DNA METHYLATION: *CIS*, *TRANS*, AND INTERGENERATIONAL

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

David Harrison Taylor

August 2016

DNA METHYLATION: *CIS*, *TRANS*, AND INTERGENERATIONAL

David Harrison Taylor, Ph.D.

Cornell University, 2016

ABSTRACT

Highly regulated epigenetic modifications, such as DNA methylation (5mC), are required for normal cell functions. The effects and interactions of *trans*-acting proteins that influence 5mC dynamics have been well studied; however, only a handful of *cis*-acting sequences that recruit these proteins and their activity have been thoroughly described. The *Rasgrf1* imprinting control region (ICR) presents a highly-regulated system of paternal-specific 5mC to study these *cis* elements. Briefly, a set of tandem repeats 30kb upstream of the *Rasgrf1* gene directs the expression of a piRNA-targeted noncoding RNA (pitRNA), which is recognized by the piRNA pathway and its associated 5mC machinery. This machinery induces 5mC on the region adjacent to the repeats known as the differentially methylated domain (DMD) specifically in the male germline, but not in the female germline, establishing allele-specific methylation.

In the first part of this work, I investigate if the *Rasgrf1* ICR is sufficient to recruit allele-specific 5mC patterns by ectopically inserting 3kb of ICR sequence at the *Wnt1* locus. While the novel insertion displays genetic reconstitution of imprinted 5mC, it does

not display imprinted expression from mutant *Wnt1* allele, arguing that imprinted expression requires a larger sequence context.

The endogenous *Rasgrf1* ICR has been shown to work only in *cis*; however, a previously characterized mutant mouse (*R2*), which bore the ICR of the *Igf2r* imprinted gene in place of the *Rasgrf1* repeats, displayed aberrant methylation and expression phenotypes both in *trans* and intergenerationally. The *R2* animals also misexpressed pitRNA both spatially and temporally, highlighting a possible mechanism for these phenotypes. In the second part of this work, I investigate the role of RNA in the aberrant expression phenotypes by injecting either the noncoding RNA or a control RNA into naive zygotes and analyzing global gene expression in the neonatal brain. While *Rasgrf1* allelic expression was not affected, I do observe three sets of molecular phenotypes: 1) the significant repression of ~10,000 genes which depends upon the RNA species being tested, 2) a global upregulation of transposable elements (TEs) which is RNA species independent, and 3) a global downward trend in all non-TE genes which is RNA species independent. Further studies will need to be conducted to better describe these expression effects.

This body of work advances our understanding of how the *cis*-acting sequences at *Rasgrf1* function and how misregulation of these sequences might cause unexpected effects. They further reveal the significance of zygotic RNA load, even through advanced stages in development.

BIOGRAPHICAL SKETCH

David Taylor was born at a very young age on March 31, 1988 to Thomas and Patricia Taylor in Richmond, VA. He wanted to be a medical doctor until, after an eighth grade project on the expectations of future careers, he ironically declared, “I don’t want to be in school that long.” He attended Nottoway County Public School system where he participated in soccer and wrestling. After graduation he promptly never played a sport again. He began defining his interest in science while attending Hampden-Sydney College, where Professor Nathaniel Hauck introduced him to the strange and fascinating subject of “epigenetics.” He pursued this interest in the summer of 2009 through a Research Education for Undergraduates (REU) internship at Cornell University under the tutelage of graduate student Haiyi Wang and professor Eric Richards. While there he tested the effects of nucleus size on heterochromatin formation and further cemented epigenetics as his future field of research. After graduating valedictorian of his class in 2010, he married his amazing wife, Esther Briscoe, and matriculated into the graduate school at Cornell University in the field of Genetics, Genomics, and Development. There he joined the lab of Paul Soloway, studying the sequences involved in RNA-mediated DNA methylation in mouse. During his graduate school tenure, he assisted in the birth of his son, Zachary Taylor. He completed his dissertation in the summer of 2016 and continued to a postdoctoral position at Yale University in the lab of Josien van Wolfswinkel, where he plans to study the effects of long non-coding RNAs in differentiation and maintenance of planarian neoblast cells.

ACKNOWLEDGEMENTS

I have reached this point in my life emotionally, spiritually, and scientifically only because of a large network of people who keep me afloat in tempestuous waters. They have formed and shaped me in profound ways that I will probably never grasp, but I know that it has been for the better. There is no way I can repay them, but I can give a small sample of what they've meant to me here. I am grateful...

To Paul Soloway, whose influence has made me a better scientist and better person. His high standards (**CONTROLS!**) combined with a supportive, easy-going style have made his lab an ideal place to learn and grow. He is an example of a well balanced life, excelling in a wide variety of hobbies outside of science that include making delicious smoked salmon, beating me in Frisbee golf, competing in triathlons, and making super hoppy batches of beer.

To my other committee members, Paula Cohen and Eric Richards, for their scientific insights and steadfast support for my interests and career goals. I would especially like to thank Eric Richards, whose example of patient and thoughtful guidance during my summer internship has shaped my ideal for the type of undergraduate mentor I strive to be. He also taught me that plants are not just leafy green things your mom makes you eat when you're four, but are also beautifully complex systems whose primary defense against environmental onslaught is gene regulation.

To the “old guard” (Patrick Murphy, Jonathan Flax, and Jim Hagarman) who have all moved on from the Soloway lab. They were patient with me as a fresh graduate student. When I knew so little, they took the time to teach me techniques and critical thinking skills without being patronizing or condescending. I would especially like to thank Jonathan Flax, whose thoughtful and eloquent discussions of science and politics would keep me engaged long after I should have caught the bus home (sorry Esther).

To the “new guard” (Erin Chu and Roman Spektor) of the Soloway lab who soldier on without me. I appreciate their infectious excitement which often takes the form of standing in front of a whiteboard and tossing project ideas around. Erin’s ambitious and hardworking attitude are inspirational, second only to her scientific intuitions. Plus she also puts up with my terrible dad jokes, which I greatly appreciate. Roman not only singlehandedly updated my technological sophistication to 2010—helping me cope with my Luddite approach to the world—but he also taught me that a deep skepticism of data, including one’s own, is an incredible scientific tool. I plan on using his “rules of benchwork” (Rule #1: Do it right the first time, aka, do pilot experiments before the real one; etc...) to teach any future trainees methodological rigor at the bench.

To Jim Putnam, the foundation of the Soloway lab, who keeps everything running smoothly. Some might choose his Frisbee golf or cooking skills as his best traits, but

I would choose his sunny, considerate, and helpful disposition that makes me find him every morning just to say hi.

To my church and Graduate Christian Fellowship (GCF) family, who have been constant sources of friendship and theological stimulation for me through all of these years. The Buskohls, the Braggs, the Cullies, the Brooks, the Felkers, and the Holdsworths have been lavish in their hospitality and generous with their time.

To my adopted Ithaca parents, Bryn and Susan Kherli, whose door is always open and who know how to make one feel at home, even in the midst of a comically large dinner party.

To Robbie Wine, whose wit, compassion, intellectual fortitude, and friendship have shaped me in the most deep and profound ways.

To my parents, Vince and Pat Taylor, who not only gave me life, but have unquestioningly supported every big decision I've made, even when it takes me far away. They are always lending a listening ear and prepared to back me up whenever I'm in need. The world would be a safer and kinder place if there were more people like them.

To my brother, John Taylor, who is my best friend. He is an insightful, intelligent, and hardworking example of masculinity who I've looked up to ever since I can remember. One day, when I grow up, I'm going to be like him.

To Zach, my son, a hilarious bright spot at the end of every day.

To my "favorite" wife Esther, who has been my ever-present encouragement and cheerleader. She is industrious, kind, gentle, and understanding in ways that are rarely found in this world. Her adventurous spirit has strong-armed me into many fun and enjoyable trips including Hawaii, the northeast, and *many* places in New York. Without her I'd definitely be a more boring character.

To Jesus, upon whom my hope is built—both in this world and the next.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION

Epigenetics

With few exceptions, every cell in the body contains the same genetic code, a remarkable fact considering there are hundreds of different cell types which display a variety of morphologies and functions that arise from this single blueprint. Not all DNA is expressed in all cells, however, and epigenetics—literally “above genetics”—is the study of heritable changes in gene expression that are not encoded in the DNA. During development these modifications can be thought as a “push” that forces a cell down one differentiation pathway versus another; or alternatively, they are the “walls” that stop a cell from changing fates after a decision is made. Epigenetics is also the mechanism by which genes respond to the environment, allowing plasticity without requiring changes in the DNA code. Common players include histone modifications, non-coding RNA, chromatin architecture, DNA methylation, and more. There is great variability amongst these marks. Some, like certain histone marks, are highly conserved throughout all eukaryotes, while others, like DNA cytosine methylation, is vital to some species but absent in many others. To keep the discussion concise for this array of marks and species, this work will focus on the mammalian system, only mentioning others when particularly relevant.

Histone modifications

DNA does not exist as a linear polymer *in vivo*, but instead is wrapped around octamers of four histone proteins: H2A, H2B, H3, and H4. Modifications to these proteins can modulate how the local DNA is interacts with transcriptional machinery (reviewed in [1]). There are hundreds of residues that can be modified both on the tail and within the globular histone structure with a variety of different marks (methylation acetylation, ubiquitination, phosphorylation, sumoylation, etc.), most of which have not been thoroughly investigated. Each modification has a variety of proteins which mediate its placement (writers), removal (erasers), and functional effects (readers). Readers are especially important as they bind to sites of modification and interact with other regulatory factors. This section will focus on the most relevant and well-established modifications.

Histone modifications are found at different *cis* regulatory features and can give insight into transcription state[1]. For instance, acetylation of lysine residues activates transcription not only by reducing the a histone's positive charge and loosening its association with the negatively charged DNA[2], but also by recruitment of accessory proteins. In *S. cerevisae*, histone acetylation recruits the Gcn5p histone acetyltransferase (HAT) protein which can acetylate adjacent histone proteins. These newly acetylated histones can also recruit Gcn5p in a feed-forward loop that propagates the spread of acetylation over the entire promoter region, inducing gene expression[3]. Some histone modifications were originally related to their genomic regulatory features

by correlation rather than mechanistic studies. The trimethylation of histone 3 lysine 3 (H3K4me3) is associated with high GC transcriptional start sites[4], while trimethylation of histone 3 lysine 36 (H3K36me3) is distributed along the gene body of actively elongating genes[5], possibly influencing splicing[6]. In yeast, the chromodomain of histone deacetylase (HDAC) protein Eaf3 recognizes H3K36me3 and removes acetyl marks on gene bodies, repressing internal transcriptional initiation[7]. Methylation of histone 3 lysine 4 (H3K4me1) is associated with enhancer regions[8], but acetylation of histone 3 lysine 27 (H3K27ac) via the HAT protein p300/CBP represents their activation[9]. H3K27ac at enhancers likely encourages the transcription of enhancer associated noncoding RNAs (eRNAs)[10], which facilitate enhancer-promoter interactions and increase transcription of their associated genes[11].

In contrast to these marks associated with gene activation, trimethylation of histone 3 lysine 9 (H3K9me3) modification denotes repressed pericentromeric heterochromatin[12,13] and is established by lysine methyltransferase (KMT) SUV39H1[14]. Much like histone acetylation, H3K9me3 can propagate its own spread by binding heterchromatin protein 1 (HP1)[15]. HP1, in turn, can interact with chromatin remodelers like BRG1[16], other HP1 proteins[17], DNA methyltransferases[18], or more SUV39H1 KMT proteins[19] to methylate adjacent histones. Another repressive mark is trimethylation of histone 3 lysine 27 (H3K27me3), which is established by the PRC2 KMT complex and induces facultative heterochromatin[20]. It holds the distinction of being bivalently colocalized with H3K4me3 in ESCs, where it signals a poised state in preparation for developmental signals[21]. Much like H3K27me3, the dimethylation of histone 3 lysine 9 is

associated with gene repression in euchromatic areas of the genome, but unlike H3K27me3 it is localized in continuous megabase-long blocks[22] which are associated with the nuclear periphery[23]. It is establishment by the G9a KMT[24], a protein which can recognize its own modification[25] and also facilitate its own spreading.

Cells must adapt their transcriptional profiles in response to development or environmental stimuli, so mechanisms for robust removal of these marks must be in place[26,27]. For instance, murine LSD1 is a lysine demethylase (KDM), which must decommission enhancers during embryonic stem cell (ESC) differentiation by removing the H3K4me1 mark[28]. During this same differentiation process the *HOXB1* promoter, which is normally coated by H3K27me3, recruits the KDM UTX to remove the repressive mark and allow for cell fate-specific expression[29]. Some mechanisms for histone mark removal even require multiple interacting partners, such as when PRC2-mediated association with H3K36me3 encourages the activity of KDM NO66—removing the H3K36me3 mark while establishing H3K27me3 during ESC development[30]. Many more erasers of epigenetic states exist, their expression and localization reflecting the immediate requirements of the cell.

Chromatin architecture

The DNA/histone “beads on a string” structure is only one facet of higher-order chromatin architecture (reviewed in [31]). Development of chromatin conformation capture

technology has unveiled megabase-scale chromosomal interaction networks known as topologically associated domains (TADs), lengths of DNA which interact with each other more frequently than at random. TADs form compartments of DNA in a similar transcriptional state, the boundaries of which are defined by a multiplicity genomic features like tRNA sequences, constitutively transcribed housekeeping genes, and binding sites of architectural proteins (CTCF, cohesin, condensin)[32]. These boundaries are stable throughout development, between cell types, and even across species[32,33], leaving inter-domain epigenetic states to control cell-type specific regulation. For instance, some TADs are repressed primarily through recruitment of H3K27me3 while others are repressed by recruitment to the nuclear periphery via interaction with H3K9me2 and the nuclear lamina[34].

The CCCTC-binding factor (CTCF) is a multifunctional protein which is particularly important for genome organization. It was once thought to act as an "enhancer blocker," since placement of CTCF sites between an enhancer and its gene was often sufficient to disrupt transcription[35–37]. A more nuanced understanding of its cellular action has emerged, however, as CTCF has been implicated in a variety of regulatory processes which are likely connected to its role in controlling chromatin architecture. Knockdown of the protein relaxes TAD boundaries[38], indicating that it is required to maintain megabase-scale topology. It also mediates DNA:DNA interactions on kilobase-scale, influencing local chromatin context by creating three dimensional loops which control gene expression and enhancer-promoter interactions[39,40]. Deletion of a CTCF site within the *HOXA* cluster in *D. melanogaster* resulted in epigenetic

disruption and active chromatin marks spreading into adjacent regions which were previously repressed[41], illustrating its function as a barrier to epigenetic spreading. Its binding displays both conserved and cell-type specific patterns[42] which have been shown to be methylation sensitive[43,44]. It also regulated by direct interaction with RNA. For instance, the interaction of CTCF with *Wrap53*, a noncoding RNA associated with the *p53* gene, is required for normal *p53* expression and response to DNA damage[45].

Non-coding RNA

The ENCODE project discovered that approximately 75% of the genome is transcribed[46], of which less than 3% codes for protein. While it is possible to have non-functional transcription, many years of research into non-coding RNA (ncRNA) have revealed a panoply of activities which are vital for cell viability. rRNA, tRNA, and other infrastructural RNA have been characterized for over forty years, but the functions of regulatory ncRNA are still being elucidated. These regulatory ncRNAs can be divided into two classes: short and long.

Short ncRNA

Short ncRNAs are <200nt RNAs which are broken into subclasses based on their biogenesis and behavior (reviewed extensively in [47,48]). Briefly, single stranded, ~22nt microRNA (miRNA) are processed by two RNAse III-like proteins, DROSHA and DICER. These RNAs degrade mRNAs which contain homologous binding sites by

recruitment of an Argonaute (Ago) protein, with miRNAs of perfect similarity inducing the Ago endonucleolytic activity and miRNAs of imperfect similarity prompting translational repression. Double stranded, ~21nt small interfering RNAs (siRNAs) can act in a similar mechanism for post-transcriptional gene silencing, but their biogenesis is not DROSHA-dependent. They have been shown to induce heterochromatin formation in plants and yeast, and affect transcriptional silencing as well. PIWI-interacting RNAs (piRNAs) are a class of particular interest to this work for reasons which will become clear later, so I will describe them in greater detail[49].

piRNAs were originally described as repeat-associated small interfering RNAs (rasiRNAs) in protozoa[50] and *D. melanogaster*[51]. Their discovery in mammals was the result of two main lines of logic. First, small RNAs had been implicated in the silencing expression of repetitive sequences to maintain the germline[52]. Second, PIWI proteins, a clade of the small-RNA associated Ago family, were also necessary for germline maintenance[53]. PIWI proteins were immunoprecipitation and their piRNAs characterized in a series of RNA-seq studies[54–56]. They have been shown to be vital for genome stability in many species including *D. melanogaster*[57], *C. elegans*[58,59], and mammals[54]. Many overlaps exist between these systems, especially between flies and mice; for instance, in both of these species piRNAs are 24-32nt in length, predominantly found in the germline, and are stabilized by 2'-O-methylation on their 3' ends[60]. In both species they are divided into two subclasses: pre-pachytene and pachytene. In mammals, pachytene piRNAs are generated from gene clusters in meiotic cells during postnatal spermatogenesis[55,56,61]. While their function is not as well characterized as pre-

pachytene RNA, they are known to associate with the Ago-clade MIWI protein to restrict mobile DNA sequences—known as transposable elements (TEs)—from activating and inducing genome instability during spermatogenesis[62]. They may also play a role in eliminating mRNAs during sperm maturation[63].

In *D. melanogaster*, pre-pachytene piRNAs are generated by Ago-clade proteins Ago3 and Aubergine (Aub) in a “ping-pong” mechanism during embryonic gametogenesis described below[64,65]. Primary piRNAs are antisense transcripts processed from large genomic clusters that bear similarity to TEs. They are loaded onto Ago3, which then recognizes and cleaves actively transcribed TEs. This sense-stranded cleavage product is then processed and loaded onto Aub, which can subsequently generate new primary piRNA transcripts in a feed-forward amplification loop, the ping-pong cycle. This pathway is similar in mice, but with some key differences. Primary piRNAs are sense-stranded compared to the TEs and derive from more dispersed regions, while secondary piRNAs are antisense[66]. Two PIWI proteins are present in murine embryonic testes, MILI and MIWI2, but the slicing activity of MIWI2 is not required for piRNA biogenesis, indicating that MILI is sufficient to generate both primary and secondary piRNAs[67]. Secondary piRNA pools are loaded onto MIWI2, which localizes to the nucleus and mediates epigenetic silencing of active transposons[66,68]. Loss of piRNA pools by the deletion of either of these proteins results in a loss of TE epigenetic regulation, increase in TE mobilization, and sterility due to spermatogenic arrest[34,48,49]. The mechanism behind how the PIWI proteins direct epigenetic regulation is currently an area of active study.

Long ncRNA

Another major type of ncRNA are the long noncoding RNAs (lncRNAs): >200nt RNAs that control enhancer activity, epigenetic landscape, and gene expression, all of which can ultimately influence cell fate and development (reviewed in [69]). Some lncRNAs are restricted to work in *cis* from their locus of origin, while others display activity in *trans* on other loci. For instance, the activity of *Xist* during X inactivation is restricted to the future inactive X from which it is transcribed[70]. This local restriction stands in stark contrast to *HOTAIR,* a lncRNA transcribed from the *HOXC* locus which regulates expression from the *HOXD* cluster on a different chromosome[71]. *Xist* and *HOTAIR* also illustrate how lncRNAs can affect local chromatin state by recruiting histone modifiers, as both recruit PRC2 to establish H3K27me3 at their respective loci[70,71]. Positive histone modification is also associated with lncRNA function. *HOTTIP* recruits the KMT2A protein, which subsequently establishes H3K4me3 and induces gene transcription along the *HOXA* locus[72]. Many other chromatin modifying factors have been shown to interact with RNAs, and future RNA precipitation experiments followed by mass spectrometry will likely identify even more.

lncRNAs can also control local DNA methylation; for example, *ecCEBPA* binds DNMT1 and inhibits its activity at the *CEBPA* gene[73]. Sequencing of RNAs associated with DNMT1 revealed that this inhibitory phenomenon was not limited to the *CEBPA* locus—in fact, thousands of RNAs bound DNMT1[73]. These DNMT1-associated RNAs negatively correlated with methylation over their gene bodies and positively correlated

with the expression of the genes themselves, consistent with a mechanism of methylation inhibition.

lncRNAs may also interact with other ncRNA, as seen when circular RNA sponge for miR-7 (ciRS-7) inhibits its miRNA target. The ciRS-7 ncRNA is expressed in the both the mouse and human brain, and has over 60 conserved binding sites for miR-7[74]. This RNA displayed inhibitory capability both *in vitro* and *in vivo*, with injection into zebrafish brain causing reduced midbrain size—the same phenotype that was observed upon morpholino knockdown of miR-7 itself[75]. LincROR has been described as maintaining ESCs in an undifferentiated state via a similar sponge mechanism—sequestering miRNAs away from *Oct4, Nanog,* and other core pluripotency transcripts[76].

Some lncRNAs sequences may have no function at all, requiring transcription only to maintain an open chromatin state. For instance, truncation of *Lockd* to a fraction of its original length induced no phenotypic changes. Knock down of this same transcript, however, resulted in repression of the neighboring *Cdkn1b* gene, likely due to the repression of an enhancer sequence close to the *Lockd* transcription start site[77]. It is possible that many lncRNAs function only to enhance access to *cis* elements, but such conclusions are difficult to reach without unlinking the act of transcription from the transcript itself. Without experiments like RNA truncations, the function of *cis*-acting lncRNAs cannot be certain. Fortunately, this confound only applies to transcripts that

act in *cis*, as *trans*-acting lncRNAs work at a distance and their effects are therefore less likely to be influenced by a local chromatin effect.

How many lncRNAs fall into these different classes is unknown, but over 75,000 high confidence lncRNAs been identified[78], of which only a handful have been studied in depth. This cornucopia of data signals the need for systematic functional studies. One such study conducted in ESC used short hairpin libraries to inhibit expression of 147 lncRNA[79]. They found that 137 of the queried lncRNAs showed significant effects on gene expression, 26 were involved in positively maintaining pluripotency state, and 30 acted as barriers to differentiation. Unfortunately, the expression of many lncRNAs is specific to certain tissues or developmental times, making more comprehensive high throughput characterization of lncRNAs more difficult since they are limited to cell culture models.

DNA methylation

Overview

There are several different types of DNA methylation present in the mammalian system. The most commonly studied DNA modification, 5-methylcytosine (5mC), is normally associated with transcriptional repression. In a process of active demethylation, the Ten-Eleven-Translocase (TET) enzyme sequentially converts 5mC into 5-hydroxymethylcytosine (5hmC)[80], then 5-formylcytosine (5fC), and finally 5-

carboxylcytosine (5caC) before being eliminated by the base excision repair pathway[81,82]. Of these conversion products, only 5hmC is thought to have an independent function because it is specifically recognized by MeCP2[83] and MBD3[84]. N(6)-methyladenine has only been recently discovered in mammals, but it appears to mark young transposons and assist in their silencing[85]. These less common marks are under extensive study, but 5mC is the most relevant mark for my studies.

In mammals, 5mC occurs mostly on CG dinucleotides (CpGs) with the exception of non-CG methylation, which correlates with active transcription in embryonic stem cells[86], but is depleted from transcriptionally active gene in neurons[87]. 60-80% of CpGs are methylated, except at CpGs islands—concentrated, hypomethylated CpGs at promoters of constitutively active genes. Proper placement of 5mC is required for silencing transposable elements, correct pairing of pericentric heterochromatin, and genomic imprinting (reviewed in [88]). It is also controls enhancer activation and can be found along the bodies of actively transcribed genes, possibly influencing splicing[89].

DNA methyltransferases

Three proteins are known to have DNA methyltransferase activity. DNMT1 associates with the replication fork via Np95 and maintains 5mC profiles on newly synthesized DNA that is only methylated on one strand, also known as hemimethylated DNA[90]. DNMT3A and DNMT3B, in contrast, are responsible for *de novo* methyltransferase activity in the cell[91]. They hold non-redundant functions in

suppressing TEs, repeats, and imprinted genes[92]. DNMT3L, a catalytically dead cofactor, is also required to stabilize DNMT3A activity[93]. None of these DNMTs have sequence specificity, so how they are recruited to specific sequences at specific times is poorly understood. Examples of recruitment by proteins or by cross-talk with other epigenetic marks exists (discussed below), but with no comprehensive understanding.

Dynamic 5mC during development

In primordial germ cells (PGCs) and preimplantation embryos the global 5mC profile is reset in preparation for the novel epigenetic profiles required for gametogenesis and embryonic development (reviewed in [88,94]). Demethylation in murine PGCs occurs at ~e7.25, after the specification and migration toward the genital ridge. It is a robust reprogramming from which no sequences escape, with the exceptions of specific TEs[95,96]. Gametic methylation establishment differs depending on the sex: male mice begin at ~e12.5, with full recovery occurring by birth, while females do not begin remethylation until the follicular growth stage after birth. These methylation profiles are critical for fertility, illustrated by the loss of DNMT3L, which results in TE mobilization and loss of the male germline[97]. The next wave of reprogramming occurs after fertilization when the male pronucleus undergoes rapid, active DNA demethylation as the sperm decondenses. This phase is followed by passive, replication-dependent demethylation of the maternal genome copy. Imprinted genes (discussed later) escape this second wave, which maintains their differential methylation profiles. Recovery of methylation begins after implantation and is critical for normal development. Embryonic stem cells, which are derived from the inner cell mass of the early embryo, are able to

retain self-renewal even when all DNA methyltransferases are removed[98]; however, they stop growing upon induction of differentiation, indicating that dynamic methylation is required for viability of differentiated cells. 5mC is likely important for locking in cell fate decisions as well. A DNMT3B null background, for instance, displays ectopic expression of pluripotency factors[99] and cells that are “differentiated” in this background experience an increased rate of reversion to a pluripotent state[100], possibly due to a lack of 5mC stabilization of heterochromatin on the pluripotency factor promoters[101]. X-inactivation is another specific early developmental event which requires DNA methylation. Expression of the ncRNA *Xist* defines the X chromosome which will be repressed, but DNMT3A/B-mediated 5mC of the *Xist* promoter of the active X chromosome is necessary to stop ectopic *Xist* expression[102].

Cross regulation with other epigenetic marks.

Epigenetic marks are not independent of each other, and reports for mechanisms of positive and negative cross-talk between DNA methylation, DNMTs, and other modifications are becoming more common. For example, DNMT1 physically interacts and functionally coordinates with G9a, the H3K9me2 KMT, to maintain facultative heterochromatin during DNA replication[103]. To maintain constitutive heterochromatin, the H3K9me3 machinery like SETDB1, SUV39H1, and HP1β interact with different subsets of DNMTs[104,105]. The interaction between H3K9me3 and DNMTs is conserved in *N. crassa,* where KMT dim-5 was shown to be required for 5mC establishment[18]. These interactions are not limited to heterochromatin either, as DNMT3A recognizes

H3K36me3 and stimulates its DNMT activity, directing 5mC onto actively transcribed gene bodies[106].

Not all of the DNMTs' interactions are cooperative; for instance, the activity of DNMTs and PRC2, the H3K27me3 KMT, are mutually exclusive and antagonistic[107–109]. Global loss of 5mC results in a global upregulation of H3K27me3, while loss of H3K27me3 results in a gain of 5mC at specific domains, like the *Rasgrf1* imprinting control region. DNMT3L has been shown to require an unmodified H3K4 position on nearby nucleosomes for full activity, and is therefore repressed by H3K4 methylation on this residue[110]. This requirement is likely important to maintain distinctions between active and repressed genes, since modifications of this residue are associated with enhancers and transcriptionally active promoters.

DNA methylation can itself interact with chromatin modifiers containing methyl-binding domains to facility transcriptional repression. Several of these methyl-binding domain containing proteins exist, each associated with a plethora of factors that encourage gene repression. MeCP2, for instance, has been shown to bind 5mC and recruit a HDACs[111], which deacetylate adjacent histones and facilitate the spreading of repressive chromatin. It also recruits H3K9 KMT activity[104] and chromatin remodeling factors [112], which are vital for maintenance of 5mC and genomic integrity[113]. It even directly binds DNMT1[114]. The repressive activity of MeCP2 is especially important in the human brain, as its loss is implicated in 96% of cases of Rhett Syndrome[115,116]. In a

manner similar to MeCP2, the MBD1 and MBD2 proteins associate with the repressive nucleosome remodeling deacetylase (NuRD) complex, a complex containing functionally equivalent subunits to those outlined for MeCP2[117]. These interactions between 5mC and other repressive chromatin modifiers assist in stabilizing heterochromatin and normal cell function.

Transposable elements

Transposable elements (TEs) are mobile sequences that hold a paradoxical relationship to their host genome (reviewed in [118]). Much like a parasite, these "selfish" sequences expand their copy number, often to the detriment of their host. On the other hand, TEs also catalyze evolution by accelerating speciation and increased regulatory complexity through increases in splicing[119], non-coding RNA transcription, and movement of regulatory sequences[120]. They are divided into two main classes, DNA transposons and retrotransposons, of which only the latter is still active within mammals. Retrotransposons operate via a "copy and paste" mechanism by reverse transcription and re-insertion into the genome, thereby greatly expanding their copy number. Within this class, elements that contain long terminal repeats (LTRs) such as intracisternal A-particles (IAPs) and murine endogenous retroviruses (MuERVs) are distinct from their non-LTR counterparts such as long interspersed nuclear elements (LINEs) and short interspersed nuclear elements (SINEs) in both structure and mechanism of repression [121]. LTR TEs, for instance, are repressed in ES cells by

H3K9me3 that covers the entire gene[122,123]. In contrast, the youngest and most active LINE elements are regulated by 5mC[124]. It is important that TEs remain tightly regulated, but the loss of 5mC during PGC formation is a particularly vulnerable time. As mentioned previously, when 5mC is reprogrammed, MILI and its slicer activity are required to destroy active TE transcripts[67]. The piRNA pathway can then re-establish 5mC-mediated transcriptional repression during gamete development. Loss of the piRNA machinery or DNMT3L results in a massive TE derepression and meiotic arrest[54,68,97,125]. Regulation by the piRNA pathway is strongly biased toward LINE-1 elements[126], possibly because IAP elements are among the very few elements that are not fully demethylated[96]. Even after 5mC is established, TE regulation is complex. For instance, one study found that TEs were also partially repressed by H3K9me2 during spermatogenesis[127]. Another found that older LINE-1 elements were marked by H3K9me3 in their 5' promoter region after fertilization[124].

Though they are potentially detrimental, TEs are also important for early development. Injection of morpholino-modified antisense oligonucleotides targeting LINE elements in the early mouse zygote causes developmental arrest at the two and four cell stage[128]. Similar injections targeting MuERV-L elements yielded similar arrest phenotypes[129]. In mice, LTRs are cis regulators of nearby zygotic developmental genes, acting as alternative promoters[130]. While no manipulation has been done in human, it has been observed that different subsets of LTR transposons show stage-specific regulation, which can be used as a marker for stage identity[131]. And MuERV-L is a marker for two-cell identity both in mouse embryos and in a subset of ES cells that

transiently cycle to a higher pluripotency state[132]. TEs can also encourage adaptation and evolution through regulatory sequence movement. For instance, many cell-type specific CTCF bound sites are due to novel insertion of SINE elements[133]. The LINE-1 class TE increases activity specifically during neuronal differentiation[134], possibly a mechanism that induces genetic mosaicism and increases plasticity in a terminally differentiated cell line.

Imprinting

Overview

While the majority of genes in the genome are expressed from both alleles, a handful of genes in the plant and mammalian systems undergo parent-of-origin specific, monoallelic expression defined as genomic imprinting. This monoallelic expression is controlled by tight regulation of parental-specific 5mC, making imprinting an excellent model system in which to study *cis*-acting mechanisms (reviewed in [135]). Evidence for imprinting originally came from experiments with Robertsonian translocations in which heterozygous animals missing large sections of chromosomal arms were bred to animals with complementary translocations[136,137]. Some of these crosses were unable to generate viable offspring, arguing that the parent-of-origin for some sections of DNA, rather than the number of alleles, was vital for early development. Further evidence shortly after this discovery came from two other

sources: First, it was found that embryos derived from either two paternal (androgenote) or two maternal (parthogenote) copies of the genome were inviable[138]. Much later, viable parthogenotes were produced, but only with deletions in key imprinted genes that regulated their dosage[139]. Second, genetic disorders in mice[140] and humans[141] were described which depended upon parent-of-origin specific inheritance of a genomic deletion. Angelman Syndrome[141], Beckwith-Wiedemann syndrome[142], and Silver-Russell syndrome[143] are just a few genetic disorders in humans currently understood to be caused by problems with imprinting.

Today there are approximately 150 mouse genes that are regulated in this manner[144], but only a handful have been deeply investigated, including *Igf2*, *Igf2r*, *Snrpn*, *Rasgrf1*, *Gnas*, *Dlk-Dio*, and *Kcnq1*. While the mechanisms regulating these loci are likely different, there are common features which unite them. First, every imprinted gene is directed by an imprinting control region (ICR) which contains a differentially methylated domain (DMD), so named because the same sequence harbors different methylation profiles between males and females. These ICRs are often intergenic sequences far from the gene or cluster of genes that they control. Second, all imprinted genes acquire their differential methylation during gametogenesis on a primary DMD, often with methylation spreading to a secondary DMD after fertilization. Third, imprinted ICRs are able to escape the wave of post-fertilization genome-wide demethylation and maintain differential methylation in adult somatic tissue. Finally, most of these regions contain lncRNAs which affect imprinted expression, either by recruiting epigenetic modifiers or by direct transcriptional inhibition of the adjacent gene. For instance,

paternal *Igf2r* gene expression is silenced by antisense transcription the *Air* noncoding RNA[145,146], while the *Kcnq1ot1* noncoding RNA directly silences paternal expression of multiple genes including *Kcnq1*, *Cdkn1c*, *Slc22a18*, and *Phlda2*[147]

It is important to point note that while mechanisms for establishment of imprinted 5mC in the germline are varied and poorly understood, a robust set of mechanisms for post-fertilization maintenance is currently being uncovered. There are two main mechanisms for this maintenance: Tripartite Motif Containing 28 (TRIM28) or Primordial Germ Cell 7 (PGC7). TRIM28 is an accessory protein that interacts with a variety of proteins which control gene silencing and heterochromatin formation including the embryonic H3K9me3 KMT ESET[148], the NuRD complex[149], and HP1[150]. It's binding partner, Zinc finger protein 57 (Zfp57), recognizes binding sites in maternally and paternally imprinted genes[151] , recruits TRIM28 specifically to the methylated allele, and protects these sequences from demethylation[151–153]. Loss of Zfp57 in ESC resulted in loss of imprinting at all tested ICRs including *Snrpn, KvDMR1, Rasgrf1, Peg3, and Gnas/Nespas*. PCG7, on the other hand, specifically recognizes imprinted genes differentially marked with H3K9me2[154] and directly inhibits the activity of theTet2 and Tet3 demethylases[155]. Loss of PGC7 in oocytes resulted in loss of imprinting in *Peg1*, *Peg3*, *Peg10*, *H19*, and *Rasgrf1* but not *Snrpn*, *Peg5*, or *Meg3*[156]. It is clear that these two mechanisms for 5mC maintenance are not mutually exclusive, as some imprinted genes require both for proper 5mC patterning. It is also possible that other mechanisms might be involved in 5mC maintenance which are yet to be discovered.

Many studies have interrogated the DNA sequences which direct DNA methylation in *cis*, and these will be discussed in Chapter 2. However, I would like to focus on the *Rasgrf1* imprinting system, a tractable model which has been extensively studied for over a decade.

Rasgrf1

The *Rasgrf1* ICR, which consists of a DMD and a set of tandem repeats, is ~30kb upstream of the *Rasgrf1* gene. Methylation at the DMD is lost at ~e12.5 during PGCs demethylation[157], but recovers 5mC via a ncRNA-mediated mechanism specifically in the male germline[158] (Figure 1.1). At approximately e12.5 the repeats induce transcription of a ~3kb ncRNA, dubbed the piRNA-targeted noncoding RNA (pitRNA), which is transcribed across the DMD sequence. The DMD contains an ancient RMER4B retrotransposon sequence, which is subsequently recognized by primary piRNAs derived from chromosome 7 and processed into secondary piRNAs in a manner that is dependent upon MILI. The piRNAs recognize at least two loci within the pitRNA—dubbed sites 1 and 2—and their interactions recruit DNMT activity to the local DMD in a currently undescribed fashion (Figure 1.2). Removal of the repeats results in the loss of pitRNA expression and DMD methylation[158,159]. Insertion of a 242kbp BAC transgene carrying the wild type ICR into repeat-deficient mice yielded normal pitRNA expression and transgene methylation, but no methylation on the endogenous alleles

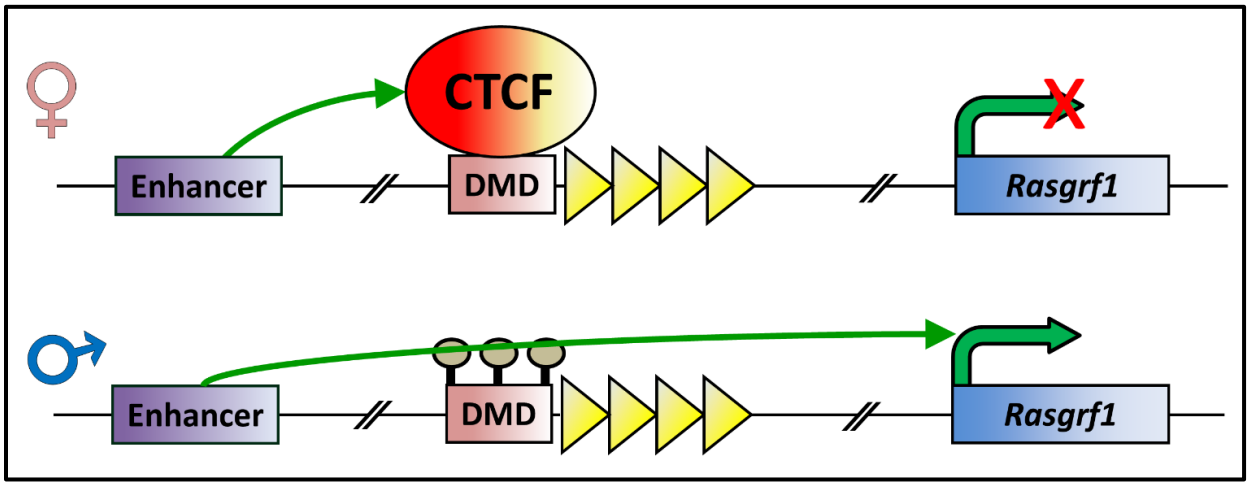

Figure 1.1 - Imprinting control of the *Rasgrf1* locus. The *Rasgrf1* ICR (DMD and repeats) regulates paternal-specific expression of *Rasgrf1* via paternal-specific methylation at the DMD. Differential methylation defines differential binding of CTCF, which subsequently controls access of the enhancer to the *Rasgrf1* gene in an allele-specific way. Yellow triangles are repeats. Closed lollipops are methylated CTCF binding sites.

which lacked pitRNA expression, illustrating a restriction to *cis*-action for the pitRNA mechanism when under the control of the repeat promoter sequence[160].

After fertilization the differential methylation is maintained by the previously described Zfp57 and Stella-mediated mechanisms. Interestingly, the repeats are also necessary to maintain methylation after fertilization but before implantation[161]. Zfp57 binding motifs and H3K9me2 are located in the ICR outside of the repeats[154] (data not shown) and would therefore be present in the repeat deletion animals, arguing that the repeat-mediated maintenance is independent of these other two mechanisms.

The *Rasgrf1* gene is robustly expressed in an imprinted pattern in neonatal brains[162]. It is important for normal body weight and olfactory learning in neonates, with loss of expression resulting in a decrease in both[163,164]. *Rasgrf1* is maternally repressed when CTCF binds enhancer blocking sequences within the unmethylated maternal DMD and abrogates contact between the gene and its enhancer[36]. Paternal methylation of the DMD, however, protects against CTCF binding and results in paternal-specific expression.

To test if the repeats and DMD were sufficient for recruitment of 5mC, these sequences were inserted into a reporter vector and mutant mice were generated by random insertion[160]. The mice displayed robust germline and somatic imprinted methylation profiles when maintained in the same line as the parent of origin; however,

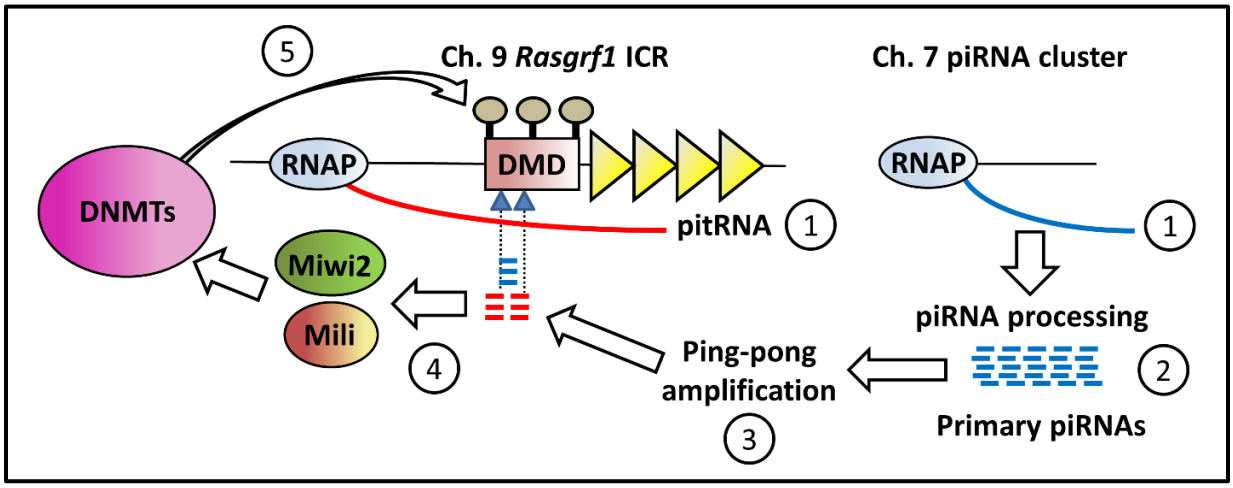

Figure 1.2 - pitRNA-mediated methylation via the piRNA pathway. 1) The pitRNA and piRNA precursor are transcribed from their respective loci. 2) Primary piRNAs are generated from the precursor before 3) ping-pong amplification, which produces secondary piRNAs. 4) PIWI proteins then use piRNA similarity to the pitRNA to recruit 5) DNMTs to methylate the associated DMD. Blue and red transcripts are primary and secondary piRNAs, respectively. Blue triangles are sites 1 and 2.

any pedigree that gained methylation upon paternal transmission of the transgene was never able to reprogram to an unmethylated state upon maternal transmission. These experiments left undefined the sequences which were sufficient for imprinted methylation, a question which I will address in Chapter 2.

The intergenerational R2 mutant

To test if a paternally imprinted gene could be converted into a maternally imprinted one, the maternally methylated ICR from region 2 of *Igf2r* was inserted in place of the repeats at the endogenous *Rasgrf1* ICR[165]. This experiment did not yield a reversal in imprinting status; however, it did result in a strange phenotype in *trans* and intergenerationally. The allele, dubbed *R2*, displayed wild type, paternal-specific imprinting when transmitted through the maternal lineage. When passed through the paternal lineage, however, the allele induced the methylation of the normally unmethylated maternal allele in *trans*. Furthermore, the genetically wild type but aberrantly methylated allele avoided reprogramming when transmitted through the female germline, exhibiting biallelic expression patterns, which indicate biallelic methylation. Some animals even displayed monoallelic maternal expression in the next generation, indicating that the naive, wild type paternal allele was being affected even in the absence of the *R2* allele. These results are reminiscent of other intergenerational effects that are RNA-mediated (discussed below), which provide guidance for possible mechanisms I will be testing in Chapter 3.

RNA induced intergenerational phenotypes

The *R2* mouse is not the only example of an intergenerational epigenetic inheritance (IEI). Here, I will use the definition of IEI espoused by Anne Ferguson-Smith[166] and Robert Martienssen[167]: a short-term epigenetic phenotype which is stable through meiosis, but is not independent of the original insult. For example, in a single exposure to a chemical, a pregnant female, her pups, and their PGCs can all be affected. All three generations may display epigenetic phenotypes, including the third generation derived from the PGCs which are yet to go through meiosis; however, IEI is not stable and can therefore not be detected in the fourth generation. In contrast, transgenerational epigenetic inheritance (TGI) has the added requirement of long term stability. It can be propagated in the absence of the original insult, sometimes indefinitely, and is therefore a rare event which is usually observed in plants. For IEI, however, there are many examples of mutations, chemical exposures, nutritional supplementations/restrictions, and behavioral traits having effects for multiple generations in animal models (reviewed in [167]). Several epidemiologic studies describe similar intergenerational effect in humans where caloric intake before puberty can lead to altered rates of diabetes or cardiovascular mortality in grandchildren[168,169], highlighting the relevance of these phenomena to human health. Mechanistic understandings for many of these phenomena are lacking; however, some productive studies have explored RNA-mediated intergenerational mechanisms. The following examples from both mice and plants in the field of RNA induced intergenerational

effects share many similarities to the aberrant *R2* phenotype seen at *Rasgrf1* and have subsequently provided informative insights for this work.

Paramutation

Paramutation is defined by two critical characteristics: 1) an allele (the paramutagenic allele) must be able to epigenetically influence its homologous allele (the paramutable allele) and 2) the newly paramutated allele must be able to be maintain its altered state through meiosis (reviewed in [170]). A good example of paramutation can be found in the maize b1 locus, consisting of two genetically identical alleles: the B-I and B' epialleles. B-I lacks DNA methylation and is actively transcribed, resulting in increased production of anthocyanin and subsequent dark purple coloration. The B' allele, on the other hand, is methylated and represses transcription of b1, resulting in a green leaf phenotype[171]. In a plant heterozygous for these epialleles, the B' allele can convert B-I into a repressed epigenetic state, denoted as B'*. B'* can subsequently convert other B-I alleles for multiple generations with high penetrance, even in the absence of the initiating B' epiallele. The conversion is mediated by the generation of small RNAs by an RNA-dependent RNA polymerase from a repetitive sequence 100kb upstream of the b1 locus[172]. The *trans* and intergenerational communication between alleles is analogous to the *Rasgrf1* system in three main ways: first, it requires transcription from tandem repeats; second, it requires the generation of small RNAs; and third, the converted allele is stable through meiosis.

miRNA transmission

RNA-mediated intergenerational phenotypes are also found in mice. A lacZ insertion into the c-Kit locus induced an intergenerational white tail tip phenotype, even in the absence of the causative allele. Injection of total RNAs from either the brain or sperm of the mutant animal into naive zygotes phenocopied the mutant, indicating that the effect is RNA-mediated. miRNAs miR-223 and miR-224 that target the Kit allele were injected in an attempt to identify the specific RNA species responsible, but the phenotype on these animals was only partially penetrant, illustrating that a more broad range of factors is required[173]. Similar experiments have also been done injecting miR-124[174], miR-1[175], or miR-19[176] into naïve zygotes, with the resulting phenotypes at least partially reproducing IEI effects of increased body size, cardiac hypertrophy, or obesity, respectively. Injection of RNA from sperm of highly stressed mice induced both behavioral and physiological stress phenotypes in the next generation[177]. DNMT2, an RNA methyltransferase, may also be important for IEI. Injection experiments of miR-223 and miR-124 into a DNMT2 null background did not result in white tails or increased body sizes above background[178], indicating that RNA modification may be important for the IEI effect.

Sperm tRNA

Some studies have done in depth analysis of RNA isolated from sperm and revealed diverse RNA species including miRNA, piRNA, and many more[179,180]. Some have even argued that RNA diversity is important for post-fertilization fitness[181]. In contrast to the miRNA-centered effects discussed previously, more recent mechanistic studies showed that truncated tRNA species in the sperm from animals on high fat[182] or low protein[183] diets can be responsible for metabolic disorder in the next generation.

Thesis overview

The purpose of this thesis is to address how *cis*-acting sequences recruit DNA methylation in the context of imprinting. The following two studies approach from different angles. Chapter 2 will address the question: Sequences have been defined as **necessary** for imprinting establishment at the *Rasgrf1* ICR, but are they actually **sufficient** for 5mC establishment and maintenance? In this study I inserted the relevant sequences from the *Rasgrf1* ICR into an ectopic locus to test if imprinted methylation, and potentially ectopic gene expression, could be reconstituted in a different genomic context. Chapter 3 will address the question: Is the presence of pitRNA in the wrong tissues sufficient to induce aberrant methylation and expression patterns seen in *R2* mice, which produced the *trans* and intergenerational phenotypes described previously? In this study we injected RNA into the early zygote to investigate if the pitRNA is able to

produce a phenotype either locally or globally. The results to both works provide valuable insights into the sequences, and their associated RNA product, which control methylation recruitment.

CHAPTER 2 – IMPRINTED DNA METHYLATION RECONSTITUTED AT A NON-IMPRINTED LOCUS

David H. Taylor, Chelsea McLean, Warren L. Wu, Alex B. Wang, Paul D. Soloway

This work has been submitted for publication.

Abstract

Background

In mammals, tight regulation of cytosine methylation is required for embryonic development and cellular differentiation. The *trans*-acting DNA methyltransferases that catalyze this modification have been identified and characterized; however, these proteins lack sequence specificity, leaving the mechanism of targeting unknown. A *cis*-acting regulator within the *Rasgrf1* imprinting control region (ICR) is necessary for establishment and maintenance of local imprinted methylation. Here we investigate whether 3kb of sequence from the *Rasgrf1* ICR is sufficient to direct appropriate imprinted methylation and target-gene expression patterns when ectopically inserted at the *Wnt1* locus.

Results

The *Rasgrf1* ICR at *Wnt1* lacked somatic methylation when maternally transmitted, and was fully methylated upon paternal transmission, consistent with its

behavior at the *Rasgrf1* locus. It was unmethylated in the female germline and was enriched for methylation in the male germline, though not to the levels seen at the endogenous *Rasgrf1* allele. *Wnt1* expression was not imprinted by our construct, likely due to additional sequences being required for this function.

Conclusion

We have identified sequences that are sufficient for partial establishment and full maintenance of the imprinted DNA methylation patterns. Because full somatic methylation can occur without full gametic methylation, we infer that somatic methylation of the *Rasgrf1* ICR is not simply a consequence of maintained gametic methylation.

Background

Cytosine methylation (5mC) is vital for the regulation of development and other essential processes such as X-chromosome inactivation, genomic imprinting, transposon silencing, and terminal differentiation. 5mC is usually associated with transcriptional repression, but is known to be a complex and dynamically regulated modification. Research into this regulation has included descriptive studies such as methylome mapping[86,184] or functional studies of the proteins that regulate methylation in *trans*, such as DNMT1[185], DNMT3a[186], and DNMT3l[187,188]. Few studies, however, have yielded much information about the *cis*-acting DNA sequences that target these

trans-acting proteins to the DNA in a locus-, tissue-, or time-specific manner. These few informative studies have used imprinted loci, which undergo highly regulated parental-specific, monoallelic methylation and expression. *Cis*-acting regulators of 5mC have been identified at *Igf2r*[189], *Snrpn*[190,191], *H19*[192–196], and *Rasgrf1*[158–160], and in some cases, their mechanisms of action have been elaborated.

The *Rasgrf1* ICR is a two part system consisting of a differentially methylated domain (DMD) and a 2kb 40-mer repetitive element that lies 30kb upstream of the paternally expressed *Rasgrf1* gene[159] (Figure 1.1). The repetitive element, which is required for proper methylation establishment and maintenance after fertilization[159,161], acts as a promoter for a piRNA-targeted noncoding RNA (pitRNA) that is transcribed across the DMD in e16.5 testes[158]. piRNAs normally silence transposable elements in the male germline; however, a subset of these primary piRNAs interact with two loci within the pitRNA, dubbed sites 1 and 2. The pitRNA is subsequently processed to secondary piRNAs, leading to *de novo* methylation of the ICR in a manner that also depends on piRNA pathway components MITOPLD and MILI.

Hypothesizing that repeat-mediated pitRNA transcription and the piRNA binding sites are sufficient to recapitulate imprinting, Park et al were able to show proper sex-specific establishment of germline 5mC in transgenic reporter mice where the ICR was inserted between the γ-globin promoter and its enhancer[160]. Proper somatic maintenance of 5mC was observed in subsequent generations if the construct was

maternally transmitted; however, this construct was not sufficient to maintain an unmethylated maternal DMD in somatic tissue if it passed through the paternal germline. Failure of maintenance was likely due to the inappropriate methylation of a neomycin resistance selectable marker (neo) within the transgene which spread onto the adjacent ICR. In support of this, methylation analysis showed somatic methylation of the transgenic neo cassette upon maternal allele transmission, even if the transgenic DMD was unmethylated. These results confounded efforts to define the minimal sequences sufficient for *Rasgrf1* imprinting.

In order to define the features of the *Rasgrf1* ICR that are sufficient for imprinting control without confounding effects of sequences from other species or ambiguity of transgene insert sites, we targeted the ICR to the non-imprinted *Wnt1* locus in mouse. Here we show that site 1 and the repeats are sufficient to establish wild type somatic imprinted methylation patterns at this ectopic locus. In keeping with all previous studies, the imprinted expression patterns could not be recapitulated, possibly due to the complexity of the local chromatin context or tissue-specific decreases in CTCF binding.

Methods

Generation of transgenic mice

The *Wnt1*^{tm1pds} vector was assembled using genomic clones from *Rasgrf1* and PCR products from *Wnt1*. It included sequence polymorphisms that distinguished the

DMDs from *Wnt1* and *Rasgrf1*, and the transcribed sequences from the wild-type and mutated *Wnt1* alleles. The vector was linearized with ZraI, electroporated into v6.5 ES cells, and placed under 200ug/ml G418 selection. Correct integration was verified by Southern blot after NdeI digestion using both a 5' (amplified by PDS497 5' - GAAGTGGGGCACATCATT and PDS492 5' - CATTTGCACTCTCGCACA) and 3' (amplified by PDS349 5' – AATATGCCTGACGCACCTTC and PDS 350 5' - CACTTCTCTCTGGGCCTCAC) external probes. The neo resistance cassette was removed using transient lipofection of the pCAGGS-flpe-puro plasmid (Addgene #20733). Cells were then microinjected into C2J blastocysts (Jax Stock No: 000058) and subsequently implanted into FVB pseudopregnant females. Germline transmission was verified by an internal PCR (PDS288 5'- TTACCCAGCTTCTCATAGGCGC and PDS1749 5'- CTGCAATTTCTGCCATCATC). Mice were then bred into the C57BL/6 background. The mutated *Wnt1* allele is referred to as *Wnt1DR*.

Swim up assay

Swim up assays were adapted from standard human protocols[197]. Briefly, sperm were isolated from cauda epididymis, washed in 1ml cell culture media supplemented with BSA, and pelleted gently at 300xg for 10 minutes, after which the supernatant was discarded. After gently adding 300ul of fresh media, the pellet was incubated for 60 minutes at 30°C to allow motile sperm to enter the supernatant. The supernatant was then gently separated from the pellet and both samples were processed for DNA methylation analysis.

Methylation

DNA for methylation analysis was bisulfite converted using the Zymo Methylation-Lightning kit (#D5030) and amplified using allele specific PCR for the *Wnt1DR* DMD (PDS405 5'- GTCGTTAAAGATAGTTTAGATATGG and PDS2172-2175 5'- ACAACRAAATACRACAATCACTAATAC) for 40 cycles. Oocyte DNA from 50 oocytes was pooled with salmon sperm as a carrier before conversion. Because of the exceedingly small amount of oocytes template, a nested approach was followed (PDS271 5'- GGAATTTTGGGGATTTTTTAGAGAGTTTATAAAGT and PDS2172-2175 5'- ACAACRAAATACRACAATCACTAATAC) for 15 additional cycles to improve yield. The bisulfite PCR products were purified (Qiagen PCR purification kit #28104), end polished (End-IT kit #ER0720) for 45 minutes, A-tailed (NEB Klenow exo- # M0212L) for 50 minutes, and ligated with TruSeq adapters. This product was subjected to 10 rounds of amplification with barcode-specific primers (PDS2700 5'- AATGATACGGCGACCACCGA and PDS2701 5'- CAAGCAGAAGACGGCATACGA). Ampure beads (Agencourt #A63880) were used to clean up between steps. The ligated product was then gel purified (Qiagen Gel extraction kit # 28704) and quantified via Qubit DNA HS kit (ThermoFisher # Q32851) before pooling and sequencing using Miseq 300nt chemistry. The resulting reads were quality controlled and trimmed, enabling the analysis of at least 1,000 high quality sequences per sample (Summary Table of read results in Additional File 3) using a local installation of the Quantitative Methylation Analysis online software (QUMA http://quma.cdb.riken.jp/).

Expression analysis

cDNA was created via random hexamer reverse transcription from Trizol extracted RNA derived from e9.5 mouse brains. *Wnt1DR* expression was measured by qRT-PCR and SYBR-green using primers specific to a multi nucleotide polymorphism in the 3' UTR of the *Wnt1DR* allele (PDS2037 5'- CTGCCTCCTCATCACTGTGTAAATA and PDS2039 5'- ATAACCGAACGCGCGCGTG). Allele-specific expression was normalized to total *Wnt1* expression (PDS2037 5'- CTGCCTCCTCATCACTGTGTAAATA and PDS2038 5'- CTGGAACCCAGCACAATAAATAGTTT).

Chromatin immunoprecipitation (ChIP)

CTCF ChIP was conducted using standard protocols. Briefly, brains from e9.5 mice carrying the *Wnt1DR* allele, transmitted either maternally or paternally, were finely minced smaller than $1mm^3$, fixed for 10 minutes in a 1.1% formaldehyde solution, quenched with 2.5M glycine for 5 minutes, disaggregated using 30 strokes of a dounce homogenizer, sonicated with the Covaris S2 Acoustic Disrupter (duty cycle = 5%, intensity = 2, cycles/burst = 200, cycle time = 3' ON/ 60" OFF, 3 cycles), and purified using 4ul of CTCF antibody[198] or IgG (Upstate #12-371) on protein A beads (Life Technologies #10003D) with multiple high stringency washes. Mouse embryonic fibroblasts (MEFs) were grown on 150mm plates (~10 million/experiment) and subjected to the same ChIP protocol, except without a dounce homogenization. CTCF binding was verified using qPCR with primers specific to the *H19* ICR as a positive control (PDS2825 5'- ATAGCCAAATCTGCACAGCG and PDS2826 5'-

CATAAGGGTCATGGGGTGGT), an intergenic region as a negative control (PDS2827 5'- AAGAAGCTGCTGAAACACCG and PDS2828 5'- TGCTGGGTGGTACTGGTATG), and at the *Wnt1DR* DMD (PDS2846 5'- CGAAGTTATATCGATAAGCTGCTG and PDS2847 5'- CTACCGCTGCGCTACAACTA).

Results

To test the sufficiency of the *Rasgrf1* ICR to impart imprinting to an ectopic locus, we targeted the repeats and DMD to the *Wnt1* locus. The *Wnt1* gene was chosen for two reasons. First, the gene is haplosufficient, so monoallelic expression induced by imprinting is not lethal[199]. Second, the expression of the gene from e9.5 to e11.5 is completely dependent upon a well-defined enhancer[200,201]; therefore, the placement of methyl-sensitive enhancer blocking sequences between the promoter and enhancer could plausibly control *Wnt1* expression. The allele was constructed as seen in Figure 2.1, positioning the ICR 3' of the *Wnt1* transcribed domain and 5' of the enhancer, which lies 3' of the gene body. The ICR was placed in the same orientation relative to the enhancer and promoter that it assumes at *Rasgrf1*. Because the neomycin resistance cassette, with its high GC enhancer and promoter, could alter regulation of the *Wnt1* locus, we removed it prior to ES cell injections. This allele was then inserted into ES cells by homologous recombination, verified by both Southern blot and PCR (Figure 2.2), and placed into mice by blastocyst injection. We named it the *Wnt1DR* (DR) allele because both the DMD and the repeats were present in the insertion.

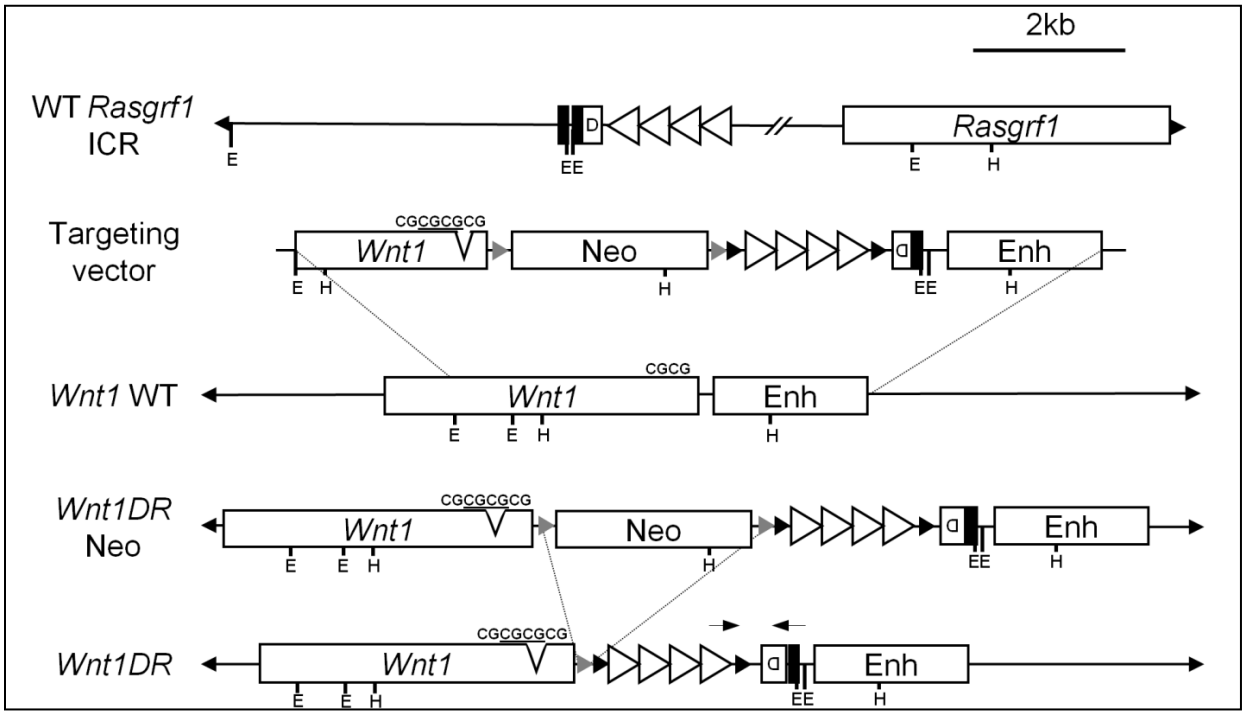

Figure 2.1 – Development of the *Wnt1DR* mutant allele. At the wild type (WT) *Rasgrf1* locus, the ICR includes the repeat region (large white triangles) directly adjacent to the DMD (D), and sites 1 and 2 (vertical black bars) from which secondary piRNAs are processed. The *Wnt1DR* targeting construct contains an FRT-flanked (small grey triangles) neomycin resistance cassette (Neo), loxP-flanked (small black triangle) repeats, the DMD, and a multi-nucleotide polymorphism in the *Wnt1* 3' UTR (divot in UTR, underlined sequence is *Wnt1DR* specific) for allele-specific expression analysis. Primers for detecting *Wnt1DR*-specific methylation are denoted with black arrows. HindIII (H) and EcoRI (E) sites.

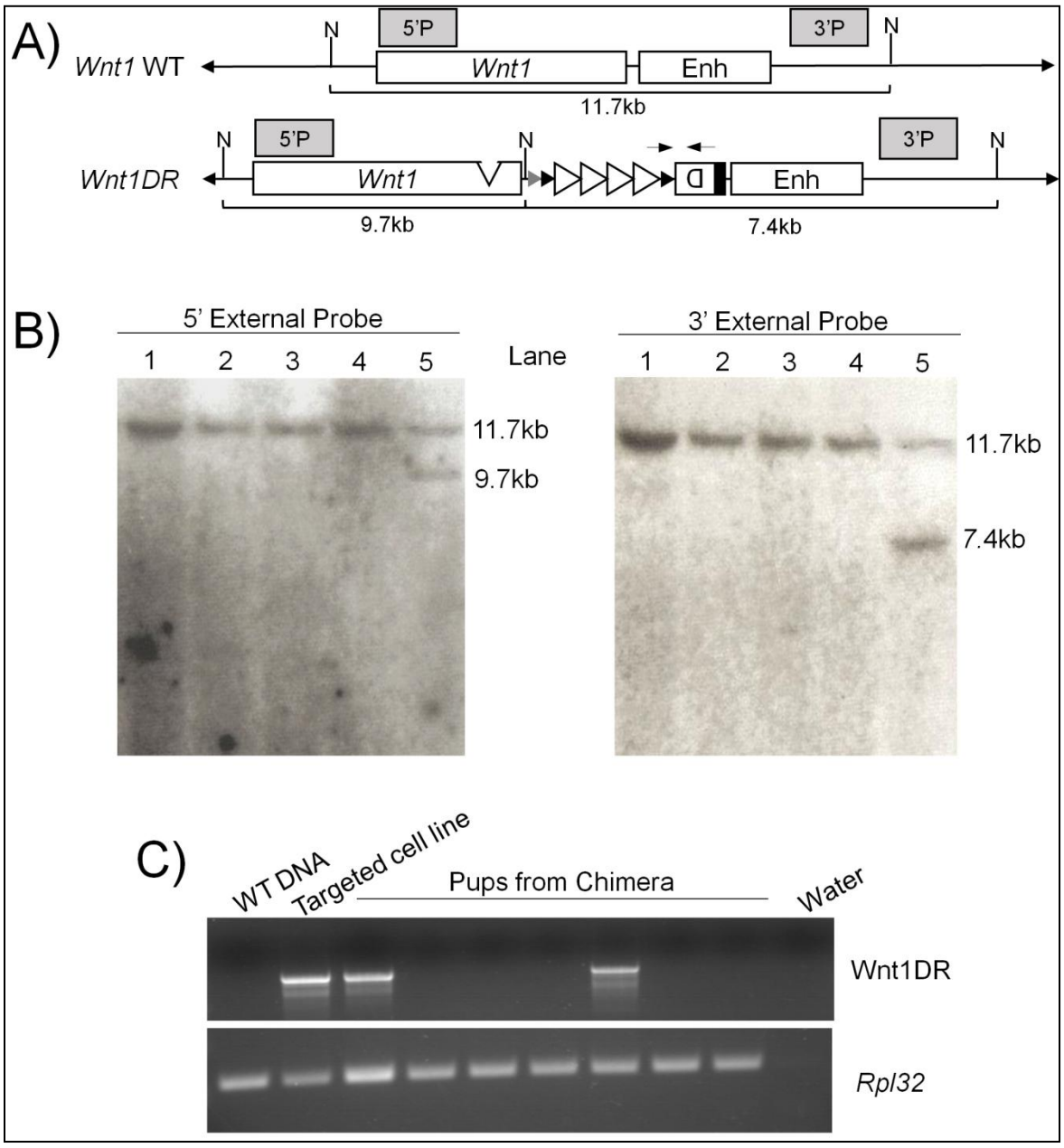

Figure 2.2 – Confirmation of correct integration of *Wnt1DR* construct.
A) Schematic of Southern probes and internal PCR used in panels B and C. B) Southern blots on *NdeI* digested ES cell DNA were probed with either an external 5' probe or an external 3' probe with the expectation of a wild type 11.74kb band and a mutant 9.76kb and 7.47kb, respectively. Example of positive samples in lane 5 of both gels. C) Internal PCR as depicted in panel A was used to follow construct transmission. Rpl32 was used as a loading control.

Since pitRNA expression is necessary for proper DMD methylation in the male germline, we analyzed the e16.5 testis of heterozygous mutant animals. Allele-specific qRT-PCR during paternal transmission revealed that pitRNA expression from the *Wnt1DR* allele is less than 2% that of the endogenous allele within the same animals (Figure 2.3). The presence of the construct did not affect expression levels at the endogenous allele, as no significant difference could be found in wild type pitRNA expression between animals that did or did not carry the allele.

We then established a crossing scheme to fulfill three goals: first, to maintain the allele for six generations of backcrosses through the female lineage; second, to maintain the allele for six generations of backcrosses through the male lineage; and third, to alternate transmission through both maternal and paternal lineages (Figure 2.4C). Analysis of these pedigrees would reveal if the *Wnt1DR* allele establishes and maintains the unmethylated state upon maternal transmission, the methylated state upon paternal transmission, and is properly reprogrammed upon each passage through the opposite sex's germline. To increase throughput for the methylation analysis, we developed a protocol for ligating TruSeq adapters onto the bisulfite PCR products for next generation sequencing. Compared to the standard bisulfite analysis pipeline of cloning followed by Sanger sequencing, this new workflow yielded, on a per sample basis, vastly increased read counts (Table 2.1), decreased cost, and decreased hands-on effort.

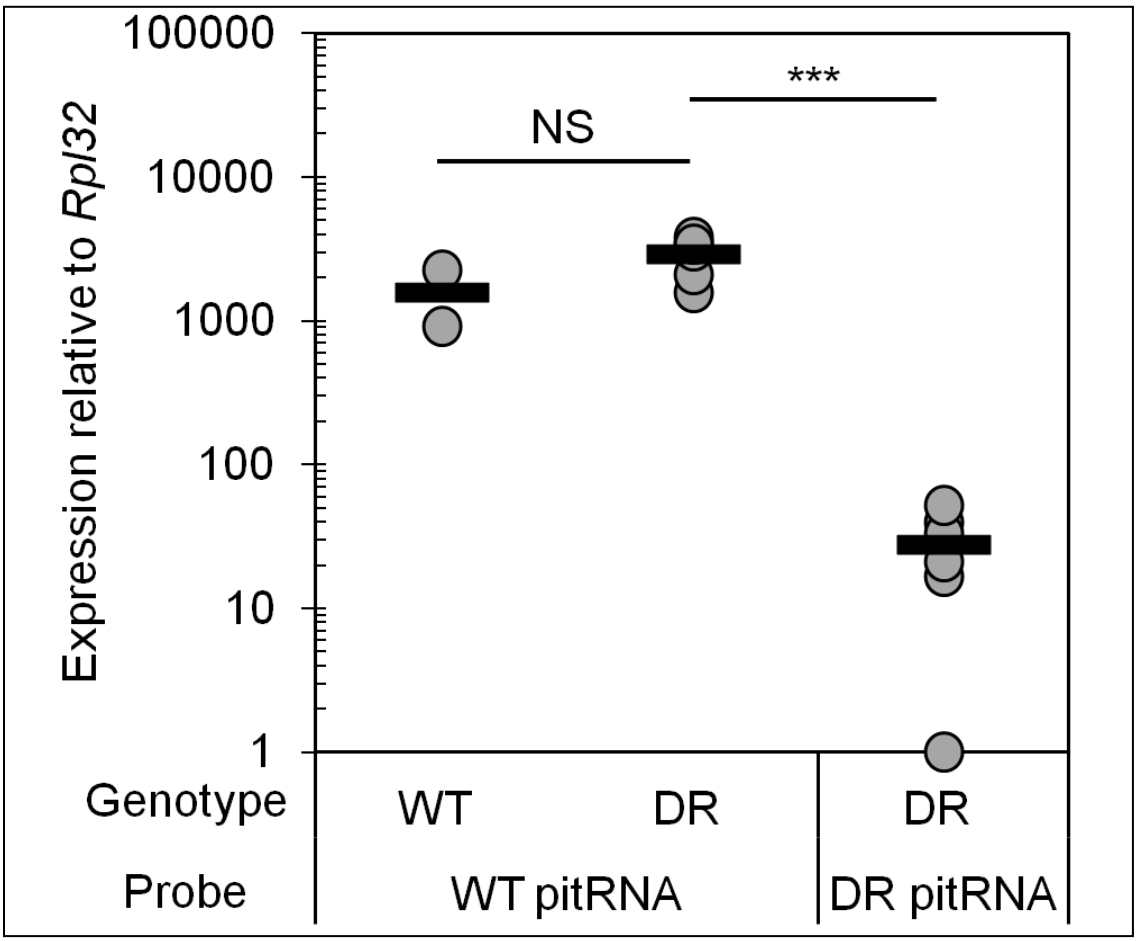

Figure 2.3 – pitRNA expression from the *Wnt1DR* allele is significantly lower than the endogenous allele. Allele-specific qRT-PCR for the *Rasgrf1*-derived pitRNA (WT pitRNA) or the *Wnt1DR*-derived pitRNA (DR pitRNA) using 6 +/DR animals or 2 WT animals. Three asterisks represent $p < 0.001$ using a two tailed t-test, NS denotes not significant, and black horizontal lines indicate the average of relevant samples.

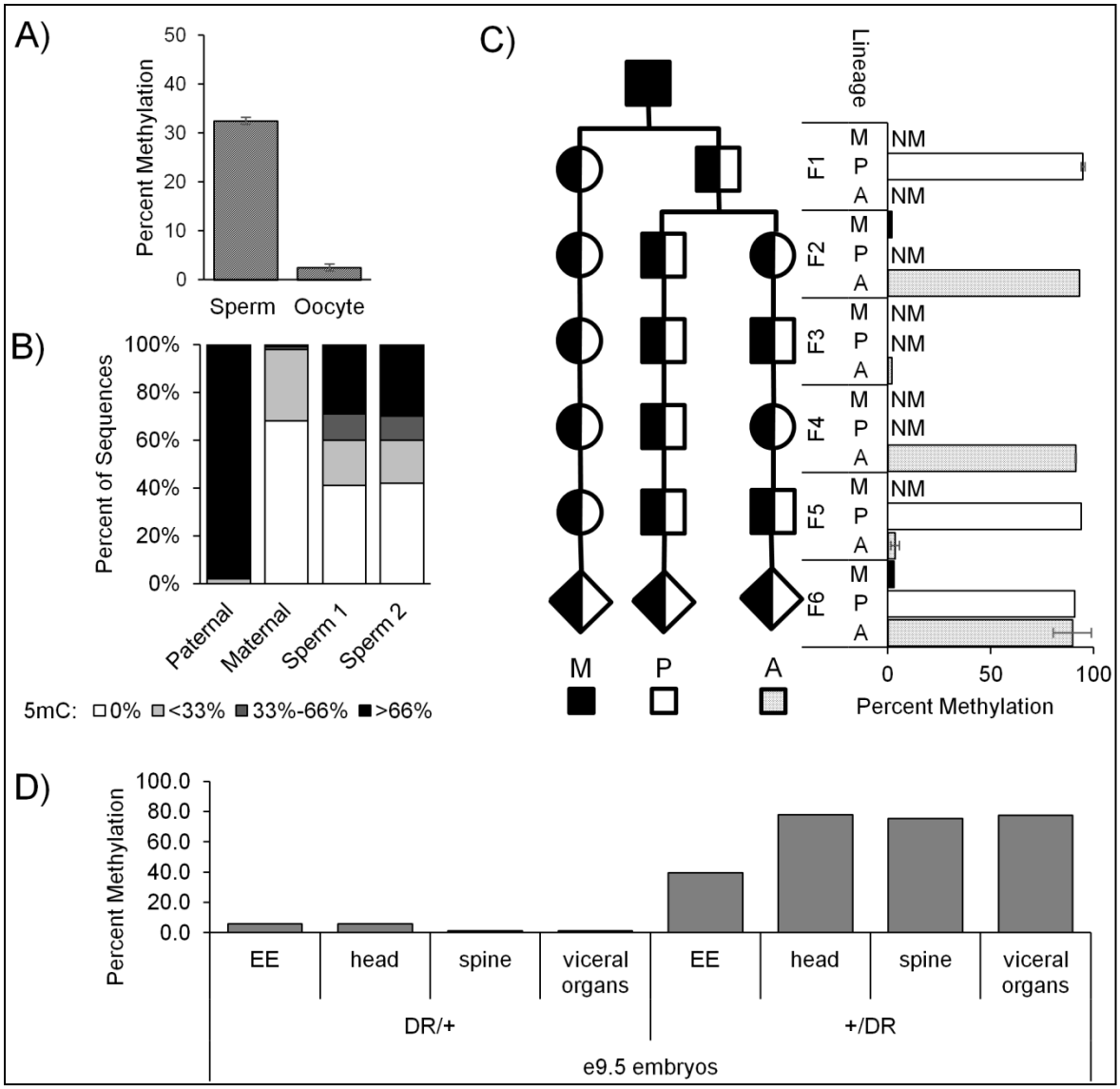

Figure 2.4 – The *Wnt1DR* allele recapitulates partial DNA germline and complete somatic imprinted DNA methylation patterns. A) Methylation levels at *Wnt1DR* in two DR/DR sperm samples and two pools of fifty DR oocytes as assayed by bisulfite sequencing. B) Distribution of methylated CpGs within a paternal lineage F6 animal from panel C, a maternal lineage F6 animal from panel C, and the two sperm samples from panel A. C) Left, mouse pedigrees with maternal (M), paternal (P), or alternating (A) transmission of *Wnt1DR*. Right, DNA methylation levels, as assayed by bisulfite sequencing, using one to four animals sampled from the pedigrees. Circles, females; squares, males; diamonds, either sex; half-filled shapes, heterozygous animals; NM, not measured. D) High throughput bisulfite analysis of extra embryonic (EE) tissue, head, spine, and visceral organs of e9.5 embryos from either maternal (DR/+) or paternal (+/DR) transmission.

Table 2.1 – Breakdown of MiSeq methylation data. Libraries were analyzed in separate runs. In the first run (*), all 6 libraries were placed into <1% of a lane while in the second run (†), each library received ~1% of a lane. Files from the second run were prohibitively large and were therefore truncated to 10MB (~38,653 sequences) before QUMA analysis. PHRED filter: 90% of the read >20. QUMA filter: >98% identity, <10 mismatches, >95% conversion, <5 unconverted CpHs. Supernatant (sup.); animal from Figure 2.4C maternal lineage (M), paternal lineage (P), and alternating lineage (A).

Animal number	Generation and lineage	Total Miseq sequences	Sequences passing PHRED filter	Sequences passing QUMA thresholds
31092*	F1 P	17,518	9,489	4,064
31150†	F1 P	239,206	151,916	3,406
31153†	F1 P	283,512	165,705	2,806
30999†	F1 P	224,213	67,996	2,700
31071*	F2 M	26,941	14,815	7,622
31075*	F2 M	26,334	15,844	9,397
31078†	F2 A	279,920	197,280	2,915
31289†	F3 A	205,607	147,078	3,527
31544†	F3 A	288,652	170,890	1,905
31508†	F4 A	707,055	306,648	7,309
31819†	F4 A	663,137	266,451	11,234
32415†	F5 A	287,049	213,814	3,743
32181†	F5 A	260,495	180,937	1,738
32178†	F5 A	238,913	87,920	3,795
32803*	F6 M	35,683	21,547	13,480
32804*	F6 M	41,468	25,831	12,543
32802*	F6 P	27,718	17,099	7,513
32339†	F6 A	227,530	161,457	3,581
32529†	F6 A	141,201	107,655	6,315
32530†	F6 A	200,173	54,830	3,080
35084†	Sperm	280,699	92,637	2,563
35090†	Sperm	184,452	125,431	4,069
35669 pellet†	Sperm	231,561	175,028	26,901
35669 sup.†	Sperm	155,302	119,684	15,679
35670 pellet†	Sperm	312,711	243,539	25,603
35670 sup.†	Sperm	212,255	167,219	16,270
Oocyte pool 1†	Oocyte	215,808	126,407	1,870
Oocyte pool 2†	Oocyte	341,780	41,574	17,548
E9.5 EE†	+/DR	593,195	490,141	8,735
E9.5 head†	+/DR	579,546	485,625	8,210
E9.5 spine†	+/DR	537,237	451,830	8,521
E9.5 organs†	+/DR	642,148	482,669	7,556
E9.5 EE†	DR/+	859,350	308,544	8,798
E9.5 head†	DR/+	679,635	355,402	7,033
E9.5 spine†	DR/+	461,931	350,100	7,898
E9.5 organs†	DR/+	486,448	393,705	7,231

To investigate correct establishment of DNA methylation at the *Wnt1DR* allele, we analyzed sperm and oocytes from mutant animals and observed significantly higher 5mC in sperm compared to oocytes (Figure 2.4A). Interestingly, at the *Wnt1DR* allele overall sperm methylation is at 30%, in stark contrast to the 100% methylation levels seen at the endogenous *Rasgrf1* locus. Methylation in sperm was not evenly distributed, with the 41% of reads containing no 5mC, 29% of the reads showing 66% 5mC or higher, and the remaining reads displaying intermediate levels (Figure 2.4B and Figure 2.5). Healthy, motile sperm were isolated by a swim up assay to ensure that motility effects were not confounding our methylation analysis. The pellet and supernatant from this assay, containing non-motile and motile sperm, respectively, yielded similar results as whole sperm isolation (Figure 2.5)—namely that both samples contained a large fraction of completely unmethylated sequences. Incomplete establishment of methylation in sperm is likely due to significantly reduced expression of pitRNA (Figure 2.3).

Analysis of methylation maintenance at the *Wnt1DR* mutant allele in adult tail DNA showed full recapitulation of both the patterns seen at the wild type *Rasgrf1* ICR. The *Wnt1DR* allele preserved its unmethylated state in the soma during five generations of passage through the female germline (Figure 2.4C). Transmission of the *Wnt1DR* allele through the male germline, despite low pitRNA expression and incompletely established germline methylation, displayed a fully methylated state in the soma throughout six generations of passage (Figure 2.4C). In depth analysis of an alternating

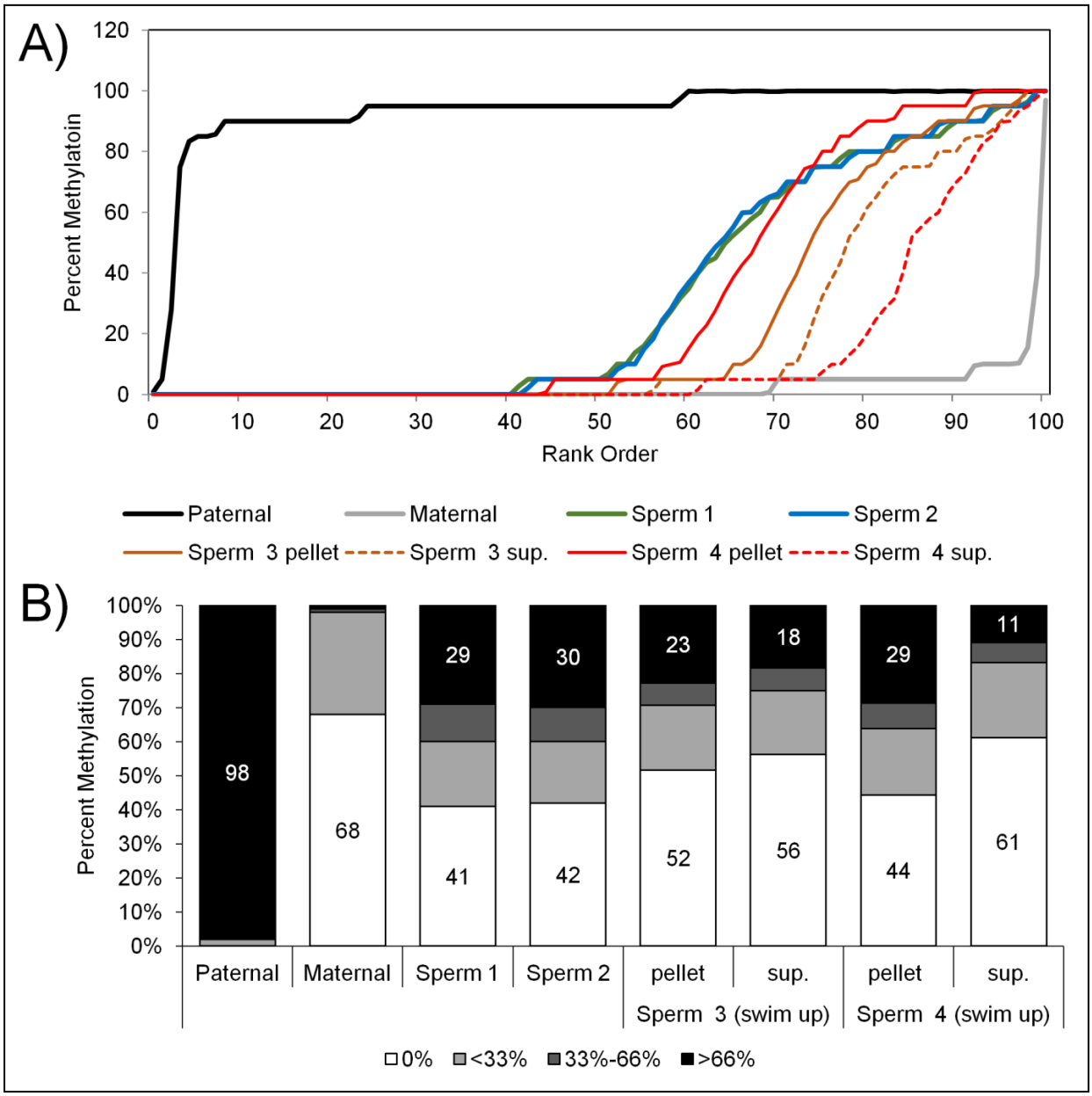

Figure 2.5 - Distribution of methylation levels in sperm DNA. Rank order analysis (A) and bar charts (B) were used to give a more detailed view of the distribution of methylation levels within Sperm 1 and Sperm 2 samples described in Figure 2.4B. Sperm 3 and Sperm 4 are derived from swim up assays that separate motile (sup.) from non-motile (pellet) sperm.

lineage illustrated that the methylation profile could be faithfully reset at multiple different generations (Figure 2.4C). These data demonstrate that the ~3kb section of the ICR is sufficient for partial establishment and full maintenance of the 5mC imprint. We extended these analyses to query methylation states of e9.5 embryos, the stage when the *Wnt1* locus is expressed. We found the same pattern when analyzing DNA of four different regions—extra embryonic tissue, head, spine, and visceral organs—namely that the *Wnt1DR* allele was significantly more methylated after paternal transmission than maternal transmission in each tissue (Figure 2.4D).

Somatic regulation of 5mC at the endogenous *Rasgrf1* DMD is required to control the methylation-sensitive binding of CTCF. Unmethylated sequences permit CTCF binding and enhancer blocking function to silence transcription from the maternal allele [36]We therefore assayed expression of *Wnt1* from our construct to determine if the faithfully imprinted 5mC at *Wnt1DR* enabled allele-specific expression. Our design of the *Wnt1DR* allele included placement of a multi-nucleotide polymorphism in the 3' UTR to facilitate allele-specific qRT-PCR analysis. Contrary to expectation, we found that maternal transmission of the allele did not repress *Wnt1* (Figure 2.6A). Interestingly, we found that the paternal allele was slightly repressed, possibly due to the proximity of the *Wnt1* gene and its enhancer to the highly methylated ICR.

There are two explanations for the inability to recapitulate imprinted expression. First, CTCF might not be bound to our mutant construct, possibly because the binding

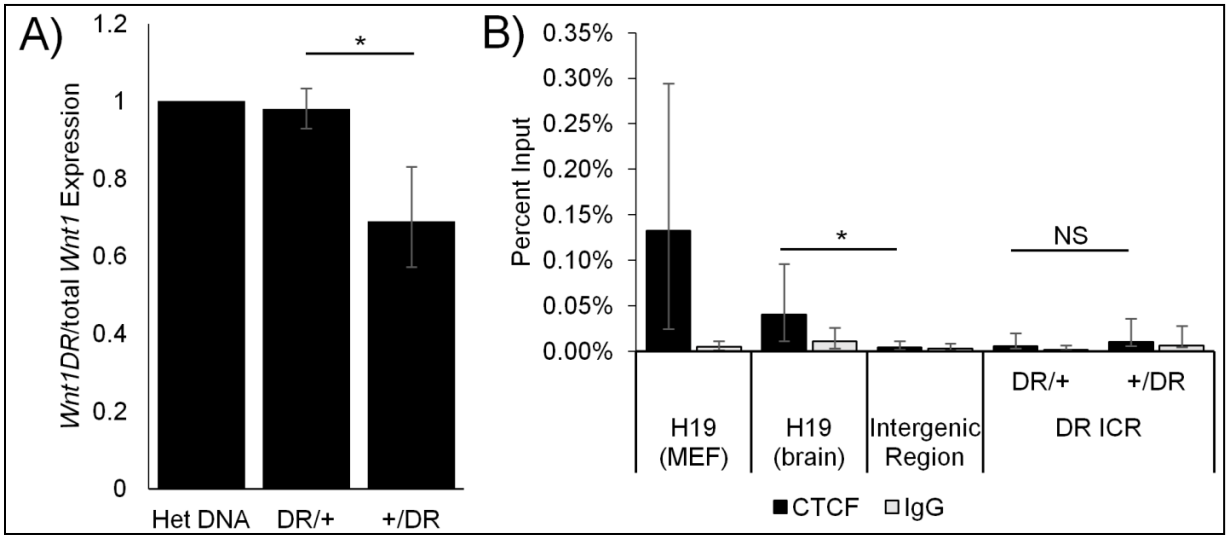

Figure 2.6 – Maternal transmission of *Wnt1DR* allele does not cause silencing or CTCF binding. A) Allele-specific qRT-PCR using e9.5 brains taken from four mice with maternal (DR/+) or paternal (+/DR) transmission of *Wnt1DR*. Data are presented as the ratio of *Wnt1DR* allele-specific expression to total *Wnt1* expression. Heterozygous DNA (Het DNA) was used as a calibrator for the 1:2 copy number of *Wnt1DR* to total *Wnt1*. Asterisk signifies $p < 0.05$ using a two tailed t-test. B) Allele-specific ChIP-qPCR querying an *H19* positive control, an intergenic region containing no CTCF sites, and the *Wnt1DR* allele. e9.5 brains from both genotypes (three biological replicates each) were used as the source of chromatin for ChIP with anti-CTCF and non-immune IgG. Assay of CTCF binding to *H19* in MEF tissue added as additional positive control. Asterisk represents $p < 0.05$ using two tailed t-test after accounting for multiple hypothesis testing using Bonferroni correction. NS, not significant. Error bars denote standard deviation with error propagation.

sites within the ICR are weak and might need external sequences for proper recruitment. CTCF has been shown to work in pairs to change the three dimensional structure of the DNA, so a second possibility might be that CTCF is correctly localized, but cannot function without interacting partner sites that can sequester the enhancer, as occurs at the globin locus control region[39]. To determine if CTCF is bound to the unmethylated, maternally-transmitted *Wnt1DR* allele, we performed ChIP-qPCR which revealed that while there was significant binding at the positive control (*H19* ICR) in both mouse embryonic fibroblasts (MEFs) and e9.5 brains, there is no significant CTCF binding to the negative control or to the *Wnt1DR* DMD (Figure 2.6B) in e9.5 brains. Therefore, we conclude that the *Wnt1DR* allele either lacks CTCF-recruiting sequences that were present at the endogenous locus, that the novel chromosomal context at the *Wnt1* locus acts to negatively regulate CTCF binding, or that the decreased binding of CTCF in e9.5 brain compared to MEFs translates to the *Wnt1DR* ICR, reducing potential binding below detectable levels.

Discussion

Temporally controlled, tissue-specific targeting of DNA methylation to specific sequences is required for normal fertility and health in mammals. Very few systems have identified sequences which are sufficient to recruit epigenetic effectors of 5mC, and even those sequences appear to act in a variety of mechanisms. Here we investigate the *Rasgrf1* imprinted region and define a 3kb portion of the ICR that is

sufficient to recapitulate many features of *Rasgrf1* imprinting when exported to the ectopic *Wnt1* locus.

In oocytes, only 2.5% of the CpGs in 19,418 bisulfite reads analyzed harbored 5mC at *Wnt1DR*, consistent with the lack of 5mC in oocytes reported at *Rasgrf1*. Similarly, upon maternal transmission of *Wnt1DR*, only 2.6% of the CpGs in 77,031 bisulfite reads were methylated, also consistent with the methylation state at the maternal copy of *Rasgrf1.*

The pitRNA, whose expression in e16.5 testes is necessary in *cis* for imprinted DNA methylation at *Rasgrf1*[158,160]*,* is expressed from *Wnt1DR*. However, its expression level is only 2% that seen from *Rasgrf1*. This weak expression may be due to undefined sequences at *Rasgrf1* that are necessary for full expression are absent from the *Wnt1DR* allele; alternatively, the necessary sequences may be present, but the chromatin context at *Wnt1* may not be permissive for their full activity. A previously published allele[160] contained fewer sequences from the endogenous ICR compared to the *Wnt1DR* construct, but still established robust sperm methylation, arguing for the latter hypothesis. Despite the low level of pitRNA expression, it was sufficient to establish methylation in the male germline at an average of 30% of the CpGs assayed at *Wnt1DR*, based on analysis of 6,632 bisulfite sequencing reads from sperm DNA. Of those reads, 41% showed no *Wnt1DR* methylation, and an equal percentage showed methylation levels between 33 and 100%. A similar pattern was revealed upon

separation of sperm by motility, eliminating the possibility that sperm were unmethylated merely because they were dead or inactivated.

Intriguingly, despite the abundance of unmethylated and partially methylated sperm, upon paternal transmission of *Wnt1DR* somatic methylation was nearly complete in twelve mice tested, with 92% of CpGs assayed in 54,923 bisulfite reads showing methylation. This result suggests that the modest levels of pitRNA expression and the partial establishment of male germline methylation are sufficient for full somatic methylation. Because we cannot know both the methylation state of a given sperm and the somatic methylation of the mouse resulting from its fertilization, it is possible that the twelve progeny assayed after paternal transmission of *Wnt1DR* arose only from richly methylated sperm. However, assuming fertilization fitness among sperm does not depend upon *Wnt1DR* methylation state, there is less than a 0.2% probability that of the twelve animals assayed for somatic methylation of their paternally transmitted *Wnt1DR* allele, none arose from fertilization by one of the 41% of sperm inferred to have an unmethylated DMD. For animals arising from fertilization by those sperm, the somatic methylation of *Wnt1DR* must have occurred by methylation spreading from the repeats, sperm-transmitted pitRNA, or another unknown mechanism. Because of their repetitive nature, the repeats are difficult to query; however, they are unlikely to exhibit a methylation profile which differs from the DMD because our assay spans the area between the repeats and pitRNA sites 1 and 2—the sequences which recruit DNA methylation. Sperm-transmitted RNAs are also unlikely, as piRNAs appear to be

depleted in mature spermatozoa[55,179]. It is formally possible, however, that a limited number of piRNAs and their effector proteins are still active after fertilization.

Another mechanism to explain the discordant methylation between sperm and soma may be allele-specific histone modifications like H3K9me2[154], H3R3me2[202], H3K9me3[203], or H4K20me2[203]; rather than 5mC, these could represent the primary imprinting mark, which subsequently enables somatic DNA methylation. Indeed, H3K9me2 is present on the *Rasgrf1* ICR in sperm and has already been shown to be vital for imprinting maintenance; it recruits PGC7 which subsequently protects against Tet-mediated demethylation of the paternal allele after[154,156]. This mark could also be sufficient to recruit *de novo* DNA methylation to the paternal allele after fertilization in the absence of a pre-existing methylation. Caution must be advised when interpreting histone immunoprecipitation experiments in sperm; however, as histones are switched for protamines in post-meiotic spermatids, leading to low total histone occupancy and inflated signal from rare epigenetic marks, compared to tissues lacking protamine substitutions. Analysis of the *Rasgrf1* ICR in particular revealed a depletion of nucleosomes, potentially complicating interpretation[204]. Our finding that somatic methylation of *Wnt1DR* may not depend on prior methylation establishment is consistent with reports of knock-in alleles of the *H19* ICR at *Afp*[195] and *CD3*[193], which also displayed incomplete germline methylation followed by complete somatic methylation. It is also consistent with the work examining methylation acquisition of *Igf2r* ICR sequences, which was done by injecting DNA into post-fertilization embryos[189], not into oocytes where DNA methylation establishment normally occurs[205].

Finally, the possibility that somatic 5mC can arise in the absence of previously established 5mC in sperm is consistent with our previous findings that sperm methylation is not sufficient for somatic maintenance of that mark[161].

A previous transgenic model to test the sufficiency of the *Rasgrf1* for control of imprinted DNA methylation failed to exhibit reversible somatic methylation upon sequential passage through the male and female germlines[160]. Analysis of that model was confounded by persistent methylation of a neo reporter cassette. In contrast, the *Wnt1DR* allele exhibited the expected methylation reversal across six generations, demonstrating the sufficiency for somatic methylation imprinting of the tested sequences.

Although somatic methylation of the *Wnt1DR* allele faithfully recapitulated what was observed on the endogenous *Rasgrf1* allele, this was not sufficient to impart imprinted expression to *Wnt1*. Imprinted expression at *Rasgrf1* requires silencing of the maternal allele by CTCF binding to the DMD, which limits activity of the maternal promoter. Methylation at the paternal allele prevents CTCF binding, and enhancer-promoter interactions are unrestricted. Its binding to *Wnt1DR* was undetectable above background, which might account for the failure of the unmethylated maternal *Wnt1DR* allele to be silenced. Lack of CTCF binding is likely due to its complex role in mediating the formation of three-dimensional chromatin structures with multiple interacting CTCF binding sites[42,206–208]. The local chromosomal context at *Wnt1* may lack appropriate

partner sites for the *Wnt1DR* ICR and thereby discourage this complex binding. Binding to the positive control in MEF samples was shown to be much higher than embryonic brain, indicating the possibility of lower CTCF binding in embryonic brain tissue in general. If true, then this decreased binding might be sufficient to explain lack of binding to the *Wnt1DR* ICR as well.

Conclusion

We have identified sequences that are sufficient for partial germline, and full somatic 5mC imprinting at an ectopic locus. *Rasgrf1* is one of the few loci whose *cis*-acting sequences have been so narrowly defined. Discordance between the levels of *Wnt1DR* methylation in sperm and somatic tissue after paternal transmission are consistent with prior findings that post-fertilization methylation is not merely a maintenance of gametic methylation; indicating the existence of distinct post-fertilization methylation mechanisms.

Authors' Contributions

DHT, CM, and PDS designed experiments. CM generated *Wnt1DR* ESC and performed the Southern Blot. DHT, WLW, and ABW performed all other experiments. DHT and PDS wrote the manuscript.

Acknowledgements

We thank Erin Chu and Roman Spektor for their stimulating conversations that benefited both the science and aesthetics of this paper. We thank the Cornell Stem Cell and Transgenic Core Facility for making the *Wnt1DR* mice and the Cornell University Biotechnology Resource Center (BRC) for their assistance in high throughput sequencing. Finally, we thank Victor Lobanenkov for the CTCF antibody.

CHAPTER 3 – CHARACTERIZING THE EFFECT OF ZYGOTIC RNA INJECTION

David H. Taylor, John McElwee, Richard He, Kevin Wei, Daniel Zinshteyn, Paul D. Soloway

Abstract

Tight control of epigenetic modifications is required for normal cell development, with epigenetic misregulation associated with a wide variety of aberrant molecular phenotypes which can result in disease. Some phenotypes can be transmitted across generations in a phenomenon known as intergenerational epigenetic inheritance (IEI), the mechanism of which is unclear. At the wild type *Rasgrf1* imprinting control region (ICR), paternal-specific DNA methylation (5mC) is regulated by expression of *cis*-acting ncRNA known as the pitRNA; however, previous work with the *Rasgrf1*$^{tm3.1pds}$ mutant animal revealed aberrant *trans*-allelic 5mC and IEI phenotypes in neonatal brain. Analysis of pitRNA expression from this mutant animal revealed misregulation both spatially and temporally, leading to the hypothesis that these effects may be due to aberrant RNA transmission to the zygote. In this study we look at the role of zygotic RNA by injecting the pitRNA, a control RNA, or buffer into naive zygotes and assay 5mC and gene expression phenotypes in the neonatal tissue. While we demonstrate that the aberrant phenotypes observed in *Rasgrf1*$^{tm3.1pds}$ animals could not be

reproduced, we do describe three surprising molecular phenotypes. Most striking is the repression of over 10,000 genes which is dependent upon the RNA species injected. The second two phenotypes are independent of RNA species injected, namely the modest upregulation of transposable elements and a global downward trend in non-TE genes. These results highlight the importance of zygotic RNAs, which can have lasting effects throughout development.

Background

Epigenetics is the study of heritable changes in gene expression that are not encoded in the DNA. Sometimes these alterations are propagated through meiosis and, in a phenomenon known as intergenerational epigenetic inheritance (IEI), can influence one or more generations. Though much more stable in plants (reviewed in[209]), mammals have shown IEI in response to many stimuli including changes to diet[210], maternal care[211], and exposure to toxins[212]. Mechanisms controlling the transmission of these phenotypes are currently unknown, but reports of IEI induced via RNA transmission have increased in the last decade, all of them with strikingly similar conclusions: 1) that parental mutations can induce phenotypes in genetically wild type progeny and 2) that transmission of aberrant pools of RNA into naïve zygotes is sufficient to phenocopy the original mutant phenotype. Indeed, injection of mixed RNAs derived from mutant tissues into wild type zygotes can result in increased body size[174], cardiac hypertrophy[175], obesity[176], stress responses[177], or a white tail tip[173]—all consistent with the phenotypes of the mutant animals. Specific RNA species like miRNA

or truncated tRNA were able to partially explain some of these effects. These studies prompted our lab to revisit a previously reported example of IEI at the *Rasgrf1* imprinted locus[165] to investigate possible RNA-mediated mechanisms.

The *Rasgrf1* imprinting control region (ICR) is a set of sequences 30kb upstream of the *Rasgrf1* coding sequence that regulates expression. Within the ICR, paternal-specific 5mC is located on the differentially methylated domain (DMD) and is controlled by an adjacent set of tandem repeats[159]. These repeats act as a promoter for a noncoding RNA (pitRNA) which enlists the piRNA machinery and their downstream 5mC activity specifically in the male germline[158]. The pitRNA and repeats have been shown to work in *cis*, as the presence of a transgenic, wild-type copy of the locus is unable to rescue endogenous sequences lacking the repeats[160]. A previous study, however, described the *Rasgrf1*$^{tm3.1pds}$ (*R2*) mutant animal, which was able to recruit methylation both in *trans* and intergenerationally[165]. The *R2* animal was generated by replacing the *cis*-acting repeat region with the “region 2” *cis*-control element of the maternally methylated *Igf2r* imprinted gene. The *R2* allele maintained an unmethylated state when transmitted through the maternal lineage; however, upon paternal transmission the *R2* allele not only recruited 5mC in *cis*, but also imparted 5mC in *trans* onto the genetically wild type, normally unmethylated maternal allele—resulting in biallelic expression of *Rasgrf1*. Surprisingly, aberrant methylation of this maternal allele also displayed IEI, maintaining its methylation to the next generation in the absence of the *R2* mutation.

If the wild type *Rasgrf1* ICR is controlled by RNA, and other instances of IEI are also controlled by RNA, then perhaps the aberrant methylation phenotypes found in the *R2* mutant animal are RNA-mediated as well. Multiple examples of such RNA-directed epigenetic control exist in the literature. In plants, RNA polymerase V uses siRNA generated by RNA polymerase IV to silence transposons and repetitive DNA in a process known as RNA-directed DNA methylation (RdDM)[213]. In *S. pombe*, siRNAs loaded onto Ago1 recruit the RNA induced transcriptional silencing complex (RITSC) to pericentromeric heterochromatic regions where they deposit H3K9me[214]. Though the mechanisms have not been thoroughly elucidated, in *C. elegans*[215,216], *D. melanogaster*[217], and *M. musculus*[68,125] piRNAs which bear similarity to transposable elements are loaded onto Ago-clade PIWI and direct epigenetic silencing via H3K9me or 5mC. In the *R2* mutant animal, two lines of evidence support the idea that the aberrant 5mC phenotypes are RNA-mediated. First, region 2 is a promoter for the *Air* noncoding RNA at the *Igf2r* ICR[145], highlighting its plausible role in misregulation of the pitRNA when inserted at the *Rasgrf1* ICR. Second, RT-PCR analysis confirmed pitRNA overexpression in both testes and ovaries of *R2* animals (Figure 3.1). Thus, we hypothesize that the *R2* promoter is responsible for aberrant transmission of the pitRNA to the next generation, resulting in the *trans* and IEI phenotypes. In this study we injected pitRNA, control RNA, and buffer into naive zygotes to investigate if zygotic RNA load can phenocopy *R2* mutant animals. We show that while RNA injection does not lead to inappropriate activation of the silent maternal *Rasgrf1* allele, it does result in significant global gene expression changes in the neonatal brain, some of which depend on the identity of the injected RNA while others are sequence independent.

Materials/Methods

Zygote injection

The pDHT5.1 (pitRNA) plasmid was generated by PCR amplification of the endogenous *Rasgrf1* DMD (PDS1934 5'- CTGTCTTTCCTCCGTATTCCAC and PDS1936 5'- AACGGTAGTGCAGCAGCAG) followed by cloning into the pC*R2*.1 backbone using the TOPO-TA cloning kit (ThermoFisher K4500-01SC). pDHT5.1 and pC*R2*.1 were linearized using HindIII and BglII, respectively, subjected to *in vitro* transcription using the HiScribe T7 Maxi Kit (E2040S), DNAse I treated (Promega M6101) for 30 minutes at 37°C, RNeasy column purified (Qiagen 74104), and eluted in sterile ddH20 for high concentration. Expected 3kb transcript lengths were validated by the Agilent 2100 Bioanalyzer provided by the Cornell University Biotechnology Resource Center (BRC). RNA was then diluted with sterile dd$H_2$0 and sterile 2x Injection Buffer (20mM Tris-HCL, pH 7.5, 0.2mM EDTA, 200mM NaCl) for a final concentration of ~400ng/ul (~200fmol) in 1x Injection Buffer. FVB females were superovulated and bred to YJC6 males[161] overnight. After 16 hours, hybrid zygotes were isolated from plugged females in a sterile environment and given, along with either RNA solution or empty 1x Injection Buffer, to the Cornell Stem Cell and Transgenic Core Facility. After injection with ~5pL of RNA solution or 1x Injection Buffer, zygotes were allowed to rest overnight in KSOM (Millipore MR-121-D) at 37°C and 5% CO2. After 16 hours, zygotes which proceeded to the two-cell stage were implanted into psuedopregnant FVBs and allowed to develop to term. Day 1 (P1) pup brain was taken

for RNA and a hind limb was taken for DNA. Tissues from FVBxYJC6 pups that did not undergo ex vivo manipulation (uninjected animals) were included as controls.

RNA analysis

cDNA was created via random hexamer reverse transcription from Trizol-extracted RNA derived from P1 brain mouse brains. Exon-spanning primers (PDS2128 5'- TGATGTCACTGGTCCATGCT and PDS2130 5'- GTTCCCTCATGCAGATGTCC) were used to amplify the *Rasgrf1* coding region. RT-PCR products were digested with AciI, which specifically cuts the FVB allele, so allelic contribution could be discriminated. LINE-1 (PDS2462 5'- AAGAAATTCCTCCCGACACA and PDS2463 5'- TGGCTTTCATAGTCTCTGGTGA), IAP (PDS2494 5'- CGGGTCGCGGTAATAAAGGT and PDS2495 5'- ACTCTCGTTCCCCAGCTGAA), and MuERV-L (PDS2492 5'- GGTGGCAGGTTGACTACGTT and PDS2493 5'- CTGTCAGTCCGTGAATGGTG) expression were also assessed via SYBR-green qPCR using Rpl32 (PDS72 5'- CATGCACACAAGCCATCTACTCA and PDS73 5'- TGCTCACAATGTGTCCTCTAAGAAC) as a reference.

RNA-seq

Strand-specific, polyA-selected libraries were prepared by an external company (polargenomics.com) and sequenced on the Nextseq 75bp platform. Galaxy (www.usegalaxy.org) was used for the workflow until the mapping. Quality was assessed by FastQC, and Tophat used for alignment to the unmasked mm9 genome,

allowing for a maximum of 100 alignments for each read. TEtranscripts[218] was used for mapping. Count tables were retrieved and DESeq2 was used for differential gene analysis and data visualization. Gene ontology analysis was done using Panther (http://pantherdb.org/). Raw sequencing results are summarized in Table 3.2. A Kolmogorov-Smirnov test was used to compare the distribution of all TEs to the distribution of all non-TEs.

Results

RT-qPCR analysis revealed significantly increased pitRNA transcription in the adult testis and ovaries from *R2* mutant animals compared to wild type animals (Figure 3.1). Robust pitRNA expression is normally only found in embryonic testis[158], illustrating how pitRNA under control of the *Igf2r* ICR is misregulated both spatially and temporally. This result also fits with the model that aberrant pitRNA expression in ovaries might be loaded into genetically wild type zygotes, potentially influencing methylation in the next generation independently of the *R2* allele.

To test this hypothesis, we injected naive zygotes with buffer, pitRNA, or with a non-mouse RNA (TOPO) created via antisense *in vitro* transcription from the empty p*CR2*.1 cloning vector. Animals from unmanipulated matings were also included to control for *ex vivo* manipulations. Brain RNA from one-day-old neonates (P1) was analyzed using RT-PCR followed by allele-specific restriction digest to assess *Rasgrf1*

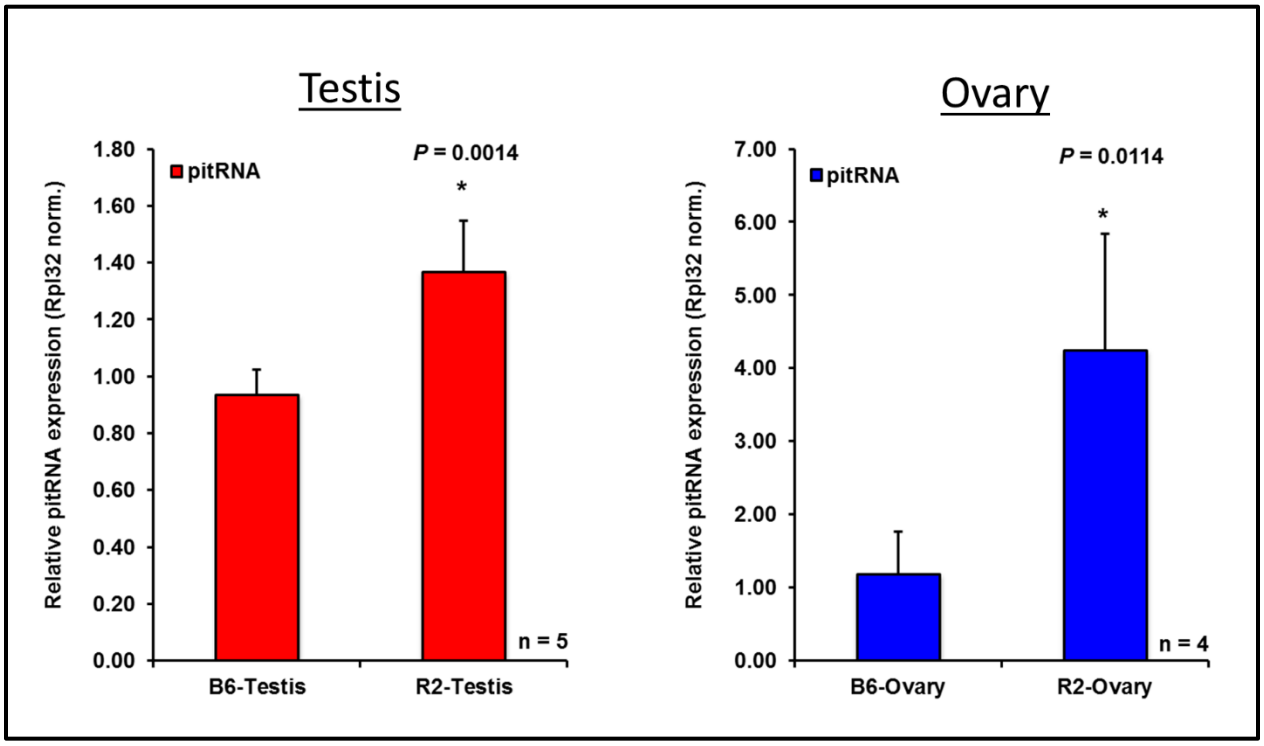

Figure 3.1 - *R2* animals display misregulation of pitRNA. qRT-PCR analysis of adult testes or ovaries in wild type or *R2* homozygous animals. Asterisk denotes p<0.05 using a two-tailed Student's t-test.

expression. Monoallelic paternal expression would indicate a wild type methylation profile, while biallelic expression would signify a *trans* methylation event. Analysis confirmed, however, that neither RNA changed the allelic expression ratios compared to uninjected brain samples of the same background (Table 3.1).

The pitRNA is involved in systems which regulate transposable elements (TEs), so expression of LINE-1, IAP, and MuERV-L elements were investigated by qRT-PCR (Figure 3.2). All three TEs in brain samples derived from pitRNA-injected zygotes were significantly upregulated compared to buffer-injected samples. This upregulation demonstrates that the presence of the pitRNA in the zygotic RNA load exerted lasting effects on gene expression, even to the neonatal brain. Interestingly, TOPO RNA injection caused significantly increased LINE-1 expression as well. IAP and MuERV-L elements in TOPO-injected samples, while not statistically significant using stringent Bonferroni correction, trended toward this same increase. It is also important to note that there was a trend toward TE upregulation in the uninjected samples as well, which is more easily interpreted when the comparison is inverted: the act of injecting buffer into zygotes causes a decrease in TE expression in the P1 brain relative to unmanipulated controls. The buffer-injected samples, however, are the most appropriate normalization when investigating the effects of zygotic RNA injection, so this convention will be maintained for the remaining analyses.

Table 3.1 – Injection of RNA into zygote induces *Rasgrf1* biallelic expression in brain. P1 brain samples displaying biallelic *Rasgrf1* expression were counted and divided by the total number of animals analyzed for percentages. N/A, not assessed.

Sample	biallelic/total animals (percent)
Buffer	N/A
Uninjected	5/25 (20%)
pitRNA	9/46 (20%)
TOPO	6/51 (12%)

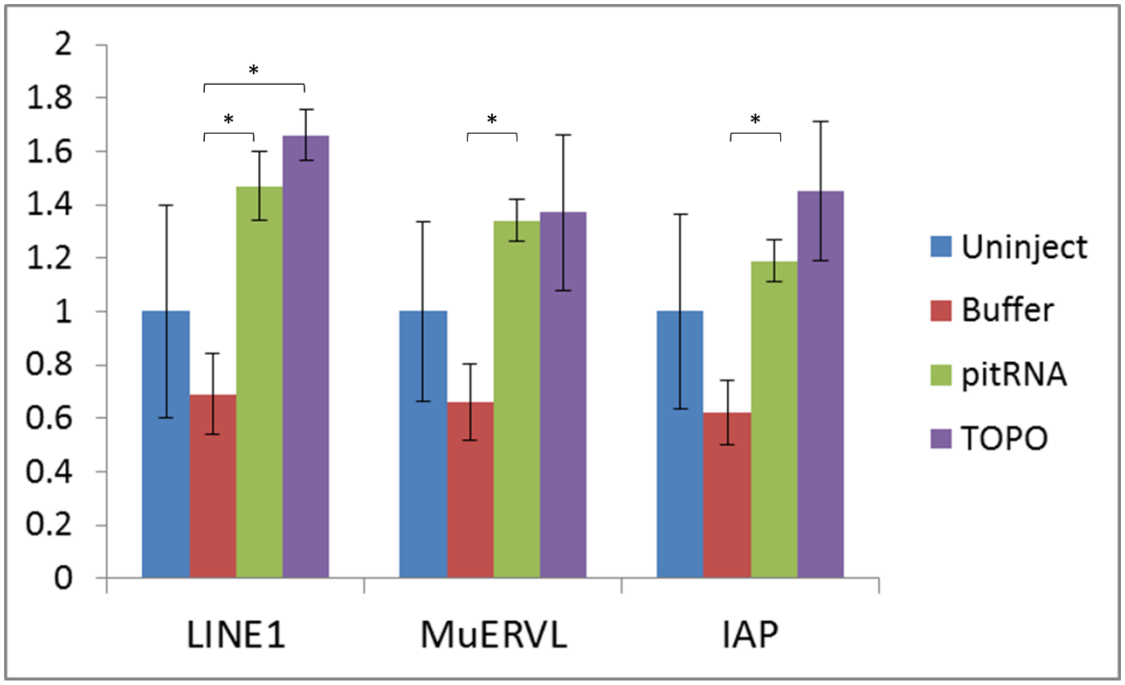

Figure 3.2 - RNA injections cause a global upregulation of TEs. Three TE subclasses were assessed (LINE-1, MuERV-L, and IAP) by qRT-PCR analysis of P1 brains after four zygotic treatments: uninjected, buffer-injected, pitRNA-injected, and TOPO-injected samples. Asterisks represents $p<0.05$ after Bonferroni correction.

Each TE class is comprised of thousands of active copies per cell, often divided into multiple subclasses that are defined by sequence divergence[219]. These different subclasses can be differentially regulated as well[220], so no single set of primers can capture all the TE expression diversity. To bypass these inherent biases and to get a better understanding of the broader effects of RNA injection, we performed RNA-seq from the same tissues in biological triplicates (Table 3.2).

Principal components analysis revealed samples falling into two major clusters (Figure 3.3). Samples derived from uninjected, buffer injection, and two of the samples from pitRNA injection clustered together tightly, indicating high similarity, while samples derived from TOPO injection and one sample from pitRNA injection clustered in a second group. These clusters are very dissimilar, with 97% of the variation present in the experiment explained by their separation. Differential gene expression analysis revealed the reason for these clusters: when compared to the buffer-injected samples, TOPO-injected samples displayed over 11,000 significantly differentially expressed (DE) genes, 1,210 upregulated and 9,888 downregulated (Figure 3.4A). As expected from the PCA plots, pitRNA samples contained very few DE genes (~10) and uninjected samples showed no differences at all (Figure 3.4A). Gene ontology analysis did not reveal enriched categories in any gene list. In contrast, TEs as a group showed significant upregulation in all samples (Figure 3.4B; for all samples $p<10^{-16}$; Uninjected D=0.33; pitRNA D=0.47; TOPO D=0.69; Two-sample Kolmogorov-Smirnov test). The "upregulation" of TEs in uninjected samples confirms the qPCR results (Figure 3.2) and

Table 3.2 – Breakdown of RNA-seq data. Raw reads from the Next-seq 75bp run.

Sample	Raw Reads (millions)
buffer #5R	18M
buffer #7R	22M
buffer #10R	18M
uninjected #2079	15M
uninjected #2080	29M
uninjected #32925	15M
pitRNA #05	36M
pitRNA #07	22M
pitRNA #16	8M
TOPO #22	15M
TOPO #45	12M
TOPO #47	63M

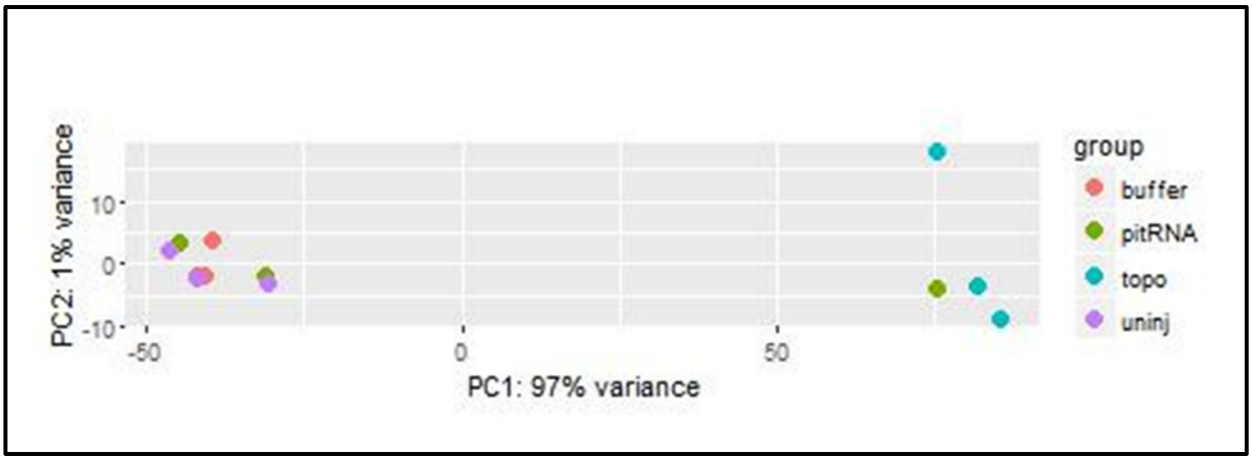

Figure 3.3 - TOPO RNA samples are outliers compared to other treatment groups. Principal components analysis (PCA) plot where 97% of the variability between samples is explained by the x-axis.

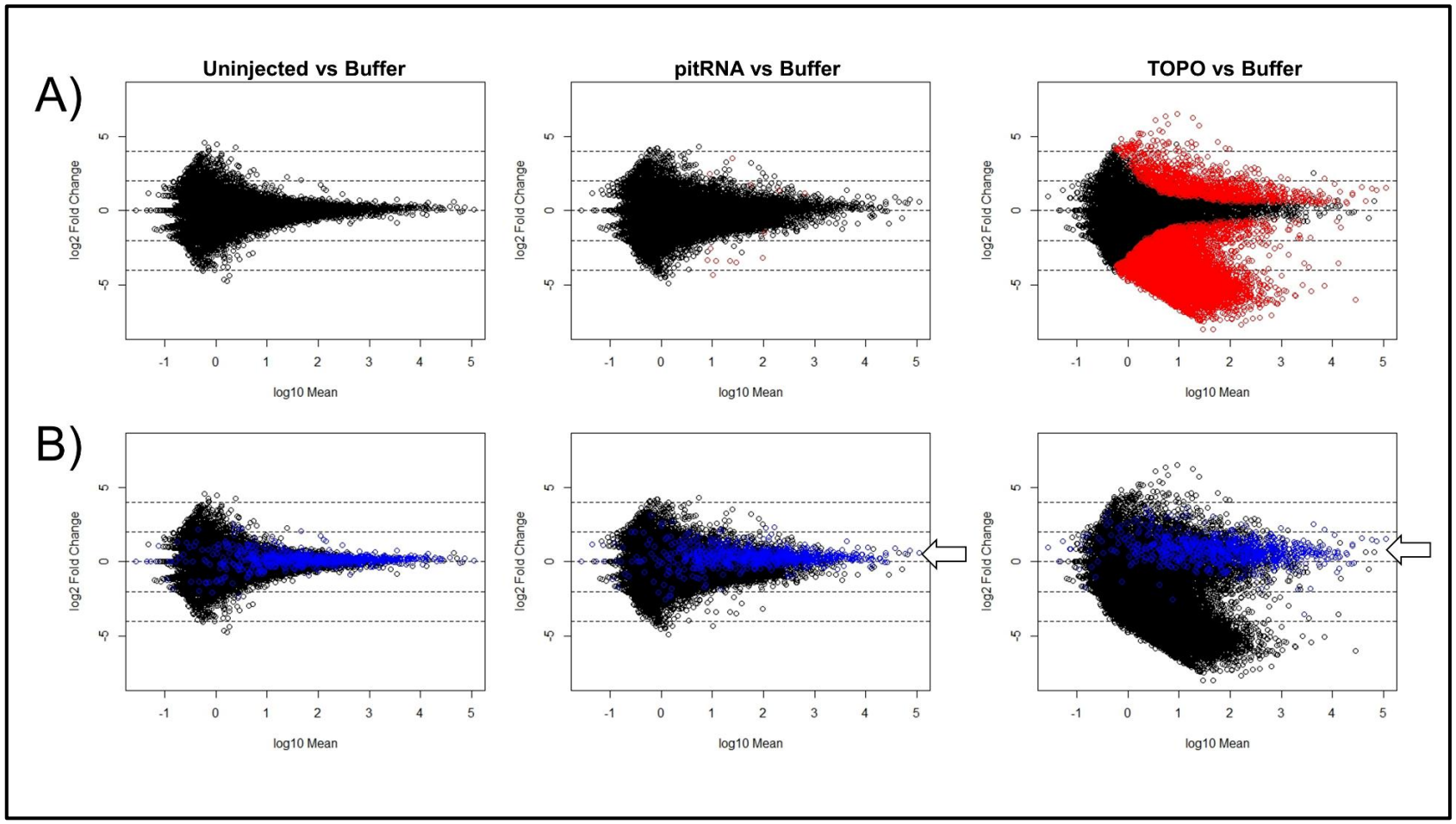

Figure 3.4 - RNA-seq displays sequence-dependent and sequence-independent effects of RNA injection. A) MA plots for buffer vs uninjected (left), buffer vs pitRNA injection (middle), and buffer vs TOPO injection (right) where red dots are individual genes passing a threshold of 5% FDR. B) Same layout as A) except blue dots are now depicting all TEs. Arrow added to highlight the global shift of these points away from zero.

reinforces the conclusion that buffer injection may dampen TE expression when compared to naturally mated samples.

Even when comparing biological replicates of the same sample, it is uncommon for genes to stay at precisely the same expression level. Often expression for genes will randomly trend either up or down in a nonsignificant manner, but since it is random the ratio of upward to downward trending genes should be equivalent (50:50). In our samples, however, the ratio became progressively skewed toward downward trending in the RNA injected samples (Figure 3.5), indicating a modest global downregulation, even in samples that did not reveal many significantly downregulated individual genes.

To characterize the 5mC profiles that might be regulating the TEs, we conducted a pilot bisulfite PCR experiment (data not shown). Tail DNA of a single animal from each of the four treatment groups was subjected to bisulfite treatment, and analyzed by PCR for LINE-1, IAP, and MuERV-L methylation. There was no correlation between expression of the element and their 5mC profiles, indicating that either 5mC was not regulating the expression differences observed in the RNA-seq analysis or that, once again, the diversity of sequences represented by TEs cannot be captured in such a targeted approach.

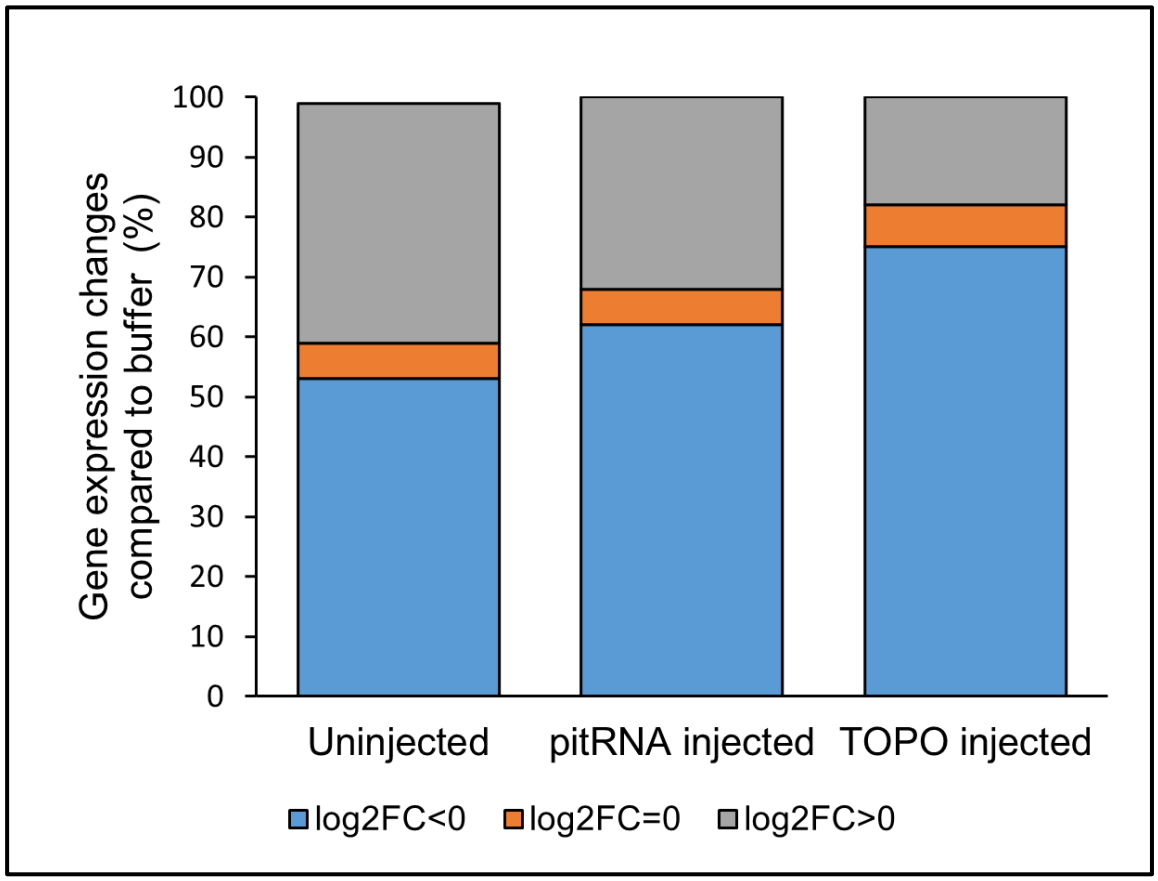

Figure 3.5 - RNA injection samples show a trend toward downregulation. All genes were categorized into increasing, stable, or decreasing expression bins, regardless of individual significance. FC is fold change.

In conclusion, these results reveal both RNA species dependent and species independent effects which were observed long after the original RNA injection, illustrating the importance of RNA in early embryonic development.

Discussion

In this study we have shown that zygotic RNA load can cause molecular phenotypes which are both independent (TE upregulation) and dependent (unique differentially expressed genes) upon the RNA species being involved. These RNAs can lead to global misregulation, the effects of which can be detected in the P1 brain.

Perhaps it is unsurprising that perturbations during the period of dynamic epigenetic zygotic reprogramming could cause significant effects. DNA methylation and many histone modifications transition from gametic profiles to a more pluripotent epigenetic state shortly after fertilization (reviewed in [221]) and may therefore be sensitive to a bolus of RNA interrupting these developmental processes. Many proteins which regulate gene expression during early development are influenced by RNA binding, so aberrant levels or species may sequester them away from their targets. For instance, H3K4me3 is established by WDR5 in the early embryo, a protein which has been shown to bind thousands of RNAs *in vivo* and maintain pluripotency[222]. Another possible candidate is PRC2, the protein complex responsible for all H3K27me3. PRC2 interacts with a broad range of RNAs *in vitro*[223]. *In vivo*, RNAs from highly expressed genes

associate with PRC2 and inhibit its catalytic activity[224]. Loss of EZH2, the catalytic subunit of PRC2, causes embryonic lethality before implantation, indicating that it is vital for normal development[225]. Bivalent H3K27me3 and H3K4me3 marks on developmental genes is important to maintain poised expression profile in preparation for differentiation[21]. It is easy to imagine that zygotic RNA load may inhibit one or more of these proteins which are so critical for development, which might lead to global misregulation and lasting effects on gene expression through the P1 stage of development. TEs would likely be more responsive to global epigenetic changes; in fact, they are most active during gametogenesis and fertilization, the times representing the most dynamic epigenetic reprogramming in development. This sensitivity to epigenetic state might explain their global upregulation in our zygotic RNA-injected brain samples.

But why would TEs be upregulated in both RNA-injected samples (Figure 3.4B) when the expression of so many genes trended downward (Figure 3.5)? It may be possible that these two events are directly linked. Perhaps increased expression of TEs decreased the pool of available transcriptional machinery. This machinery would normally be available to all active genes, but a slight deficit due to TE-mediated titration could increase the probability that the expression of a random gene would trend downward. Evidence for this can be seen in the relationship between the global downward trends in Figure 3.5 and the D-scores of the Kolmogorov-Smirnov (KS) test used to statistically assess the TEs. A KS test compares the distributions of two samples, finds the area of maximum divergence between their distributions, and calculates both the difference (the D-score) and the likelihood of that difference

occurring by chance (p-value). In our samples the values of the D-score increase—representing greater TE divergence from the distribution of the rest of the genes—proportionally to the increase in percentage of downward trending genes. Further analysis is needed however, as the D-score from the zygotic TOPO-injected brain samples may be confounded by its asymmetric distribution.

An alternative explanation for the global downward expression trend might lie in interconnected nature of TEs and their host gene landscape. The heterochromatin state of TEs has been shown to both positively[226] and negatively[227] affect local gene expression and global gene regulatory networks[228,229]. Some TEs may even double as enhancers, as evidenced by their tissue-specific hypomethylation, gain of H3K4me1, p300 occupancy, and strong correlation with expression of nearby tissue-specific genes[230]. It is therefore possible that the increased expression of TEs might inhibit global expression through indirect effects on unknown regulatory pathways.

The RNA-mediated inhibition of *trans*-acting proteins would explain the shared misregulation patterns between the two RNA species: the TE upregulation (Figure 3.4B) and the downward shifting expression ratios (Figure 3.5); however, many thousands of genes are misregulated specifically upon TOPO injection (Figure 3.4A), indicating that key differences exist either in sequence or structure between the two RNAs. Perhaps TOPO RNA acts as an endogenous competing RNA, much like *LINC-ROR*[76] or *H19*[231], competing away miRNAs which would normally repress translation of a transcriptional

master regulator which, when free from repression, may activate or repress large numbers of downstream genes. The secondary structure might also have significant regulatory consequences. Riboswitches, for instance, are RNAs which can change secondary structure to bind both DNA and proteins and can affect a variety of cellular functions including transcription, termination, changes in translation rate, splicing, and mRNA decay (reviewed in [232]).

There are concerns of experimental design that warrant caution in making conclusions from these unexpected results. The first concern regards the molarity of the injected transcripts. The pitRNA and TOPO RNA were size-matched to maintain equivalent molarity; however, during *in vitro* transcription the TOPO plasmid always produced an additional 1kb RNA product, possibly from an ectopic T7 site within the plasmid backbone. The heterogeneous nature of the TOPO RNA, therefore, represented a higher molarity of RNA which might induce the robust effects we see in the zygotic TOPO-injection brain sample, namely the massive misexpression of more than 11,000 genes. Perhaps any RNA injected at similar concentration would show the same profile after reaching a specific threshold. Indeed, evidence from within our dataset may actually support this conclusion: the PCA plot depicts two of the pitRNA samples clustering tightly with the buffer-injected and uninjected samples, while the third pitRNA sample clusters tightly with the TOPO. The application of injection media is imprecise—based solely on the appearance of pronuclear “swelling”—so it is possible that one pitRNA sample received an RNA dose which surpassed the threshold and now

displays “TOPO-specific” effects. Future injections with varying concentrations of TOPO RNA will need to be conducted to resolve this threshold hypothesis.

A second concern lies in the potential bias of batch effect. Three separate injection experiments were carried out with a different RNA species injected on each day. In hindsight, a better experimental design would call for the injection of all three treatments on each day so that batch effects based on injection day are eliminated. This design would also create “biological replicates” which would incorporate variation between days and permit more confident statistical analysis. Our current data make the batch effect an unavoidable confounding factor in the downstream analysis. It is therefore difficult to claim that gene expression differences from animals injected with different RNAs on different days is solely due to the RNAs themselves. Our next experiment should be to repeat at least one more injection with the same RNAs injected on the same day. If we observe the same PCA clustering patterns as seen in Figure 3.3, then we can confidently conclude that our results are biological rather than a technical artifact due to batch effect. Even within one day, however, many developmental events can occur during the course of the experiment, potentially introducing a batch effect in developmental timing between the first and last round of injections. An ideal experiment, therefore, would inject mature oocytes which could then be fertilized nearly simultaneously with *in vitro* techniques, mitigating this timing issue.

It is important to highlight that the pilot study of 5mC at TEs not only had poor statistical power, but also might be unable to capture the diversity of sequences that TEs represent. A possible next step might be to do whole genome bisulfite sequencing (WGBS) on these samples[233]. TEs make up ~40% of the genome, so even a WGBS experiment at low coverage should be sufficient to assess 5mC of TE consensus sequences, which can represent thousands of individual transposon sites.

In conclusion, we have shown that zygotic RNA load might have a lasting effect on gene expression even after birth. These results also fit with the hypothesis that IEI can be mediated by RNA from either the male or female gametes.

Authors' Contributions

DHT, JM, and PDS designed experiments. JM and RH performed experiments for Figure 3.1. DHT performed all other experiments. DHT, KW, and DZ performed analysis. DHT and PDS wrote the manuscript.

Acknowledgements

We thank Erin Chu and Roman Spektor for their stimulating conversations that benefited both the science and aesthetics of this paper. We thank the Cornell Stem Cell and Transgenic Core Facility for their work injecting zygotes and the Cornell University

Biotechnology Resource Center (BRC) for their assistance in high throughput sequencing.

CHAPTER 4 - EXTENDED DISCUSSION

In the first part of this work I showed that a 3kb sequence from the *Rasgrf1* ICR was sufficient to recapitulate imprinted methylation patterns at an ectopic locus. I also showed that sperm lacking methylation at the *Wnt1DR* locus were sufficient to induce robust somatic methylation, providing indirect evidence for a novel post-fertilization mechanism. These findings have been submitted for publication and will likely be in press before the end of the year. In the second part of this work, I investigated how the misregulation of these very same sequences—namely through misregulation of pitRNA expression—might drive aberrant methylation effects both in *trans* and intergenerationally. Instead of finding *Rasgrf1*-specific effects, however, I described how zygotic RNA load might induce global effects, some of which are dependent upon the RNA species being queried and some of which are RNA species independent. More computational analysis, which will be described below, is required before these findings will be considered complete. Altogether this body of work advances our understanding of how the *cis*-acting sequences at *Rasgrf1* function and how misregulation of these sequences might cause unexpected effects.

A considerable body of knowledge has been built describing the sequences and mechanisms regulating imprinted 5mC and expression at the *Rasgrf1* ICR, however, many more questions remain. In this section I will probe deeper into the mysteries surrounding the *Rasgrf1* locus and the field of *cis*-acting recruitment of methylation in

general. Tissue derived from e16.5 embryos are diminishingly small, so some of the experiments I discuss would only be feasible in a cell culture system. A perfect culture system model to test germline based methylation events would contain pitRNA expression, robust piRNA biogenesis, and properly established DNA methylation on a transfected wild type *Rasgrf1* ICR; our lab is characterizing multiple cell lines which might be suitable for the task.

Establishment of 5mC on the *Rasgrf1* ICR during gametogenesis

Regulation of the repeats and pitRNA

The repeats

The repeats at the *Rasgrf1* ICR have been shown to act as a promoter for pitRNA expression specifically during PGC development, with virtually no expression observed outside of this time and tissue[158]. This begs the question: How are these sequences so tightly regulated? An unbiased approach to identify relevant transcription factors (TFs) would involve a gel shift assay using repeat sequences in cell extracts, followed by pull-down and mass spectrometry to ascertain bound proteins. Another approach would look for TFs with differential expression between male and female PGCs that have binding sites within the ICR, searching for either a positive regulator in the male PGCs or a negative regulator in female PGCs. SP1 has a canonical binding site within the repeats and is implicated in pitRNA expression (unpublished data),

however, this protein is broadly expressed and therefore cannot explain the restriction of pitRNA expression to male PGCs without another factor. Of course, the factor may be a chromatin feature, like a histone modification, and would therefore not have a binding motif at all. In this case it may be valuable to determine the epigenetic profile of the male and female PGCs by chromatin immunoprecipitation followed by qPCR (ChIP-qPCR) at the ICR to assess if they might be granting male germline specificity.

There might also be non-protein elements involved in this specificity as well. For instance, the repeats may be regulated by differential G-quadraplex (G4) formation. G4s are stretches of G-rich DNA which use non-Watson Crick bonding to fold into a four sided secondary structure, also described as a G-quartet. They are enriched at regulatory sequences and are implicated in gene regulation (reviewed in [234]). The repeats are enriched for sequences which might generate G4s, but the formation and regulation of G4s at this region is currently unknown. It is important to note the relationship between 5mC and G4. While *in vitro* studies argue that 5mC stabilizes G4s[235], *in vivo* studies suggest 5mC has an inhibitory effect on G4 formation. Comparisons across tissues depict a negative correlation between the two[236], and direct chemical or nutritional inhibition of 5mC results in fewer G4s[237]. This generates a model in which the unmethylated profile in PGCs allows for the formation of G4s in the male germline. Proteins which specifically bind G4s might then display male-specific activation of pitRNA expression. Interestingly, SP1 appears as a likely candidate in this regard because it has been shown to bind G4s in a non-canonical fashion[238]. The pitRNA would then repress its own transcription by recruiting methylation and

weakening G4s on the repeats. This hypothesis would require binding of a G4 repressive factor in female PGCs however, since both sexes have the same unmethylated profile at this time point. This repressive factor need not be restricted to the female PGCs since the critical time for pitRNA-mediated methylation is narrow.

It would also be valuable to assess if the repeats are merely a promoter for the pitRNA, or if they have other cis-acting functions. If it is only a promoter for the pitRNA, then replacement of the repeats with an inducible promoter, in conjunction with an experimental design that replicates the pitRNA expression profile, should be sufficient to display wild type imprinting at *Rasgrf1*. If, however, the repeats have additional functions outside of its role as promoter to recruit 5mC then an inducible promoter might be unable to phenocopy the wild type imprinting pattern.

Outside of the questions of necessity and alternative functionality in PGCs, we can certainly conclude that the repeats are not sufficient to recruit wild type 5mC patterns in sperm. Evidence for this stems from the results of Chapter 2 in which the repeats inserted into the *Wnt1* locus showed incomplete methylation in the sperm. This is likely due to the weak expression of pitRNA from the *Wnt1DR* allele and argues that either a more expansive set of sequences from the ICR is required or that the chromosomal context (histone modifications, interfering transcription of *Wnt1*, etc.) is more important than we realized. Use of a broader range of repeat sequence inserted at the *Wnt1* locus could address the first concern, potentially providing sequences for important TFs which can reproduce the wild type pitRNA expression level and timing.

Profiling chromosomal context such as histone modifications in the wild type and the *Wnt1DR* alleles might yield some clues as to the weak expression. Alternatively, insertion of the same sequences into an array of new loci would avoid the particulars of the *Wnt1* locus altogether and address the effects of chromatin context in a much more robust way. For instance, comparing pitRNA expression, 5mC level, and local chromatin marks at different insertion sites could yield interesting correlations.

The pitRNA

We have observed that robust pitRNA expression occurs between e12.5 and e16.5 in the male germline, expression which correlates with piRNA-induced remethylation; however, we have not defined the developmental window in which the pitRNA is necessary. Perhaps expression at earlier or later times may still be able to induce wild type 5mC patterns. The previously described inducible promoter system would also be a useful tool to address this question. Turning the promoter on and off at different time points during gametogenesis and checking the methylation of the mature sperm would reveal what expression timing and expression levels are vital. It would also query the effects of aberrant expression, a question I will address in a later section.

Rather than just asking if the pitRNA is necessary, we could also ask if it is sufficient for 5mC. Work from Chapter 3 revealed that the presence of the pitRNA after fertilization does not influence 5mC at *Rasgrf1* in a stable way; however, since the locus has been shown to act in *cis*, it is likely that active transcription or close association with the locus is required. Recent work has detailed successful experiments in which RNA

was tethered to an ectopic locus using a catalytically dead Cas9 protein and a special guide RNA fused to the lncRNA being tested for function[239]. This system could be used to tether the pitRNA to a repeat-less ICR to see if the presence of the RNA sequence, rather than its transcription, is sufficient for 5mC. We could also tether it to ectopic loci with either high or low similarity to the DMD to assess if DNA similarity is a requirement, or if the specificity derives solely from the RNA:RNA base pairing that recruits PIWI proteins. This approach, however, may not work if 5mC requires the DNA:RNA hybrids that are a product of active transcription.

Another outstanding question is how the pitRNA only acts in *cis* when under the control of the repeats in the embryonic testes, whereas aberrant regulation effected by the *R2* allele enables 5mC in *trans*. It is possible that this is due to the level of transcription. pitRNA transcription may be stoichiometrically calibrated with piRNA levels such that co-transcriptional processing leaves no excess pitRNA to function in *trans*. If this were the case, overexpression of pitRNA might overwhelm the piRNA machinery, possibly resulting in aberrant *trans* methylation. *Cis*-action might also be due to structural features associated with the act of transcription, like R-loops, which cannot be “exported” to another locus. Short RNA:DNA hybrids are a normal part of transcription, but in special cases more extended, stable regions are maintained in this hybrid complex, forming R-loops[240]. R-loop structures have surprising, if not paradoxical, mechanisms and functions, facilitating both active transcription by protection against DNMTs at transcription start sites[241] and efficient termination mediated by H3K9me2 recruitment and heterochromatinization at termination sites[242].

Tethering the RNA to the locus via non-Watson Crick base pairing would ensure that only the expressed allele is affected, and the stability of these hybrids might extend the time in which the RNA is “presented” to its interacting proteins or RNA.

The requirement of the piRNA pathway: questions and clues

In the introduction I described many papers that explored the piRNA pathway, but none of them have determined the mechanism by which piRNA proteins recruit DNA methylation. If it is PIWI regulated, then the pitRNA is uniquely situated as a powerful tool to explore this question because it is a single copy locus implicated in robust, piRNA-mediated methylation. Biochemical studies similar to those mentioned previously (pulldown and mass spectrometry) would certainly be a valuable direction of investigation. After identifying the proteins that interact with the pitRNA, broader questions can be asked, like whether these proteins act at all TEs, if their overexpression would cause any aberrant methylation patterns, or if they have any other functions outside of TE repression. It is possible that important factors for 5mC establishment would interact only with the PIWI proteins, and that this RNA-focused approach would be ineffective. Direct interaction between 5mC machinery and PIWIs is unlikely, however, as a thorough immunoprecipitation and mass spectrometry experiment of all murine PIWIs revealed none of these factors[243].

Though strong, the evidence for piRNA and pitRNA in 5mC recruitment is not unassailable. The relevant conclusions for piRNA-induced 5mC at the ICR are derived

from spermatogonia in animals with no piRNA machinery (MILI and MIWI2 null). Male animals lacking this machinery do not progress through meiosis due to genome damage and apoptosis incurred by TE mobilization[125,244]. These defects could be due to problems that arose in the germline stem cells that enables their own survival, but not the survival of differentiating derivatives (spermatogonia). Spermatogonia of mice lacking the piRNA pathway, therefore, might not adequately represent the mechanisms that function at the *Rasgrf1* locus in wild type cells. For instance, perhaps the 5mC mechanisms that function specifically at the *Rasgrf1* ICR are delayed due to the stressed state of the cells. This delay might be resolved by sperm maturation, except these cells die before reaching that critical time. An alternative piRNA-independent mechanism for *Rasgrf1* 5mC may reside in the MORC1 ATPase, a protein which has already been shown to be essential for *Rasgrf1* 5mC. Elimination of this protein does not affect piRNA biogenesis, but does lead to derepression of subsets of TEs and, most importantly for this work, loss of methylation at the ICR[220]. It is also possible that this protein is not independent, but instead acts downstream of the piRNA pathway; however, this relationship is currently unclear.

In an orthogonal approach to assess if 5mC at this locus is piRNA independent, I could mutate piRNA sites within the pitRNA without affecting its transcription. 5mC recruitment in the absence of these sites would signal a piRNA-independent mechanism. Additionally, by disrupting the pitRNA instead of the piRNA pathway, we could characterize their interactions in cells which are not destined to undergo apoptosis. The piRNA profiles of these animals would have to be rigorously checked,

however, to ensure that piRNAs with multiple mismatches were not mapping to other sections of the *Rasgrf1* ICR. Two other alternative experiments are: 1) Delete the primary piRNA clusters from chromosome 7 to see if 5mC occurs in their absence, and 2) Insert an ectopic polyadenylation site between the TSS and the piRNA sites, effectively asking the same question by truncating the transcript.

Previous work has shown that unique primary piRNAs are able to mediate the processing of the pitRNA into secondary piRNAs[158], however it is not known if this processing is required for 5mC recruitment. Methylation occurs in the nucleus, likely in a co-transcriptional manner, and yet MIWI2—the only PIWI protein that localizes into the nucleus[66]—does not require slicer activity for normal piRNA production[67]. MIWI2 is required for TE methylation[68]. It is possible, therefore, that the function of MIWI2 is primarily as a platform for epigenetic regulation, and that pitRNA processing might be inconsequential for 5mC recruitment. Purification of processed piRNAs from an animal without repeats would yield a wild type piRNA profile, except without pitRNA-derived secondary piRNAs. If the “ideal” cell culture system that I outlined earlier exists, then lipofection of this RNA pool into MILI/MIWI catalytic mutant cells, which cannot create their own piRNA, could test if the catalytically dead MIWI2 protein is able to recruit methylation in the absence of secondary piRNA processing. This question is important because it could decouple two tightly connected PIWI protein functions: piRNA biogenesis and epigenetic regulation. A caveat must be made to this model, however, as modest 5mC is found on the ICR in MIWI2 null spermatogonia, highlighting the mystery surrounding the proteins regulating this region.

Maintenance of 5mC on the *Rasgrf1* ICR after fertilization

I had previously mentioned finding a cell culture system which models e16.5 PGCs so that *Rasgrf1* 5mC establishment could be biochemically assessed. A cell culture system that accurately represents the post-fertilization time point would also be valuable, as zygotes are even less abundant than PGCs. The repeats are necessary to maintain 5mC after fertilization, so a good cell culture proxy must show a loss of 5mC upon repeat deletion. It is possible that embryonic stem cells (ESC) might be such a system since they are derived from preimplantation embryos; however, they have not yet been validated.

The repeats are required between fertilization and implantation to maintain 5mC on the ICR, indicating that either the repeats or the pitRNA have a vital role to play at this time point[161]. Fortunately, many of the experiments outlined above can be used to address questions in the post-fertilization embryo as well. For instance, the previously mentioned inducible promoter system would be able to assess if expression the pitRNA is sufficient for 5mC maintenance. If it is not sufficient, then it is likely that the repeats are acting in some other capacity after fertilization. Assessing alternative roles of the repeats would be difficult to determine *in vivo*, since alterations to the repeats might also influence the establishment of 5mC in the germline as well. Therefore, it would be valuable to assess cell culture systems that reflect the post fertilization state, such as

the ESCs mentioned previously. Such a system could be used to biochemically identify interacting proteins mediating the repeats' protective function.

What if, however, the inducible expression of the pitRNA is sufficient for 5mC maintenance? None of the mechanisms for imprinting maintenance (Stella and Zfp57) have been shown to be RNA-dependent, and the piRNA pathway is not known to be active outside of the male germline, so any such pitRNA-dependent maintenance mechanism would certainly be novel. Here again, a validated post-fertilization cell culture system would be valuable for biochemical identification of pitRNA-interacting proteins which would open a whole new line of questions. We could also ask what sequences of the pitRNA are sufficient for its activity by truncations/mutations within the transcript. Finally, we could also use the inducible promoter system to more tightly define, *in vivo*, the developmental stage in which the pitRNA is required. The animals could also be used to investigate the effects of aberrant regulation, which I will discuss later.

Another interesting area of investigation stems from the results of Chapter 2, where I hypothesized that the robust somatic paternal methylation recovery observed after incomplete sperm methylation might be due to a secondary post-fertilization establishment mechanism. Of course, the alternative explanations of sperm fitness effects and auxiliary methylation sites would have to be ruled out first before we consider an entirely new mechanism. Our study found that motility was unlikely to be a biasing factor toward methylated sperm; however, sperm motility is not the same as

sperm functionality. It is possible that unmethylated sperm with good motility may be selected against, perhaps due to inhibited chemotaxis or an inefficient acrosomal reaction. These fitness hypotheses could be assessed by *in vitro* fertilization (IVF) studies. If unmethylated *Wnt1DR* sperm cannot fertilize eggs due to reduced functionality, then we would expect litters derived from these sperm to show a similar range of incomplete methylation phenotypes: some pups with high 5mC and some with no 5mC. It is also possible that unmethylated sperm could fertilize efficiently, but that they are unable to generate a healthy embryo due to some early developmental defect. The IVF experiments above could test this hypothesis up to the blastocyst stage, with later stages checked at various time points after implantation of the IVF blastocyst into a pseudopregnant female. Inviability at these later stages should result in higher levels of embryo resorption than a wild type control and lower levels of methylation than healthy littermates. The hypothesis for alternative methylation sites would be more difficult to query, as the repeats—the most likely location for methylation due to its high GC content—are too large and repetitive for bisulfite sequencing methods and do not contain methylation sensitive restriction sites. A novel approach to assess 5mC at the locus might be to use sequence capture followed by modern sequencing technologies like Nanopore[245] or single molecule, real-time (SMRT) sequencing[246] which both claim direct detection of 5mC without bisulfite treatment. These technologies also claim longer read lengths which would be valuable for allele-specificity.

These other alternatives eliminated, I would investigate the secondary post-fertilization establishment hypothesis. An easy first experiment might better define the

timing of methylation of the *Wnt1DR* ICR after fertilization, with the expectation that many zygotes would have no methylation while time points after implantation would display increasing methylation acquisition. The inducible promoter animal could also be used to test the post-fertilization methylation theory, as it might be possible to recruit 5mC after fertilization in the absence of correctly established sperm 5mC. A result like this would get into similar questions of 1) necessity, 2) sufficiency, and 3) allelic discrimination regarding ICR components that I have previously addressed and will therefore not restate. The difficult issue for these experiments, be they *in vivo* or in cell culture, is that the germline establishment of methylation on the ICR must be abrogated to test post-fertilization establishment mechanisms. Cell culture systems would be easier of course, as chemical inhibitors of 5mC or DNMT knockout/replacement systems could induce this erasure. Discovery of a post-fertilization 5mC mechanism would be of high impact, as it goes against the current imprinting paradigm. If such a mechanism were elucidated at the *Rasgrf1* ICR, I would also want to test other imprinted loci. Ectopic insertions of the *H19* ICR showed similar phenotypes to those seen in my *Wnt1DR* study: incomplete sperm gametic methylation resulting in robust somatic methylation[193,195]. These results argue that such a mechanism may be more widespread than just the *Rasgrf1* locus.

R2 transgenerational effects

I showed in Chapter 3 that zygotic RNA load can have lasting effects on gene expression in neonatal brain, with substantial differences that vary depending upon the

RNA species used and the target genes affected. As is often the case, these results generate more question than answers. The following sections will address questions regarding both the results of zygotic RNA injection and issues with the original *R2* experiments which might be informative for future experiments.

Lack of *R2* phenotype upon pitRNA injection

The experiments in Chapter 3 were motivated by the hypothesis that mislocalization and misexpression of the pitRNA might be inducing the *trans* and intergenerational effects seen the *R2* mutant animals, but our results show that zygotic RNA injection was not sufficient to reproduce these phenotypes. It is possible that our hypothesis is incorrect; however, it is important to note experimental caveats that might provide alternate explanations. For instance, the RNAs in these experiments were not modified in any way before injection, and the modification state of the *R2*-derived pitRNA might be important for *trans* and intergenerational effects. RNAs can be capped or polyadenylated to increase its stability, or they can be internally modified in over 100 different ways (reviewed in [247]). It has also been shown that an RNA methyltransferase, DNMT2, is required for at least two different RNA-mediated intergenerational effects to occur[178], indicating that RNA modification is likely important for intergenerational stability. Our qPCR detected higher levels of pitRNA in the *R2* animals, but this assay only measures steady state RNA levels; therefore, this increase might represent lower degradation rates than wild type transcription rather than increased transcription due to misregulation of the *R2* allele. No efforts have gone into characterizing the modifications

of either the endogenous pitRNA or the *R2*-pitRNA, so a more thorough study of their modifications might inform an RNA design which would phenocopy the *R2* mutant upon injection into zygotes.

Another explanation for the failure of the pitRNA to recapitulate the *R2* phenotype may lie in the sequence of the pitRNA itself. The pitRNA used in these experiments was derived from the wild type allele. The *R2* transcript, in contrast, is not wild type. Transcription starts within the *R2* promoter, so the *R2* pitRNA is actually chimeric in nature: a section of *IGF2R* sequence followed by the wild type pitRNA. This aberrant transcript, either by its sequence identity or an unknown secondary structure, may itself recruit factors that are not present at the wild type ICR to mediate its aberrant 5mC, a phenotype that could not be reproduced with the wild type pitRNA. It is possible that injection experiments with the chimeric *R2* RNA would be able to phenocopy *R2* 5mC.

A third concern is that the timing of *R2*-derived pitRNA misexpression may be key to the *trans* and IEI effects of the *R2* mutant, but this expression profile is poorly characterized. The earliest time point that the paternal *R2* allele could induce 5mC in *trans* onto the maternal allele is at fertilization, and the repeats are dispensable for maintaining 5mC at the wild type ICR after implantation, so we hypothesized that this narrow window was likely when the *trans* effect was occurring. Similarly, the IEI effect may have nothing to do with the presence of pitRNA in the zygote, but may be due to inappropriate expression inhibiting 5mC reprogramming on the maternal allele during

PGC development. Simply put, the *trans* and IEI effects may rely heavily on the timing of pitRNA activation from the *R2* allele, but this timing is poorly defined. Querying the pitRNA misexpression profile of *R2* animals across different times and tissues by qPCR would provide descriptive, rather than mechanistic, information. A more useful approach would be the inducible promoter experiment as outlined previously. Control over timing and amount of pitRNA may allow us to more tightly define the circumstances of the *R2* phenotype both in *trans* after fertilization and intergenerationally during gametogenesis.

Fertilization initiates a series of rapid and dynamic developmental processes, including the immediate hardening of the zona pellucida, the resumption of meiosis in the oocyte genome, the initiation of active 5mC reprogramming in the male pronucleus, protamine exchange in the male pronucleus, and pronuclear fusion (reviewed in [248]). Another concern is that our RNA injection experiments utilized zygotes from natural matings, a time point at which many of these processes have already begun. If our RNA injections occurred before fertilization, however, perhaps in the mature oocyte, then we might introduce the RNA during these early events at a time of greatest effect. These oocytes could then be subjected to *in vitro* fertilization to better control the initiation of development, brought to term, and tested for the *trans* and IEI phenotypes.

Global effects of zygotic RNA load

Injection of the pitRNA and TOPO RNA revealed interesting global misexpression phenomena, but a full characterization of the phenomena may shed

more light onto potential mechanisms. Future injection experiments may test many variables to better define the effects that were RNA species dependent or independent. For instance, injection of RNA species that have low or high GC content may have different protein binding affinities and therefore have different inhibitory capability; the same can be said for sequences with differing secondary structures. Different lengths of RNA would also be valuable to test, as it is known that some proteins that promiscuously bind RNA, such as PRC2, are length dependent[223]. Perhaps this length dependency explains why the ~20nt miRNA injections of the other previously described IEI experiments do not induce these effects. It would also be valuable to test a variety of RNA concentrations to evaluate the threshold hypothesis laid out in Chapter 3. Perhaps the pitRNA and TOPO PCA clustering would switch positions if pitRNA concentration were increased to 450ng/ul and TOPO concentration were decreased to 350ng/ul, illustrating that the “TOPO-specific” effects are actually concentration dependent effects rather than RNA species dependent. Testing some of these these variables would allow us to tease apart the multiple overlapping phenotypes—TE upregulation, global trends toward downregulation, and the 11,000 “TOPO-specific” DE genes. These effects may all be due to misregulation of a single master regulator, but they may also derive from independent mechanisms which would respond differently to the variables we decide to test.

We chose to assay RNA from P1 brain in our study because that is where *Rasgrf1* displays imprinted expression, however, many other tissue types from the same time point may see similar misregulation. Analyzing gene expression from

different tissues (lung, liver, heart, skin, etc.) in the injections we planned above would give us insight into how widespread the effects are. It is also possible that these zygotic RNA injection effects are specific to the brain. MECP2, a protein implicated in many neurodevelopmental diseases, has been shown to repress LINE-1 transposition specifically in the brain[249] and hippocampus has been shown to upregulate LINE-1, presumably to increase plasticity in somatic cells[134]. These data argue that the brain may have a higher likelihood for TE misexpression and may be more sensitive to zygotic perturbation than other tissues.

It is also vital to mine the rich RNA-seq dataset that we already possess for mechanistic insights. Intersection of our dataset with ENCODE datasets might reveal enrichment in particular chromatin marks or transcription factors. For instance, if a TF is enriched near the transcription start site of the downregulated genes, but its expression profile is unaffected, then this might signal RNA-mediated post-transcriptional inhibition. A more nuanced approach might use weighted gene co-expression network analysis, a method which identifies highly connected nodes as potential upstream regulators by defining groups of genes with similar expression changes[250,251]. It is also possible that the RNA injections may have also affected alternative splicing, so a program like mixture-of-isoforms (MISO) might be able to determine aberrant splice forms that would lead to misregulation[252]. If our injected RNA is acting like a miRNA sponge, however, then we might create a cross-referenced list of miRNAs which 1) bind our RNAs and 2) are expressed during early development. We would expect mRNA targets of these miRNAs to be upregulated in the RNA-injected samples when compared to buffer-

injected samples in our RNA-seq datasets. miRNAs might even regulate some TEs directly[253], making it possible that the TE upregulation effects seen in our data might be due to miRNA depletion.

If a master regulator were implicated, we would expect two main characteristics: First, it would be regulated by RNA. *In vitro* binding and activity assays either with or without competing RNA would detect if the inhibition is direct. If the master regulator were affected indirectly, such as through aberrant splicing, then the same phenotypes should arise from a cell line that expresses the master regulator. For instance, lipofection with the RNA might induce comparable aberrant splicing and downstream misexpression effects as observed in our data. Second, a master regulator, if inhibited by alternative means, should phenocopy our results. Therefore, shRNA knockdown in zygotes, followed by RNA expression analysis in the P1 brain should cause the same global effects.

Loss of the *R2* phenotype

A less risky approach to the question asked by Chapter 3 might be to sequence oocytes or early post-fertilization embryos from *R2* animals and assess any aberrant RNA transcripts directly before proceeding to RNA injections designed to phenocopy the intergenerational effect. Unfortunately, the intergenerational effect is no longer present in these animals, so sequencing would probably not be informative. When the

Soloway lab was at Roswell Park Cancer Institute in Buffalo, the *R2* phenotype was robust; however, upon moving to Cornell three phenotypes changed: 1) The intergenerational phenotype was lost, 2) The *trans* methylation phenotype became less penetrant, and 3) Mutant animals with no repeats, which were normally unmethylated at the DMD, began to stochastically methylate a small proportion of their alleles. Similar phenomena in which environment can significantly alter epigenetic and phenotypic states have been observed by other labs as well. For instance, rat pups that experience attentive maternal care (licking, grooming, and arched-back nursing) gained methylation at the glucocorticoid receptor and had less adult anxiety than animals who did not receive such treatment[211]. Exposure to chemicals like bisphenol-A, a molecule commonly found in plastics, have been shown to change methylation state of a LTR transposable element, which subsequently affects expression of the downstream *Agouti* gene[212]. Differential *Agouti* expression results in a continuum of coat color from brown to yellow. Diet can change 5mC as well. Supplementation with vitamins involved in methyl donor synthesis has been shown to affect LTR methylation and *Agouti* expression[210].

So what might be the difference between Roswell Park and Cornell? A pilot study was performed on *R2* mice to investigate if dietary supplementation could rescue the lost phenotypes (data not shown). Normal, choline depleted, choline enriched, and high fat diets were given to mice in the same experimental design as those described in the study by Herman et al, but there was no significant alteration of *Rasgrf1* monoallelic expression, indicating that the ICR methylation state was likely unaltered as well. It is

possible that some other environmental stimulus is responsible for the loss in phenotype (sleep/wake cycles, noise level, air recirculation, chemical exposure, etc.), but we have decided to not pursue these questions as there are more variables than time.

REFERENCES

1. Zhou, V. W., Goren, A. & Bernstein, B. E. Charting histone modifications and the functional organization of mammalian genomes. *Nat. Rev. Genet.* **12,** 7–18 (2011).
2. Vidali, G., Gershey, E. L. & Allfrey, V. G. Chemical Studies of Histone Acetylation THE DISTRIBUTION OF ε-N-ACETYLLYSINE IN CALF THYMUS HISTONES. *J. Biol. Chem.* **243,** 6361–6366 (1968).
3. Kuo, M. H., Zhou, J., Jambeck, P., Churchill, M. E. & Allis, C. D. Histone acetyltransferase activity of yeast Gcn5p is required for the activation of target genes in vivo. *Genes Dev.* **12,** 627–639 (1998).
4. Bernstein, B. E. *et al.* Genomic Maps and Comparative Analysis of Histone Modifications in Human and Mouse. *Cell* **120,** 169–181 (2005).
5. Guenther, M. G., Levine, S. S., Boyer, L. A., Jaenisch, R. & Young, R. A. A chromatin landmark and transcription initiation at most promoters in human cells. *Cell* **130,** 77–88 (2007).
6. Luco, R. F. *et al.* Regulation of alternative splicing by histone modifications. *Science* **327,** 996–1000 (2010).
7. Joshi, A. A. & Struhl, K. Eaf3 chromodomain interaction with methylated H3-K36 links histone deacetylation to Pol II elongation. *Mol. Cell* **20,** 971–978 (2005).
8. Heintzman, N. D. *et al.* Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat. Genet.* **39,** 311–318 (2007).

9. Creyghton, M. P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 21931–21936 (2010).
10. Kim, T.-K. *et al.* Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465,** 182–187 (2010).
11. Lai, F. *et al.* Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. *Nature* **494,** 497–501 (2013).
12. Martens, J. H. A. *et al.* The profile of repeat-associated histone lysine methylation states in the mouse epigenome. *EMBO J.* **24,** 800–812 (2005).
13. Nakayama, J., Rice, J. C., Strahl, B. D., Allis, C. D. & Grewal, S. I. Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. *Science* **292,** 110–113 (2001).
14. Rea, S. *et al.* Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* **406,** 593–599 (2000).
15. Lachner, M., O'Carroll, D., Rea, S., Mechtler, K. & Jenuwein, T. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* **410,** 116–120 (2001).
16. Nielsen, A. L. *et al.* Selective interaction between the chromatin-remodeling factor BRG1 and the heterochromatin-associated protein HP1α. *EMBO J.* **21,** 5797–5806 (2002).
17. Brasher, S. V. *et al.* The structure of mouse HP1 suggests a unique mode of single peptide recognition by the shadow chromo domain dimer. *EMBO J.* **19,** 1587–1597 (2000).

18. Tamaru, H. & Selker, E. U. A histone H3 methyltransferase controls DNA methylation in Neurospora crassa. *Nature* **414,** 277–283 (2001).
19. Yamamoto, K. & Sonoda, M. Self-interaction of heterochromatin protein 1 is required for direct binding to histone methyltransferase, SUV39H1. *Biochem. Biophys. Res. Commun.* **301,** 287–292 (2003).
20. Kuzmichev, A., Nishioka, K., Erdjument-Bromage, H., Tempst, P. & Reinberg, D. Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. *Genes Dev.* **16,** 2893–2905 (2002).
21. Bernstein, B. E. *et al.* A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells. *Cell* **125,** 315–326 (2006).
22. Wen, B., Wu, H., Shinkai, Y., Irizarry, R. A. & Feinberg, A. P. Large histone H3 lysine 9 dimethylated chromatin blocks distinguish differentiated from embryonic stem cells. *Nat. Genet.* **41,** 246–250 (2009).
23. Yokochi, T. *et al.* G9a selectively represses a class of late-replicating genes at the nuclear periphery. *Proc. Natl. Acad. Sci.* **106,** 19363–19368 (2009).
24. Rice, J. C. *et al.* Histone Methyltransferases Direct Different Degrees of Methylation to Define Distinct Chromatin Domains. *Mol. Cell* **12,** 1591–1598 (2003).
25. Collins, R. E. *et al.* The ankyrin repeats of G9a and GLP histone methyltransferases are mono- and dimethyllysine binding modules. *Nat. Struct. Mol. Biol.* **15,** 245–250 (2008).
26. Dimitrova, E., Turberfield, A. H. & Klose, R. J. Histone demethylases in chromatin biology and beyond. *EMBO Rep.* **16,** 1620–1639 (2015).

27. Zhang, T., Cooper, S. & Brockdorff, N. The interplay of histone modifications - writers that read. *EMBO Rep.* **16,** 1467–1481 (2015).
28. Whyte, W. A. *et al.* Enhancer decommissioning by LSD1 during embryonic stem cell differentiation. *Nature* **482,** 221–225 (2012).
29. Agger, K. *et al.* UTX and JMJD3 are histone H3K27 demethylases involved in HOX gene regulation and development. *Nature* **449,** 731–734 (2007).
30. Brien, G. L. *et al.* Polycomb PHF19 binds H3K36me3 and recruits PRC2 and demethylase NO66 to embryonic stem cell genes during differentiation. *Nat. Struct. Mol. Biol.* **19,** 1273–1281 (2012).
31. Ciabrelli, F. & Cavalli, G. Chromatin-driven behavior of topologically associating domains. *J. Mol. Biol.* **427,** 608–625 (2015).
32. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485,** 376–380 (2012).
33. Vietri Rudan, M. *et al.* Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture. *Cell Rep.* **10,** 1297–1309 (2015).
34. Kind, J. *et al.* Single-cell dynamics of genome-nuclear lamina interactions. *Cell* **153,** 178–192 (2013).
35. Hark, A. T. *et al.* CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. *Nature* **405,** 486–489 (2000).
36. Yoon, B. *et al.* Rasgrf1 Imprinting Is Regulated by a CTCF-Dependent Methylation-Sensitive Enhancer Blocker. *Mol. Cell. Biol.* **25,** 11184–11190 (2005).
37. Bell, A. C., West, A. G. & Felsenfeld, G. The Protein CTCF Is Required for the Enhancer Blocking Activity of Vertebrate Insulators. *Cell* **98,** 387–396 (1999).

38. Zuin, J. *et al.* Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 996–1001 (2014).
39. Splinter, E. *et al.* CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. *Genes Dev.* **20,** 2349–2354 (2006).
40. Ong, C.-T. & Corces, V. G. CTCF: an architectural protein bridging genome topology and function. *Nat. Rev. Genet.* **15,** 234–246 (2014).
41. Narendra, V. *et al.* CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation. *Science* **347,** 1017–1021 (2015).
42. Cuddapah, S. *et al.* Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. *Genome Res.* **19,** 24–32 (2009).
43. Bell, A. C. & Felsenfeld, G. Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature* **405,** 482–485 (2000).
44. Wang, H. *et al.* Widespread plasticity in CTCF occupancy linked to DNA methylation. *Genome Res.* **22,** 1680–1688 (2012).
45. Saldaña-Meyer, R. *et al.* CTCF regulates the human p53 gene through direct interaction with its natural antisense transcript, Wrap53. *Genes Dev.* **28,** 723–734 (2014).
46. Djebali, S. *et al.* Landscape of transcription in human cells. *Nature* **489,** 101–108 (2012).
47. Kim, V. N., Han, J. & Siomi, M. C. Biogenesis of small RNAs in animals. *Nat. Rev. Mol. Cell Biol.* **10,** 126–139 (2009).

48. Carthew, R. W. & Sontheimer, E. J. Origins and Mechanisms of miRNAs and siRNAs. *Cell* **136,** 642–655 (2009).
49. Weick, E.-M. & Miska, E. A. piRNAs: from biogenesis to function. *Dev. Camb. Engl.* **141,** 3458–3471 (2014).
50. Djikeng, A., Shi, H., Tschudi, C. & Ullu, E. RNA interference in Trypanosoma brucei: cloning of small interfering RNAs provides evidence for retroposon-derived 24-26-nucleotide RNAs. *RNA N. Y. N* **7,** 1522–1530 (2001).
51. Aravin, A. A. *et al.* The small RNA profile during Drosophila melanogaster development. *Dev. Cell* **5,** 337–350 (2003).
52. Aravin, A. A. *et al.* Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the D. melanogaster germline. *Curr. Biol. CB* **11,** 1017–1027 (2001).
53. Deng, W. & Lin, H. miwi, a Murine Homolog of piwi, Encodes a Cytoplasmic Protein Essential for Spermatogenesis. *Dev. Cell* **2,** 819–830 (2002).
54. Aravin, A. A., Sachidanandam, R., Girard, A., Fejes-Toth, K. & Hannon, G. J. Developmentally Regulated piRNA Clusters Implicate MILI in Transposon Control. *Science* **316,** 744–747 (2007).
55. Grivna, S. T., Beyret, E., Wang, Z. & Lin, H. A novel class of small RNAs in mouse spermatogenic cells. *Genes Dev.* **20,** 1709–1714 (2006).
56. Girard, A., Sachidanandam, R., Hannon, G. J. & Carmell, M. A. A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature* **442,** 199–202 (2006).
57. Vagin, V. V. *et al.* A Distinct Small RNA Pathway Silences Selfish Genetic Elements in the Germline. *Science* **313,** 320–324 (2006).

58. Batista, P. J. *et al.* PRG-1 and 21U-RNAs interact to form the piRNA complex required for fertility in C. elegans. *Mol. Cell* **31,** 67–78 (2008).
59. Das, P. P. *et al.* Piwi and piRNAs act upstream of an endogenous siRNA pathway to suppress Tc3 transposon mobility in the Caenorhabditis elegans germline. *Mol. Cell* **31,** 79–90 (2008).
60. Kirino, Y. & Mourelatos, Z. Mouse Piwi-interacting RNAs are 2'-O-methylated at their 3' termini. *Nat. Struct. Mol. Biol.* **14,** 347–348 (2007).
61. Li, X. Z. *et al.* An ancient transcription factor initiates the burst of piRNA production during early meiosis in mouse testes. *Mol. Cell* **50,** 67–81 (2013).
62. Reuter, M. *et al.* Miwi catalysis is required for piRNA amplification-independent LINE1 transposon silencing. *Nature* **480,** 264–267 (2011).
63. Gou, L.-T. *et al.* Pachytene piRNAs instruct massive mRNA elimination during late spermiogenesis. *Cell Res.* **24,** 680–700 (2014).
64. Brennecke, J. *et al.* Discrete Small RNA-Generating Loci as Master Regulators of Transposon Activity in Drosophila. *Cell* **128,** 1089–1103 (2007).
65. Gunawardane, L. S. *et al.* A Slicer-Mediated Mechanism for Repeat-Associated siRNA 5' End Formation in Drosophila. *Science* **315,** 1587–1590 (2007).
66. Aravin, A. A. *et al.* A piRNA Pathway Primed by Individual Transposons Is Linked to De Novo DNA Methylation in Mice. *Mol. Cell* **31,** 785–799 (2008).
67. Fazio, S. D. *et al.* The endonuclease activity of Mili fuels piRNA amplification that silences LINE1 elements. *Nature* **480,** 259–263 (2011).

68. Kuramochi-Miyagawa, S. *et al.* DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. *Genes Dev.* **22,** 908–917 (2008).

69. Taylor, D. H., Chu, E. T.-J., Spektor, R. & Soloway, P. D. Long non-coding RNA regulation of reproduction and development. *Mol. Reprod. Dev.* **82,** 932–956 (2015).

70. Zhao, J., Sun, B. K., Erwin, J. A., Song, J.-J. & Lee, J. T. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* **322,** 750–756 (2008).

71. Rinn, J. L. *et al.* Functional Demarcation of Active and Silent Chromatin Domains in Human HOX Loci by Noncoding RNAs. *Cell* **129,** 1311–1323 (2007).

72. Wang, K. C. *et al.* A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature* **472,** 120–124 (2011).

73. Di Ruscio, A. *et al.* DNMT1-interacting RNAs block gene-specific DNA methylation. *Nature* **advance online publication,** (2013).

74. Hansen, T. B. *et al.* Natural RNA circles function as efficient microRNA sponges. *Nature* **495,** 384–388 (2013).

75. Memczak, S. *et al.* Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* **495,** 333–338 (2013).

76. Wang, Y. *et al.* Endogenous miRNA sponge lincRNA-RoR regulates Oct4, Nanog, and Sox2 in human embryonic stem cell self-renewal. *Dev. Cell* **25,** 69–80 (2013).

77. Paralkar, V. R. *et al.* Unlinking an lncRNA from Its Associated cis Element. *Mol. Cell* **62,** 104–110 (2016).

78. Volders, P. J. *et al.* An update on LNCipedia: a database for annotated human lncRNA sequences. *Nucleic Acids Res.* **43,** 4363–4364 (2015).
79. Guttman, M. *et al.* lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature* **477,** 295–300 (2011).
80. Tahiliani, M. *et al.* Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324,** 930–935 (2009).
81. Ito, S. *et al.* Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **333,** 1300–1303 (2011).
82. He, Y.-F. *et al.* Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* **333,** 1303–1307 (2011).
83. Mellén, M., Ayata, P., Dewell, S., Kriaucionis, S. & Heintz, N. MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system. *Cell* **151,** 1417–1430 (2012).
84. Yildirim, O. *et al.* Mbd3/NURD complex regulates expression of 5-hydroxymethylcytosine marked genes in embryonic stem cells. *Cell* **147,** 1498–1510 (2011).
85. Wu, T. P. *et al.* DNA methylation on N(6)-adenine in mammalian embryonic stem cells. *Nature* **532,** 329–333 (2016).
86. Lister, R. *et al.* Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462,** 315–322 (2009).
87. Lister, R. *et al.* Global epigenomic reconfiguration during mammalian brain development. *Science* **341,** 1237905 (2013).

88. Smith, Z. D. & Meissner, A. DNA methylation: roles in mammalian development. *Nat. Rev. Genet.* **14,** 204–220 (2013).

89. Jones, P. A. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat. Rev. Genet.* **13,** 484–492 (2012).

90. Sharif, J. *et al.* The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. *Nature* **450,** 908–912 (2007).

91. Okano, M., Bell, D. W., Haber, D. A. & Li, E. DNA Methyltransferases Dnmt3a and Dnmt3b Are Essential for De Novo Methylation and Mammalian Development. *Cell* **99,** 247–257 (1999).

92. Kato, Y. *et al.* Role of the Dnmt3 Family in De Novo Methylation of Imprinted and Repetitive Sequences During Male Germ Cell Development in the Mouse. *Hum. Mol. Genet.* **16,** 2272–2280 (2007).

93. Jia, D., Jurkowska, R. Z., Zhang, X., Jeltsch, A. & Cheng, X. Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. *Nature* **449,** 248–251 (2007).

94. Smallwood, S. A. & Kelsey, G. De novo DNA methylation: a germ cell perspective. *Trends Genet. TIG* **28,** 33–42 (2012).

95. Lane, N. *et al.* Resistance of IAPs to methylation reprogramming may provide a mechanism for epigenetic inheritance in the mouse. *genesis* **35,** 88–93 (2003).

96. Seisenberger, S. *et al.* The Dynamics of Genome-wide DNA Methylation Reprogramming in Mouse Primordial Germ Cells. *Mol. Cell* **48,** 849–862 (2012).

97. Bourc'his, D. & Bestor, T. H. Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature* **431,** 96–99 (2004).

98. Tsumura, A. *et al.* Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. *Genes Cells Devoted Mol. Cell. Mech.* **11,** 805–814 (2006).

99. Borgel, J. *et al.* Targets and dynamics of promoter DNA methylation during early mouse development. *Nat. Genet.* **42,** 1093–1100 (2010).

100. Feldman, N. *et al.* G9a-mediated irreversible epigenetic inactivation of Oct-3/4 during early embryogenesis. *Nat. Cell Biol.* **8,** 188–194 (2006).

101. Hathaway, N. A. *et al.* Dynamics and Memory of Heterochromatin in Living Cells. *Cell* **149,** 1447–1460 (2012).

102. Sado, T., Okano, M., Li, E. & Sasaki, H. De novo DNA methylation is dispensable for the initiation and propagation of X chromosome inactivation. *Dev. Camb. Engl.* **131,** 975–982 (2004).

103. Estève, P.-O. *et al.* Direct interaction between DNMT1 and G9a coordinates DNA and histone methylation during replication. *Genes Dev.* **20,** 3089–3103 (2006).

104. Fuks, F., Hurd, P. J., Deplus, R. & Kouzarides, T. The DNA methyltransferases associate with HP1 and the SUV39H1 histone methyltransferase. *Nucleic Acids Res.* **31,** 2305–2312 (2003).

105. Li, H. *et al.* The histone methyltransferase SETDB1 and the DNA methyltransferase DNMT3A interact directly and localize to promoters silenced in cancer cells. *J. Biol. Chem.* **281,** 19489–19500 (2006).

106. Dhayalan, A. *et al.* The Dnmt3a PWWP domain reads histone 3 lysine 36 trimethylation and guides DNA methylation. *J. Biol. Chem.* **285,** 26114–26120 (2010).

107. Lindroth, A. M. *et al.* Antagonism between DNA and H3K27 Methylation at the Imprinted Rasgrf1 Locus. *PLoS Genet* **4,** e1000145 (2008).

108. Hagarman, J. A., Motley, M. P., Kristjansdottir, K. & Soloway, P. D. Coordinate regulation of DNA methylation and H3K27me3 in mouse embryonic stem cells. *PloS One* **8,** e53880 (2013).

109. Brinkman, A. B. *et al.* Sequential ChIP-bisulfite sequencing enables direct genome-scale investigation of chromatin and DNA methylation cross-talk. *Genome Res.* **22,** 1128–1138 (2012).

110. Ooi, S. K. T. *et al.* DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature* **448,** 714–717 (2007).

111. Nan, X. *et al.* Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature* **393,** 386–389 (1998).

112. Zhang, Y., LeRoy, G., Seelig, H. P., Lane, W. S. & Reinberg, D. The dermatomyositis-specific autoantigen Mi2 is a component of a complex containing histone deacetylase and nucleosome remodeling activities. *Cell* **95,** 279–289 (1998).

113. Jeddeloh, J. A., Stokes, T. L. & Richards, E. J. Maintenance of genomic methylation requires a SWI2/SNF2-like protein. *Nat. Genet.* **22,** 94–97 (1999).

114. Kimura, H. & Shiota, K. Methyl-CpG-binding protein, MeCP2, is a target molecule for maintenance DNA methyltransferase, Dnmt1. *J. Biol. Chem.* **278,** 4806–4812 (2003).

115. Amir, R. E. *et al.* Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nat. Genet.* **23,** 185–188 (1999).

116. Kavalali, E. T., Nelson, E. D. & Monteggia, L. M. Role of MeCP2, DNA methylation, and HDACs in regulating synapse function. *J. Neurodev. Disord.* **3,** 250–256 (2011).

117. Zhang, Y. *et al.* Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. *Genes Dev.* **13,** 1924–1935 (1999).

118. Fedoroff, N. V. Presidential address. Transposable elements, epigenetics, and genome evolution. *Science* **338,** 758–767 (2012).

119. Li, W. H., Gu, Z., Wang, H. & Nekrutenko, A. Evolutionary analyses of the human genome. *Nature* **409,** 847–849 (2001).

120. Bourque, G. *et al.* Evolution of the mammalian transcription factor binding repertoire via transposable elements. *Genome Res.* **18,** 1752–1762 (2008).

121. Gifford, W. D., Pfaff, S. L. & Macfarlan, T. S. Transposable elements as genetic regulatory substrates in early development. *Trends Cell Biol.* **23,** 218–226 (2013).

122. Rowe, H. M. *et al.* KAP1 controls endogenous retroviruses in embryonic stem cells. *Nature* **463,** 237–240 (2010).

123. Bulut-Karslioglu, A. *et al.* Suv39h-dependent H3K9me3 marks intact retrotransposons and silences LINE elements in mouse embryonic stem cells. *Mol. Cell* **55,** 277–290 (2014).

124. Castro-Diaz, N. *et al.* Evolutionally dynamic L1 regulation in embryonic stem cells. *Genes Dev.* (2014). doi:10.1101/gad.241661.114

125. Carmell, M. A. *et al.* MIWI2 Is Essential for Spermatogenesis and Repression of Transposons in the Mouse Male Germline. *Dev. Cell* **12,** 503–514 (2007).

126. Nagamori, I. *et al.* Comprehensive DNA Methylation Analysis of Retrotransposons in Male Germ Cells. *Cell Rep.* **12,** 1541–1547 (2015).
127. Di Giacomo, M. *et al.* Multiple epigenetic mechanisms and the piRNA pathway enforce LINE1 silencing during adult spermatogenesis. *Mol. Cell* **50,** 601–608 (2013).
128. Beraldi, R., Pittoggi, C., Sciamanna, I., Mattei, E. & Spadafora, C. Expression of LINE-1 retroposons is essential for murine preimplantation development. *Mol. Reprod. Dev.* **73,** 279–287 (2006).
129. Kigami, D., Minami, N., Takayama, H. & Imai, H. MuERV-L is one of the earliest transcribed genes in mouse one-cell embryos. *Biol. Reprod.* **68,** 651–654 (2003).
130. Peaston, A. E. *et al.* Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos. *Dev. Cell* **7,** 597–606 (2004).
131. Göke, J. *et al.* Dynamic transcription of distinct classes of endogenous retroviral elements marks specific populations of early human embryonic cells. *Cell Stem Cell* **16,** 135–141 (2015).
132. Macfarlan, T. S. *et al.* Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature* **487,** 57–63 (2012).
133. Schmidt, D. *et al.* Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. *Cell* **148,** 335–348 (2012).
134. Muotri, A. R. *et al.* Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. *Nature* **435,** 903–910 (2005).

135. Ferguson-Smith, A. C. Genomic imprinting: the emergence of an epigenetic paradigm. *Nat. Rev. Genet.* **12,** 565–575 (2011).
136. Searle, A. G. & Beechey, C. V. Complementation studies with mouse translocations. *Cytogenet. Cell Genet.* **20,** 282–303 (1978).
137. Searle, A. G. & Beechey, C. V. Genome imprinting phenomena on mouse chromosome 7. *Genet. Res.* **56,** 237–244 (1990).
138. McGrath, J. & Solter, D. Completion of mouse embryogenesis requires both the maternal and paternal genomes. *Cell* **37,** 179–183 (1984).
139. Kono, T. *et al.* Birth of parthenogenetic mice that can develop to adulthood. *Nature* **428,** 860–864 (2004).
140. Cattanach, B. M. & Kirk, M. Differential activity of maternally and paternally derived chromosome regions in mice. *Nature* **315,** 496–498 (1985).
141. Knoll, J. H. *et al.* Angelman and Prader-Willi syndromes share a common chromosome 15 deletion but differ in parental origin of the deletion. *Am. J. Med. Genet.* **32,** 285–290 (1989).
142. Koufos, A. *et al.* Familial Wiedemann-Beckwith syndrome and a second Wilms tumor locus both map to 11p15.5. *Am. J. Hum. Genet.* **44,** 711–719 (1989).
143. Bartholdi, D. *et al.* Epigenetic mutations of the imprinted IGF2-H19 domain in Silver-Russell syndrome (SRS): results from a large cohort of patients with SRS and SRS-like phenotypes. *J. Med. Genet.* **46,** 192–197 (2009).
144. Imprinting Loci | MouseBook. Available at: http://www.mousebook.org/imprinting-gene-list. (Accessed: 30th July 2016)

145. Sleutels, F., Zwart, R. & Barlow, D. P. The non-coding Air RNA is required for silencing autosomal imprinted genes. *Nature* **415,** 810–813 (2002).
146. Santoro, F. *et al.* Imprinted Igf2r silencing depends on continuous Airn lncRNA expression and is not restricted to a developmental window. *Development* **140,** 1184–1195 (2013).
147. Mohammad, F. *et al.* Long noncoding RNA-mediated maintenance of DNA methylation and transcriptional gene silencing. *Development* (2012). doi:10.1242/dev.079566
148. Schultz, D. C., Ayyanathan, K., Negorev, D., Maul, G. G. & Rauscher, F. J. 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 Dev.* **16,** 919–932 (2002).
149. Schultz, D. C., Friedman, J. R. & Rauscher, F. J. Targeting histone deacetylase complexes via KRAB-zinc finger proteins: the PHD and bromodomains of KAP-1 form a cooperative unit that recruits a novel isoform of the Mi-2alpha subunit of NuRD. *Genes Dev.* **15,** 428–443 (2001).
150. Ryan, R. F. *et al.* KAP-1 corepressor protein interacts and colocalizes with heterochromatic and euchromatic HP1 proteins: a potential role for Krüppel-associated box-zinc finger proteins in heterochromatin-mediated gene silencing. *Mol. Cell. Biol.* **19,** 4366–4378 (1999).
151. Li, X. *et al.* A Maternal-Zygotic Effect Gene, Zfp57, Maintains Both Maternal and Paternal Imprints. *Dev. Cell* **15,** 547–557 (2008).

152. Quenneville, S. *et al.* In Embryonic Stem Cells, ZFP57/KAP1 Recognize a Methylated Hexanucleotide to Affect Chromatin and DNA Methylation of Imprinting Control Regions. *Mol. Cell* **44,** 361–372 (2011).
153. Quenneville, S. *et al.* The KRAB-ZFP/KAP1 System Contributes to the Early Embryonic Establishment of Site-Specific DNA Methylation Patterns Maintained during Development. *Cell Rep.* **2,** 766–773 (2012).
154. Nakamura, T. *et al.* PGC7 binds histone H3K9me2 to protect against conversion of 5mC to 5hmC in early embryos. *Nature* (2012). doi:10.1038/nature11093
155. Bian, C. & Yu, X. PGC7 suppresses TET3 for protecting DNA methylation. *Nucleic Acids Res.* **42,** 2893–2905 (2014).
156. Nakamura, T. *et al.* PGC7/Stella protects against DNA demethylation in early embryogenesis. *Nat. Cell Biol.* **9,** 64–71 (2006).
157. Li, J.-Y., Lees-Murdock, D. J., Xu, G.-L. & Walsh, C. P. Timing of establishment of paternal methylation imprints in the mouse. *Genomics* **84,** 952–960 (2004).
158. Watanabe, T. *et al.* Role for piRNAs and Noncoding RNA in De Novo DNA Methylation of the Imprinted Mouse Rasgrf1 Locus. *Science* **332,** 848–852 (2011).
159. Yoon, B. J. *et al.* Regulation of DNA methylation of Rasgrf1. *Nat. Genet.* **30,** 92–96 (2002).
160. Park, Y. J. *et al.* Sequences Sufficient for Programming Imprinted Germline DNA Methylation Defined. *PLoS ONE* **7,** e33024 (2012).
161. Holmes, R., Chang, Y. & Soloway, P. D. Timing and Sequence Requirements Defined for Embryonic Maintenance of Imprinted DNA Methylation at Rasgrf1. *Mol. Cell. Biol.* **26,** 9564–9570 (2006).

162. Plass, C. *et al.* Identification of Grf1 on mouse chromosome 9 as an imprinted gene by RLGS|[ndash]|M. *Nat. Genet.* **14,** 106–109 (1996).
163. Drake, N., Park, Y., Shirali, A., Cleland, T. & Soloway, P. Imprint switch mutations at *Rasgrf1* support conflict hypothesis of imprinting and define a growth control mechanism upstream of IGF1. *Mamm. Genome* **20,** 654–663 (2009).
164. Drake, N. M., DeVito, L. M., Cleland, T. A. & Soloway, P. D. Imprinted Rasgrf1 expression in neonatal mice affects olfactory learning and memory. *Genes Brain Behav.* **10,** 392–403 (2011).
165. Herman, H. *et al.* Trans allele methylation and paramutation-like effects in mice. *Nat. Genet.* **34,** 199–202 (2003).
166. Grossniklaus, U., Kelly, B., Ferguson-Smith, A. C., Pembrey, M. & Lindquist, S. Transgenerational epigenetic inheritance: how important is it? *Nat. Rev. Genet.* **14,** 228–235 (2013).
167. Heard, E. & Martienssen, R. A. Transgenerational epigenetic inheritance: myths and mechanisms. *Cell* **157,** 95–109 (2014).
168. Kaati, G., Bygren, L. O. & Edvinsson, S. Cardiovascular and diabetes mortality determined by nutrition during parents' and grandparents' slow growth period. *Publ. Online 29 Oct. 2002 Doi101038sjejhg5200859* **10,** (2002).
169. Bygren, L. O. *et al.* Change in paternal grandmothers' early food supply influenced cardiovascular mortality of the female grandchildren. *BMC Genet.* **15,** 12 (2014).
170. Chandler, V. L. Paramutation: From Maize to Mice. *Cell* **128,** 641–645 (2007).
171. Haring, M. *et al.* The role of DNA methylation, nucleosome occupancy and histone modifications in paramutation. *Plant J. Cell Mol. Biol.* **63,** 366–378 (2010).

172. Arteaga-Vazquez, M. *et al.* RNA-mediated trans-communication can establish paramutation at the b1 locus in maize. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 12986–12991 (2010).
173. Rassoulzadegan, M. *et al.* RNA-mediated non-mendelian inheritance of an epigenetic change in the mouse. *Nature* **441,** 469–474 (2006).
174. Grandjean, V. *et al.* The miR-124-Sox9 paramutation: RNA-mediated epigenetic control of embryonic and adult growth. *Development* **136,** 3647–3655 (2009).
175. Wagner, K. D. *et al.* RNA Induction and Inheritance of Epigenetic Cardiac Hypertrophy in the Mouse. *Dev. Cell* **14,** 962–969 (2008).
176. Grandjean, V. *et al.* RNA-mediated paternal heredity of diet-induced obesity and metabolic disorders. *Sci. Rep.* **5,** 18193 (2015).
177. Gapp, K. *et al.* Implication of sperm RNAs in transgenerational inheritance of the effects of early trauma in mice. *Nat. Neurosci.* **17,** 667–669 (2014).
178. Kiani, J. *et al.* RNA–Mediated Epigenetic Heredity Requires the Cytosine Methyltransferase Dnmt2. *PLoS Genet* **9,** e1003498 (2013).
179. Peng, H. *et al.* A novel class of tRNA-derived small RNAs extremely enriched in mature mouse sperm. *Cell Res.* **22,** 1609–1612 (2012).
180. Kawano, M., Kawaji, H., Grandjean, V., Kiani, J. & Rassoulzadegan, M. Novel small noncoding RNAs in mouse spermatozoa, zygotes and early embryos. *PloS One* **7,** e44542 (2012).
181. Yuan, S. *et al.* Sperm-borne miRNAs and endo-siRNAs are important for fertilization and preimplantation embryonic development. *Dev. Camb. Engl.* **143,** 635–647 (2016).

182. Chen, Q. *et al.* Sperm tsRNAs contribute to intergenerational inheritance of an acquired metabolic disorder. *Science* **351,** 397–400 (2016).
183. Sharma, U. *et al.* Biogenesis and function of tRNA fragments during sperm maturation and fertilization in mammals. *Science* **351,** 391–396 (2016).
184. Xie, W. *et al.* Base-Resolution Analyses of Sequence and Parent-of-Origin Dependent DNA Methylation in the Mouse Genome. *Cell* **148,** 816–831 (2012).
185. Howell, C. Y. *et al.* Genomic imprinting disrupted by a maternal effect mutation in the Dnmt1 gene. *Cell* **104,** 829–838 (2001).
186. Kaneda, M. *et al.* Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. *Nature* **429,** 900–903 (2004).
187. Bourc'his, D., Xu, G. L., Lin, C. S., Bollman, B. & Bestor, T. H. Dnmt3L and the establishment of maternal genomic imprints. *Science* **294,** 2536–2539 (2001).
188. Hata, K., Okano, M., Lei, H. & Li, E. Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice. *Dev. Camb. Engl.* **129,** 1983–1993 (2002).
189. Birger, Y., Shemer, R., Perk, J. & Razin, A. The imprinting box of the mouse Igf2r gene. *Nature* **397,** 84–88 (1999).
190. Kantor, B., Makedonski, K., Green-Finberg, Y., Shemer, R. & Razin, A. Control elements within the PWS/AS imprinting box and their function in the imprinting process. *Hum. Mol. Genet.* **13,** 751–762 (2004).
191. Shemer, R. *et al.* The imprinting box of the Prader-Willi/Angelman syndrome domain. *Nat. Genet.* **26,** 440–443 (2000).

192. Cranston, M. J., Spinka, T. L., Elson, D. A. & Bartolomei, M. S. Elucidation of the minimal sequence required to imprint H19 transgenes. *Genomics* **73,** 98–107 (2001).

193. Gebert, C., Kunkel, D., Grinberg, A. & Pfeifer, K. H19 imprinting control region methylation requires an imprinted environment only in the male germ line. *Mol. Cell. Biol.* **30,** 1108–1115 (2010).

194. Matsuzaki, H., Okamura, E., Shimotsuma, M., Fukamizu, A. & Tanimoto, K. A randomly integrated transgenic H19 imprinting control region acquires methylation imprinting independently of its establishment in germ cells. *Mol. Cell. Biol.* **29,** 4595–4603 (2009).

195. Park, K.-Y., Sellars, E. A., Grinberg, A., Huang, S.-P. & Pfeifer, K. The H19 differentially methylated region marks the parental origin of a heterologous locus without gametic DNA methylation. *Mol. Cell. Biol.* **24,** 3588–3595 (2004).

196. Tanimoto, K. *et al.* Genomic imprinting recapitulated in the human beta-globin locus. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 10250–10255 (2005).

197. Jameel, T. Sperm swim-up: a simple and effective technique of semen processing for intrauterine insemination. *JPMA J. Pak. Med. Assoc.* **58,** 71–74 (2008).

198. Kim, T. H. *et al.* Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell* **128,** 1231–1245 (2007).

199. McMahon, A. P. & Bradley, A. The Wnt-1 (int-1) proto-oncogene is required for development of a large region of the mouse brain. *Cell* **62,** 1073–1085 (1990).

200. Danielian, P. S., Echelard, Y., Vassileva, G. & McMahon, A. P. A 5.5-kb Enhancer Is both Necessary and Sufficient for Regulation ofWnt-1Transcriptionin Vivo. *Dev. Biol.* **192,** 300–309 (1997).

201. Echelard, Y., Vassileva, G. & McMahon, A. P. Cis-Acting Regulatory Sequences Governing Wnt-1 Expression in the Developing Mouse CNS. *Development* **120,** 2213–2224 (1994).

202. Girardot, M. *et al.* PRMT5-mediated histone H4 arginine-3 symmetrical dimethylation marks chromatin at G + C-rich regions of the mouse genome. *Nucleic Acids Res.* **42,** 235–248 (2014).

203. Delaval, K. *et al.* Differential histone modifications mark mouse imprinting control regions during spermatogenesis. *EMBO J.* **26,** 720–729 (2007).

204. Erkek, S. *et al.* Molecular determinants of nucleosome retention at CpG-rich sequences in mouse spermatozoa. *Nat. Struct. Mol. Biol.* **20,** 868–875 (2013).

205. Stoger, R. *et al.* Maternal-specific methylation of the imprinted mouse Igf2r locus identifies the expressed locus as carrying the imprinting signal. *Cell* **73,** 61–71 (1993).

206. Gosalia, N., Neems, D., Kerschner, J. L., Kosak, S. T. & Harris, A. Architectural proteins CTCF and cohesin have distinct roles in modulating the higher order structure and expression of the CFTR locus. *Nucleic Acids Res.* **42,** 9612–9622 (2014).

207. Handoko, L. *et al.* CTCF-mediated functional chromatin interactome in pluripotent cells. *Nat. Genet.* **43,** 630–638 (2011).

208. Ling, J. Q. *et al.* CTCF mediates interchromosomal colocalization between Igf2/H19 and Wsb1/Nf1. *Science* **312,** 269–272 (2006).

209. Chandler, V. L. & Stam, M. Chromatin conversations: mechanisms and implications of paramutation. *Nat. Rev. Genet.* **5,** 532–544 (2004).

210. Cooney, C. A., Dave, A. A. & Wolff, G. L. Maternal methyl supplements in mice affect epigenetic variation and DNA methylation of offspring. *J. Nutr.* **132,** 2393S–2400S (2002).

211. Weaver, I. C. G. *et al.* Epigenetic programming by maternal behavior. *Nat. Neurosci.* **7,** 847–854 (2004).

212. Dolinoy, D. C., Huang, D. & Jirtle, R. L. Maternal nutrient supplementation counteracts bisphenol A-induced DNA hypomethylation in early development. *Proc. Natl. Acad. Sci.* **104,** 13056–13061 (2007).

213. Matzke, M. A. & Mosher, R. A. RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. *Nat. Rev. Genet.* **15,** 394–408 (2014).

214. Castel, S. E. & Martienssen, R. A. RNA interference in the nucleus: roles for small RNAs in transcription, epigenetics and beyond. *Nat. Rev. Genet.* **14,** 100–112 (2013).

215. Shirayama, M. *et al.* piRNAs Initiate an Epigenetic Memory of Nonself RNA in the C. elegans Germline. *Cell* doi:10.1016/j.cell.2012.06.015

216. Ashe, A. *et al.* piRNAs Can Trigger a Multigenerational Epigenetic Memory in the Germline of C. elegans. *Cell* **150,** 88–99 (2012).

217. Thomas, A. L. *et al.* Piwi induces piRNA-guided transcriptional silencing and establishment of a repressive chromatin state. *Genes Dev.* **27,** 390–399 (2013).

218. Jin, Y., Tam, O. H., Paniagua, E. & Hammell, M. TEtranscripts: a package for including transposable elements in differential expression analysis of RNA-seq datasets. *Bioinforma. Oxf. Engl.* **31,** 3593–3599 (2015).
219. Sookdeo, A., Hepp, C. M., McClure, M. A. & Boissinot, S. Revisiting the evolution of mouse LINE-1 in the genomic era. *Mob. DNA* **4,** 3 (2013).
220. Pastor, W. A. *et al.* MORC1 represses transposable elements in the mouse male germline. *Nat. Commun.* **5,** 5795 (2014).
221. Cantone, I. & Fisher, A. G. Epigenetic programming and reprogramming during development. *Nat. Struct. Mol. Biol.* **20,** 282–289 (2013).
222. Yang, Y. W. *et al.* Essential role of lncRNA binding for WDR5 maintenance of active chromatin and embryonic stem cell pluripotency. *eLife* **3,** e02046 (2014).
223. Davidovich, C. *et al.* Toward a consensus on the binding specificity and promiscuity of PRC2 for RNA. *Mol. Cell* **57,** 552–558 (2015).
224. Kaneko, S. *et al.* Interactions between JARID2 and noncoding RNAs regulate PRC2 recruitment to chromatin. *Mol. Cell* **53,** 290–300 (2014).
225. O'Carroll, D. *et al.* The polycomb-group gene Ezh2 is required for early mouse development. *Mol. Cell. Biol.* **21,** 4330–4336 (2001).
226. Le, T. N., Miyazaki, Y., Takuno, S. & Saze, H. Epigenetic regulation of intragenic transposable elements impacts gene transcription in Arabidopsis thaliana. *Nucleic Acids Res.* gkv258 (2015). doi:10.1093/nar/gkv258
227. Waterland, R. A. & Jirtle, R. L. Transposable Elements: Targets for Early Nutritional Effects on Epigenetic Gene Regulation. *Mol. Cell. Biol.* **23,** 5293–5300 (2003).

228. Upadhyay, M. *et al.* Transposon Dysregulation Modulates dWnt4 Signaling to Control Germline Stem Cell Differentiation in Drosophila. *PLoS Genet.* **12,** (2016).
229. Cowley, M. & Oakey, R. J. Transposable elements re-wire and fine-tune the transcriptome. *PLoS Genet.* **9,** e1003234 (2013).
230. Xie, M. *et al.* DNA hypomethylation within specific transposable element families associates with tissue-specific enhancer landscape. *Nat. Genet.* **45,** 836–841 (2013).
231. Kallen, A. N. *et al.* The imprinted H19 lncRNA antagonizes let-7 microRNAs. *Mol. Cell* **52,** 101–112 (2013).
232. Wan, Y., Kertesz, M., Spitale, R. C., Segal, E. & Chang, H. Y. Understanding the transcriptome through RNA structure. *Nat. Rev. Genet.* **12,** 641–655 (2011).
233. Urich, M. A., Nery, J. R., Lister, R., Schmitz, R. J. & Ecker, J. R. MethylC-seq library preparation for base-resolution whole-genome bisulfite sequencing. *Nat. Protoc.* **10,** 475–483 (2015).
234. Rhodes, D. & Lipps, H. J. G-quadruplexes and their regulatory roles in biology. *Nucleic Acids Res.* **43,** 8627–8637 (2015).
235. Lin, J. *et al.* Stabilization of G-quadruplex DNA by C-5-methyl-cytosine in bcl-2 promoter: implications for epigenetic regulation. *Biochem. Biophys. Res. Commun.* **433,** 368–373 (2013).
236. Halder, R. *et al.* Guanine quadruplex DNA structure restricts methylation of CpG dinucleotides genome-wide. *Mol. Biosyst.* **6,** 2439–2447 (2010).

237. François, M., Leifert, W. R., Tellam, R. & Fenech, M. F. Folate deficiency and DNA-methyltransferase inhibition modulate G-quadruplex frequency. *Mutagenesis* (2016). doi:10.1093/mutage/gev088

238. Raiber, E.-A., Kranaster, R., Lam, E., Nikan, M. & Balasubramanian, S. A non-canonical DNA structure is a binding motif for the transcription factor SP1 in vitro. *Nucleic Acids Res.* **40,** 1499–1508 (2012).

239. Shechner, D. M., Hacisuleyman, E., Younger, S. T. & Rinn, J. L. Multiplexable, locus-specific targeting of long RNAs with CRISPR-Display. *Nat. Methods* **12,** 664–670 (2015).

240. Skourti-Stathaki, K. & Proudfoot, N. J. A double-edged sword: R loops as threats to genome integrity and powerful regulators of gene expression. *Genes Dev.* **28,** 1384–1396 (2014).

241. Ginno, P. A., Lott, P. L., Christensen, H. C., Korf, I. & Chédin, F. R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. *Mol. Cell* **45,** 814–825 (2012).

242. Skourti-Stathaki, K., Kamieniarz-Gdula, K. & Proudfoot, N. J. R-loops induce repressive chromatin marks over mammalian gene terminators. *Nature* **516,** 436–439 (2014).

243. Vagin, V. V. *et al.* Proteomic analysis of murine Piwi proteins reveals a role for arginine methylation in specifying interaction with Tudor family members. *Genes Dev.* **23,** 1749–1762 (2009).

244. Watanabe, T. *et al.* MITOPLD Is a Mitochondrial Protein Essential for Nuage Formation and piRNA Biogenesis in the Mouse Germline. *Dev. Cell* **20,** 364–375 (2011).

245. Laszlo, A. H. *et al.* Detection and mapping of 5-methylcytosine and 5-hydroxymethylcytosine with nanopore MspA. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 18904–18909 (2013).

246. Flusberg, B. A. *et al.* Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat. Methods* **7,** 461–465 (2010).

247. Motorin, Y. & Helm, M. RNA nucleotide methylation. *Wiley Interdiscip. Rev. RNA* **2,** 611–631 (2011).

248. Clift, D. & Schuh, M. Restarting life: fertilization and the transition from meiosis to mitosis. *Nat. Rev. Mol. Cell Biol.* **14,** 549–562 (2013).

249. Muotri, A. R. *et al.* L1 retrotransposition in neurons is modulated by MeCP2. *Nature* **468,** 443–446 (2010).

250. Zhang, B. & Horvath, S. A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol. Biol.* **4,** Article17 (2005).

251. Horvath, S. *et al.* Analysis of oncogenic signaling networks in glioblastoma identifies ASPM as a molecular target. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 17402–17407 (2006).

252. Katz, Y., Wang, E. T., Airoldi, E. M. & Burge, C. B. Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nat. Methods* **7,** 1009–1015 (2010).

253. Shalgi, R., Pilpel, Y. & Oren, M. Repression of transposable-elements - a microRNA anti-cancer defense mechanism? *Trends Genet. TIG* **26,** 253–259 (2010).