynucleotide kinase: an RNA repair enzyme.', *The EMBO journal*. European Molecular Biology Organization, 21(14), pp. 3873–80. doi: 10.1093/emboj/cdf397.
Wang, Q. *et al.* (2013) 'Identification and functional characterization of tRNA-derived RNA fragments (tRFs) in respiratory syncytial virus infection.', *Molecular therapy : the journal of the American Society of Gene Therapy*, 21(2), pp. 368–79. doi: 10.1038/mt.2012.237.
Wang, Q. and Rio, D. C. (2018) 'JUM is a computational method for comprehensive annotation-free analysis of alternative pre-mRNA splicing patterns.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 115(35), pp. E8181–E8190. doi: 10.1073/pnas.1806018115.
Watanabe, T. *et al.* (2008) 'Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes.', *Nature*, 453(7194), pp. 539–543. doi: 10.1038/nature06908.
Watanabe, T. *et al.* (2011) 'Role for piRNAs and noncoding RNA in de novo DNA methylation of the imprinted mouse Rasgrf1 locus.', *Science (New York, N.Y.)*, 332(6031), pp. 848–852. doi: 10.1126/science.1203919.
Waterland, R. A. *et al.* (2006) 'Maternal methyl supplements increase offspring DNA methylation atAxin fused', *genesis*, 44(9), pp. 401–406. doi: 10.1002/dvg.20230.
Waterland, R. A. and Jirtle, R. L. (2003) 'Transposable elements: targets for early nutritional effects on epigenetic gene regulation.', *Molecular and cellular biology*. American Society for Microbiology Journals, 23(15), pp. 5293–300. doi: 10.1128/mcb.23.15.5293-5300.2003.
Weintraub, A. S. *et al.* (2017) 'YY1 Is a Structural Regulator of Enhancer-Promoter Loops'. doi: 10.1016/j.cell.2017.11.008.
Westmuckett, A. D. *et al.* (2014) 'Impaired Sperm Maturation in Rnase9 Knockout Mice1', *Biology of Reproduction*. Narnia, 90(6). doi: 10.1095/biolreprod.113.116863.
Whiddon, J. L. *et al.* (2017) 'Conservation and innovation in the DUX4-family

gene network', *Nature Genetics*. Nature Publishing Group, 49(6), pp. 935–940. doi: 10.1038/ng.3846.
White, A. E. *et al.* (2011) 'Drosophila histone locus bodies form by hierarchical recruitment of components', *The Journal of Cell Biology*. Rockefeller University Press, 193(4), pp. 677–694. doi: 10.1083/JCB.201012077.
Whitfield, M. L. *et al.* (2000) 'Stem-loop binding protein, the protein that binds the 3' end of histone mRNA, is cell cycle regulated by both translational and posttranslational mechanisms.', *Molecular and cellular biology*. American Society for Microbiology Journals, 20(12), pp. 4188–98. doi: 10.1128/MCB.20.12.4188-4198.2000.
Wibowo, A. *et al.* (2016) 'Hyperosmotic stress memory in Arabidopsis is mediated by distinct epigenetically labile sites in the genome and is restricted in the male germline by DNA glycosylase activity', *eLife*, 5. doi: 10.7554/eLife.13546.
Windhager, L. *et al.* (2012) 'Ultrashort and progressive 4sU-tagging reveals key characteristics of RNA processing at nucleotide resolution.', *Genome research*. Cold Spring Harbor Laboratory Press, 22(10), pp. 2031–42. doi: 10.1101/gr.131847.111.
Wolf, D. and Goff, S. P. (2007) 'TRIM28 Mediates Primer Binding Site-Targeted Silencing of Murine Leukemia Virus in Embryonic Cells', *Cell*. Cell Press, 131(1), pp. 46–57. doi: 10.1016/J.CELL.2007.07.026.
Wolf, D. and Goff, S. P. (2009) 'Embryonic stem cells use ZFP809 to silence retroviral DNAs', *Nature*. Nature Publishing Group, 458(7242), pp. 1201–1204. doi: 10.1038/nature07844.
Woods, L., Perez-Garcia, V. and Hemberger, M. (2018) 'Regulation of Placental Development and Its Impact on Fetal Growth—New Insights From Mouse Models', *Frontiers in Endocrinology*. Frontiers, 9, p. 570. doi: 10.3389/fendo.2018.00570.
Wu, C. H. and Gall, J. G. (1993) 'U7 small nuclear RNA in C snurposomes of the Xenopus germinal vesicle.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 90(13), pp. 6257–9. doi: 10.1073/PNAS.90.13.6257.
Wu, J. *et al.* (2016) 'The landscape of accessible chromatin in mammalian preimplantation embryos', *Nature*. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved., advance on. Available at: http://dx.doi.org/10.1038/nature18606.
Yamaguchi, K. *et al.* (2018) 'Re-evaluating the Localization of Sperm-Retained Histones Revealed the Modification-Dependent Accumulation in Specific Genome Regions', *CellReports*, 23, pp. 3920–3932. doi: 10.1016/j.celrep.2018.05.094.
Yamasaki, S., Ivanov, P., Hu, G.-F., *et al.* (2009) 'Angiogenin cleaves tRNA and promotes stress-induced translational repression.', *The Journal of cell biology*, 185(1), pp. 35–42. doi: 10.1083/jcb.200811106.
Yamasaki, S., Ivanov, P., Hu, G. F., *et al.* (2009) 'Angiogenin cleaves tRNA and promotes stress-induced translational repression', *Journal of Cell Biology*, 185(1), pp. 35–42. doi: 10.1083/jcb.200811106.

Yamazaki, T. *et al.* (2018) ‘TCF3 alternative splicing controlled by hnRNP H/F regulates E-cadherin expression and hESC pluripotency.’, *Genes & development*. Cold Spring Harbor Laboratory Press, 32(17–18), pp. 1161–1174. doi: 10.1101/gad.316984.118.
Yanagimachi, R. *et al.* (1985) ‘Maturation of spermatozoa in the epididymis of the Chinese hamster’, *American Journal of Anatomy*, 172(4), pp. 317–330. doi: 10.1002/aja.1001720406.
Yang, B. X. *et al.* (2015) ‘Systematic Identification of Factors for Provirus Silencing in Embryonic Stem Cells’, *Cell*. Elsevier, 163(1), pp. 230–45. doi: 10.1016/j.cell.2015.08.037.
Yang, Q. *et al.* (2016) ‘Highly sensitive sequencing reveals dynamic modifications and activities of small RNAs in mouse oocytes and early embryos.’, *Science advances*. American Association for the Advancement of Science, 2(6), p. e1501482. doi: 10.1126/sciadv.1501482.
Yang, X.-J. (2004) ‘The diverse superfamily of lysine acetyltransferases and their roles in leukemia and other diseases’, *Nucleic Acids Research*. Narnia, 32(3), pp. 959–976. doi: 10.1093/nar/gkh252.
Yang, X. *et al.* (2006) ‘Characterization of 3’hExo, a 3’ exonuclease specifically interacting with the 3’ end of histone mRNA.’, *The Journal of biological chemistry*. American Society for Biochemistry and Molecular Biology, 281(41), pp. 30447–54. doi: 10.1074/jbc.M602947200.
Yang, X. *et al.* (2014) ‘A conserved interaction that is essential for the biogenesis of histone locus bodies.’, *The Journal of biological chemistry*. American Society for Biochemistry and Molecular Biology, 289(49), pp. 33767–82. doi: 10.1074/jbc.M114.616466.
Ye, X. *et al.* (2003) ‘The cyclin E/Cdk2 substrate p220(NPAT) is required for S-phase entry, histone gene expression, and Cajal body maintenance in human somatic cells.’, *Molecular and cellular biology*, 23(23), pp. 8586–600. Available at: http://www.ncbi.nlm.nih.gov/pubmed/14612403 (Accessed: 15 April 2019).
Yeung, M. L. *et al.* (2009) ‘Pyrosequencing of small non-coding RNAs in HIV-1 infected cells: evidence for the processing of a viral-cellular double-stranded RNA hybrid.’, *Nucleic acids research*, 37(19), pp. 6575–86. doi: 10.1093/nar/gkp707.
Yeung, W. K. A. *et al.* (2019) ‘Histone H3K9 Methyltransferase G9a in Oocytes Is Essential for Preimplantation Development but Dispensable for CG Methylation Protection’, *Cell Reports*, 27, pp. 282–293. doi: 10.1016/j.celrep.2019.03.002.
Yoder, J. A., Walsh, C. P. and Bestor, T. H. (1997) ‘Cytosine methylation and the ecology of intragenomic parasites’, *Trends in Genetics*. Elsevier Current Trends, 13(8), pp. 335–340. doi: 10.1016/S0168-9525(97)01181-5.
Yoshida, K. *et al.* (2018) ‘Mapping of histone-binding sites in histone replacement-completed spermatozoa’, *Nature Communications*. Nature Publishing Group, 9(1), p. 3885. doi: 10.1038/s41467-018-06243-9.
Yuan, S. *et al.* (2014) ‘Methylation by NSun2 Represses the Levels and Function of MicroRNA 125b’, *Molecular and Cellular Biology*, 34(19), pp. 3630–3641. doi: 10.1128/MCB.00243-14.
Yuan, S. *et al.* (2016) ‘Sperm-borne miRNAs and endo-siRNAs are important for

fertilization and preimplantation embryonic development', *Development*. Oxford University Press for The Company of Biologists Limited, 143(4), pp. 635–647. doi: 10.1242/DEV.131755.
Zalzman, M. *et al.* (2010) 'Zscan4 regulates telomere elongation and genomic stability in ES cells', *Nature*. Nature Publishing Group, 464(7290), pp. 858–863. doi: 10.1038/nature08882.
Zamudio, N. *et al.* (2015) 'DNA methylation restrains transposons from adopting a chromatin signature permissive for meiotic recombination.', *Genes & development*. Cold Spring Harbor Laboratory Press, 29(12), pp. 1256–70. doi: 10.1101/gad.257840.114.
Zemach, A. *et al.* (2013) 'The Arabidopsis Nucleosome Remodeler DDM1 Allows DNA Methyltransferases to Access H1-Containing Heterochromatin', *Cell*. Cell Press, 153(1), pp. 193–205. doi: 10.1016/J.CELL.2013.02.033.
Zhang, B. *et al.* (2016) 'Allelic reprogramming of the histone modification H3K4me3 in early mammalian development', *Nature*. Nature Publishing Group, 537(7621), pp. 553–557. doi: 10.1038/nature19361.
Zhang, P. *et al.* (2015) 'MIWI and piRNA-mediated cleavage of messenger RNAs in mouse testes', *Cell Research*, 25(2), pp. 193–207. doi: 10.1038/cr.2015.4.
Zhang, S., Sun, L. and Kragler, F. (2009a) 'The phloem-delivered RNA pool contains small noncoding RNAs and interferes with translation.', *Plant physiology*. American Society of Plant Biologists, 150(1), pp. 378–87. doi: 10.1104/pp.108.134767.
Zhang, S., Sun, L. and Kragler, F. (2009b) 'The phloem-delivered RNA pool contains small noncoding RNAs and interferes with translation.', *Plant physiology*, 150(1), pp. 378–87. doi: 10.1104/pp.108.134767.
Zhang, Yunfang *et al.* (2018) 'Dnmt2 mediates intergenerational transmission of paternally acquired metabolic disorders through sperm small non-coding RNAs', *Nature Cell Biology*. Nature Publishing Group, 20(5), pp. 535–540. doi: 10.1038/s41556-018-0087-2.
Zhang, Z. *et al.* (2014) 'The HP1 Homolog Rhino Anchors a Nuclear Complex that Suppresses piRNA Precursor Splicing', *Cell*. Cell Press, 157(6), pp. 1353–1363. doi: 10.1016/J.CELL.2014.04.030.
Zhao, J. *et al.* (2000) 'NPAT links cyclin E-Cdk2 to the regulation of replication-dependent histone gene transcription.', *Genes & development*. Cold Spring Harbor Laboratory Press, 14(18), pp. 2283–97. doi: 10.1101/GAD.827700.
Zheng, H. *et al.* (2016) 'Resetting Epigenetic Memory by Reprogramming of Histone Modifications in Mammals', *Molecular Cell*, 63, pp. 1066–1079. doi: 10.1016/j.molcel.2016.08.032.
Zheng, K. *et al.* (2010) 'Mouse MOV10L1 associates with Piwi proteins and is an essential component of the Piwi-interacting RNA (piRNA) pathway', *Proceedings of the National Academy of Sciences*, 107(26), pp. 11841–11846. doi: 10.1073/pnas.1003953107.
Zheng, L., Roeder, R. G. and Luo, Y. (2003) 'S Phase Activation of the Histone H2B Promoter by OCA-S, a Coactivator Complex that Contains GAPDH as a Key Component', *Cell*. Elsevier, 114(2), pp. 255–266. doi: 10.1016/S0092-8674(03)00552-X.