# MECHANISMS AND CONSEQUENCES OF DNA METHYLATION IN TWO MODEL SPECIES

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

Erin Tsi-Jia Chu

December 2017

# MECHANISMS AND CONSEQUENCES OF DNA METHYLATION IN TWO MODEL SPECIES

Erin Tsi-Jia Chu, D.V.M., Ph. D.

Cornell University 2017

## DOCTORAL ABSTRACT

Epigenetic modifications are known to regulate gene expression in a heritable manner, and can broadly be divided into three interacting classes: DNA methylation, histone modifications, and chromatin interactions. While the *trans* acting factors that establish, maintain, and remove DNA methylation are well-known, the *cis* acting mechanisms that direct DNA methylation to specific genomic locations remain elusive. Two gene classes offer insights into *cis*-acting mechanisms for DNA methylation: imprinted loci, and transposable elements. A locus spanning both is the murine *Rasgrf1* locus. *Rasgrf1* has a *cis* element, a series of tandem repeats, required for DNA methylation, but also harbors a long noncoding RNA, the pitRNA. The pitRNA is driven by the repeats and is targeted by piRNAs, small RNAs that mediate transposable element methylation in the mammalian male germline. However, the effects of the pitRNA versus the repeats have not yet been separated. My work, where I used CRISPR/Cas9 genome editing to generate a targeted mutant system permitting inducible control of the pitRNA, is the first to query the sufficiency of the pitRNA independently. Using quantitative qPCR and targeted bisulfite sequencing, I demonstrated that expression of the lncRNA at physiological levels in the male germline is insufficient to impart DNA methylation at

*Rasgrf1*. These findings were complimented by additional *in vitro* studies, where I identified Sp1 as a transcription factor that binds the repeats and is required for pitRNA expression. Sp1 binds secondary DNA structure and has recently been identified as a regulating factor at another imprinted gene. Together, these findings support an alternative, critical role for the repeats beyond their known role in regulating pitRNA expression.

Beyond mechanism, DNA methylation in the context of disease are an area of active study, though its utility in non-traditional model organisms is nascent. The second focus of my thesis speaks to this. I performed reduced representation bisulfite analysis on two dog breeds with highly diverse morphology and disease risks. While this work is largely preliminary, two differentially methylated regions have direct association with differential disease risk between these two breeds, suggesting that the canine methylome could be used as method of disease surveillance.

## BIOGRAPHICAL SKETCH

Erin Chu was born on Thanksgiving Day, November 24th, 1988 to Ying and Randy Chu in Poughkeepsie, New York. She attended schools in three countries before graduating from the International School of Beijing with an International Baccalaureate She returned to upstate New York for her undergraduate education at Cornell University, where she first became interested in basic science research. In 2008, she began work as an undergraduate research assistant in the lab of Dr. Nate Sutter. She continued her work in the Sutter Lab as a 2008 Hughes Summer Scholar investigating the genetics of body size diversity in the domestic dog, then laying the foundation for a study examining the genetics of body size diversity in the domestic horse in collaboration with the lab of Dr. Samantha Brooks through 2010. In the summer of 2009, she was a Paul W. Zuccaire Summer Fellow in the lab of Dr. Phil Avner at the Institut Pasteur in Paris, France. Here, she began to explore the field of epigenetics. Erin graduated with her Bachelor of Science in Biological Sciences, Magna Cum Laude with Honors in Research, in 2010 and matriculated to the Cornell University New York State College of Veterinary Medicine as a Combined DVM/PhD student. She graduated with her Doctor of Veterinary Medicine in 2014, and completed her PhD in 2017 in the lab of Dr. Paul Soloway. Following her graduation, Erin is pursuing her goal to work in biotechnology research and development as a Senior Veterinary Geneticist at Embark Veterinary, Inc.

“But I don’t want to go among mad people,” Alice remarked.

“Oh, you can’t help that,” said the Cat:

“We’re all mad here. I’m mad. You’re mad.”

“How do you know I’m mad?” said Alice.

“You must be, said the Cat, “or you wouldn’t have come here.”

- Lewis Carroll, *Alice in Wonderland*

ACKNOWLEDGEMENTS

I am lucky enough to have met some truly wonderful friends, colleagues, and mentors during my time at Cornell.

My advisor, Dr. Paul Soloway, was the reason I got into epigenetics to begin with. Thank you for always giving it your best.

My excellent Graduate Committee, Drs. Paula Cohen, Natasza Kurpios, and John Schimenti, has been incredibly communicative and supportive throughout my graduate education. I thank them for their valuable insights and advice.

Roman Spektor has been the best and worst person to share an office wall, lab space, and brain with for the past three years.

I have never met a kinder colleague than Dr. David Taylor, and have missed you in the office this year. The plants do, too.

I am thankful to have known Jim Putnam, Drs. John McElwee, Byong-Ryul Hyun, Jim Hagarman, Jonathan Flax, and Patrick Murphy. My hardworking undergraduates Ephrath Tesfaye, Margo Motulsky, and Meg Hofstedt have produced great work and have been the best way for me to stay up to date on pop culture.

I am so proud of all of my comrades in Cornell's Combined DVM/PhD Program. In particular, I could not have done this without my year-mate and partner-in-crime, Dr. Jess Brown.

From my veterinary training: Drs. Evin Adolph '14, Jenn Romano Bucholtz '14, and Jamie Zhen '14 were the only reason I looked forward to studying for finals, and got me through some tough times. I also thank the veterinary faculty and staff, who in my opinion are Cornell Vet's greatest asset. Drs. Brian Collins, Paul Maza, Ryane Englar, Dorothy Ainsworth, and Araceli

Lucio-Forster have been equal parts mentors and friends. The amazing smiles and hugs of Jean Mint and Laura Hobbes, Margie Vail, and Molly Cooper have sustained me.

Being at Cornell for as long as I've been, I must also thank the Office of Undergraduate Biology and all the amazing undergraduate staff that have stayed in touch over the past eleven (!) years: Pamela Davis Acey, Laurel Southard, and Wendy Aquadro have always had a smile and some great advice for me.

My gym family, Dr. Casey Garland, Liv Weihe, and everyone at Ultimate Athletics, were why I left lab at 5:30 most weekdays.

My barn family has saved me more times than I can count. Thank you, Helen Evans and Dannielle Cornelius, for being my mothers away from home, and Sarah Morris and Sophia Rocco for countless adventures.

My future husband, Dr. Evan Buntrock, perpetually challenges and balances me. It bodes for an interesting life together.

I have known my best girls Veronica Fischmann, Sarah Yang, Kirsten Sargent, and Marin Cherry for what seems like a lifetime. How lucky am I to have had you on my side all these years?

My four-legged companions Shiloh, Honor, and Clementiger have been a source of constant comfort and consternation.

Finally, my mother, Ying Chu, my godparents, Chung and Jim McArdle, and my cousin, Brendan McArdle, are the light of my life. I have not lived a day without knowing how much you love me, and how proud you are of me. Thank you for supporting me through it all.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES



.

# LIST OF ABBREVIATIONS

5mC – 5'methylcytosine
5hmC – 5'hydroxymethylcytosine
5fC – 5'formylcytosine
5caC – 5'carboxylcytosine
AGO2 – Argonaute RISC Catalytic Component 2
AID/APOBEC – Activation Induced Cytosine Demininase/Apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like
*ANRIL/CDKN2B-AS1* – CDKN2B Antisense RNA 1
AR – androgen receptor
*Airn* – Antisense Of IGF2R Non-Protein Coding RNA
ATP – adenosine triphosphate
*BCAR4* – Breast Cancer Anti-Estrogen Resistance 4
BLIMP/PRDM1 – PR Domain Containing 1, With ZNF Domain
BOLL/BOULE – Boule-Like RNA-Binding Protein
CBX5/HP1 – Chromobox Homolog 5 / Heterochromatin Protein 1
CBX7 – Chromobox Homolog 7
CC – cumulus cells
CEBPA – CCAAT/Enhancer Binding Protein (C/EBP), Alpha
CENPA – Centromere Protein A
ceRNA - competing endogenous RNA
ChIP – Chromatin Immunoprecipitation
circRNA – circular RNA
CLIP – cross-linking and immunoprecipitation assay
COC – cumulus-oocyte complex
CSF – Codon Substitution Frequency
*CTBP1-AS1* – C-Terminal Binding Protein 1 antisense transcript 1
CTCF – CCCTC-Binding Factor
DAZ – Deleted In Azoospermia
DAZ1 – Deleted In Azoospermia 1
DAZL – Deleted In Azoospermia-Like
DDX4/VASA – DEAD (Asp-Glu-Ala-Asp) Box Polypeptide 4
DMRT1 – Doublesex And Mab-3 Related Transcription Factor 1
DNA – Deoxyribonucleic acid
dpc – days post coitum
dpp – days post partum
DNMT1 – DNA (Cytosine-5-)-Methyltransferase 1
DNMT2 – tRNA Aspartic Acid Methyltransferase 1
DNMT3B – DNA (Cytosine-5-)-Methyltransferase 3 Beta
DROSHA – Drosha, Ribonuclease Type III
ecRNA - extra coding RNA
ESR1 – estrogen receptor 1
eRNA – enhancer RNA
ESC – embryonic stem cells
EZH2 – Enhancer Of Zeste 2 Polycomb Repressive Complex 2 Subunit
*Fendrr* – FOXF1 Adjacent Non-Coding Developmental Regulatory RNA
*FGFR2* – Fibroblast Growth Factor Receptor 2

*FIRRE* – *Firre* Intergenic Repeating RNA Element
FSH/LH – follicle stimulating hormone / leutenizing hormone
*GAS5* – Growth Arrest-Specific 5
GFP – green fluorescent protein
H3K27me3 – histone H3 lysine 27 trimethylation
H3K36me3– histone H3 lysine 36 trimethylation
H3K4me3– histone H3 lysine 4 trimethylation
H3K9me2 – histone H3 lysine 9 dimethylation
H3K9me3– histone H3 lysine 9 trimethylation
HDAC1 – Histone deacetylase 1
HDAC2 – Histone deacetylase 2
HDAC3 – Histone deacetylase 3
HELLP – hemolysis, elevated liver enzymes, and low platelets
HGRNPU/SAF-A – Heterogeneous Nuclear Ribonucleoprotein U (Scaffold Attachment Factor A)
HJURP – Holliday Junction Recognition Protein
*HOXC* – Homeobox C cluster
*HOXD* – Homeobox D cluster
*HOTAIR* – HOX Transcript Antisense RNA
ICR – Imprinting Control Region
*Igf2r* – Insulin-Like Growth Factor 2 Receptor
IME1 – inducer of meiosis 1
IME4 – inducer of meiosis 4
*IRT1* – IME1 Regulatory Transcript
IUGR – intrauterine growth restriction
*Kcnq1* – Potassium Channel, Voltage Gated KQT-Like Subfamily Q, Member 1
*Kcnq1ot1* - KCNQ1 Opposite Strand/Antisense Transcript 1
KIT – V-Kit Hardy-Zuckerman 4 Feline Sarcoma Viral Oncogene Homolog
KLK2 – Kallikrein-Related Peptidase 2
KLK3 – Kallikrein-Related Peptidase 3
KDM1A/LSD1 - Lysine (K)-Specific Demethylase 1A
LBR – Lamin B Receptor
lincRNA – long intergenic noncoding RNA
lncRNA – long noncoding RNA
m6A – N6-methyladenine
*lincMD1* – Long Intergenic Non-Protein Coding RNA, Muscle Differentiation 1
*MALAT1* - Metastasis Associated Lung Adenocarcinoma Transcript 1
Mei2 – RNA-binding protein involved in meiosis
*MesP1* – Mesoderm Posterior Basic Helix-Loop-Helix Transcription Factor 1
METTL3 – Methyltransferase Like 3
*MHRT* – Myosin Heavy Chain-Associated RNA Transcript
MII – meiosis II
miRNA – micro RNA
MLL – Mixed-Lineage Leukemia
*Mmi1* – Meiotic mRNA Interception
MOP1 – Modifier of paramutation 1
MOV10L – Mov10 RISC Complex RNA Helicase Like 1
*NEAT1* – Nuclear Paraspeckle Assembly Transcript 1
ncRNA – noncoding RNA
NOMO1 – NODAL Modulator 1

NR – Nuclear receptor
nt – nucleotides
pancRNA – promoter associated noncoding RNA
*PANDA* – Promoter Of CDKN1A Antisense DNA Damage Activated
PBMC – Peripheral blood mononuclear cell
PCGEM1 – Prostate-Specific Transcript
PCOS – polycystic ovary syndrome
PCR – polymerase chain reaction
PGC – primordial germ cell
PGR – Progesterone receptor
piRNA – piwi interacting RNA
pitRNA – piRNA targeted RNA
PIWI – P-element induced wimpy testes
PIWIL1/MIWI – P-element induced wimpy testes like 1
PIWIL2/MILI – P-element induced wimpy testes like 2
Poly(A)$^{+}$ – polyadenylated
Poly(A)$^{-}$ – not polyadenylated
PPROM – premature rupture of the placental membranes
PR – progesterone receptor
PRC2 – polycomb repressive complex 2
PRMD14 – PR Domain Containing 14
PRNCR1 – Prostate Cancer Associated Non-Coding RNA 1
QPCR – Quantitative polymerase chain reaction
*Rasgrf1* – Ras Protein-Specific Guanine Nucleotide-Releasing Factor 1
RCOR1/CoREST – REST Corepressor 1
*RepA* – Repeat A of *Xist*
RME1 – Regulator of Meiosis 1
RME2 – Regulator of Meiosis 2
RNA – Ribonucleic acid
RoR – Regulator Of Reprogramming
rRNA – ribosomal RNA
RRBS – reduced representation bisulfite sequencing
*SCHLAP1* – SWI/SNF Complex Antagonist Associated With Prostate Cancer 1
SET3 – SET domain-containing 3
SHARP/SPEN – Spen Family Transcriptional Repressor
shRNA – short hairpin RNA
SINEB2 – short interspersed nuclear element B2
siRNA – small interfering RNA
SIX3 – SIX Homeobox 3
*Slc22a2* – Solute Carrier Family 22 (Organic Cation Transporter), Member 2
*Slc22a3* – Solute Carrier Family 22 (Organic Cation Transporter), Member 3
SMARC4A/BRG1 – SWI/SNF Related, Matrix Associated, Actin Dependent Regulator Of Chromatin, Subfamily A, Member 4
SMARCB1– SWI/SNF Related, Matrix Associated, Actin Dependent Regulator Of Chromatin, Subfamily B, Member 1
sme3 – Suppressor of MEi2
SMRT/NCOR2 – Nuclear Receptor Corepressor 2
snoRNA – small nucleolar RNA
snRNA – small nuclear RNA
SOX9 – SRY (Sex Determining Region Y)-Box 9

*SPRY4-IT1* – SPRY4 Intronic Transcript 1
SR – Splicing Factors, Arginine/Serine-Rich
*SRA* – Steroid Receptor RNA Activator
SRAP/SRA1 – Steroid Receptor RNA Activator 1
SRY – Sex Determining Region Y
*SWI/SNF* – Switch / Sucrose non-fermenter
*Sxl* – Sex lethal
TBCA – Tubulin cofactor A
TEI – transgenerational epigenetic inheritance
TET – Ten eleven translocase
TFAP2C/AP2γ – Transcription Factor AP-2 Gamma (Activating Enhancer Binding Protein 2 Gamma)
*tie1AS* – Tyrosine Kinase With Immunoglobulin-Like And EGF-Like Domains 1 antisense transcript
tRNA – transfer RNA
*TUNAR* – TUNA (Tcl1 Upstream Neuron-Associated lincRNA
uaRNA – UTR associated RNA
*Uchl1* – Ubiquitin Carboxyl-Terminal Esterase L1 (Ubiquitin Thiolesterase)
uPGR – unliganded progesterone receptor
UTR – untranslated region
WRD5 – WD Repeat Domain 5
*Xist* – Inactive X-chromosome specific transcript
*ZFAS1* – *ZNFX1* Antisense RNA 1

# I INTRODUCTION

## I.A Epigenetic state: DNA methylation

Epigenetic state is defined as heritable patterns of gene expression without changes to the DNA sequence. A hallmark epigenetic modification is DNA methylation, which in mammals occurs in a CpG dinucleotide context. The methylome is established early during embryogenesis and changes dynamically throughout early development, accompanying diverging cell types—indeed, mature tissues, even composed of heterogenous cell types, can be distinguished by differential methylation (Ziller *et al*, 2013; Rakyan 2008).

Genome-wide, the DNA methylome remains fairly stable throughout the lifetime of an organism (reviewed in Bestor *et al*, 2015). However, changes to the methylome accompany the aging process (Chen *et al*, 2016; Horvath *et al*, 2013; Hannum *et al*, 2013) as well as disease states. Perhaps the best known link between the DNA methylome and disease can be found in the cancer epigenetics field. For example, global hypomethylation of benign and malignant colorectal cancer samples was observed in the 1980s (Goelz *et al*, 1985; Feinberg and Vogelstein, 1983a). Concerning specific genes, hypomethylation of the c-Ha-ras and c-kit-ras oncogenes was observed in colorectal adenocarcinomas and a small cell lung carcinoma compared to normal adjacent tissues (Feinberg and Vogelstein 1983b). Hypermethylation of tumor suppressor, including *p16/CDKN* (Herman *et al*, 1995; Gonzalez-

Zulueta *et al*, 1995; Merlo *et al*, 1995) and *VHL* (Herman *et al*, 1994).

Though it remains unclear whether DNA methylation changes are causative or consequential of many diseases, other examples that link DNA methylation changes can be found in **Table I.1**.

**Table I.1. DNA methylation changes in prevalent diseases.** The first five diseases listed are among the top ten causes of death worldwide (World Health Organization 2017). PBMCs, peripheral blood mononuclear cells.

| **Disease** | **Perturbance (Reference)** |
|---|---|
| Ischemic heart disease and stroke | Genome wide methylome changes including many factors for angiogenesis and vascular remodeling (Zaina *et al*, 2014); LINE1 element hypomethylation (Baccarelli *et al*, 2010); MCT3 hypermethylation (Zhu *et al*, 2005) |
| Chronic obstructive pulmonary disease | Aberrant methylation of immune and inflammatory response genes (Qiu *et al*, 2011; Vucic *et al*, 2014) |
| Diabetes mellitus | Whole methylome changes in diabetic pancreatic islets (Volkmar *et al*, 2012; Dayeh *et al*, 2014) and PBMCs (Toperoff *et al*, 2011) |
| Alzheimer's Disease | Early changes in *ANK1*, *BIN1*, *RHBDF2* methylation (De Jager *et al*, 2014); global, age-related changes in human brain (Coppieters *et al*, 2013) |
| Male infertility | Loss of imprinting in sperm (Hammoud *et al*, 2010; Kobayashi *et al*, 2007; Kobayashi *et al*, 2009); Global hypermethylation (Houshdaran *et al*, 2007) |
| Asthma | Hypermethylation of Th1 and Th2 cytokine genes (Brand *et al*, 2012); hypomethylation of IL-4 and IFNgamma promoters (Kwon *et al*, 2008); Genetic-epigenetic interactions at *IL4R* (Soto-Ramirez *et al*, 2012) |
| Obesity | Differential methylation of *CLOCK, PER2, BMAL1* (Milagro *et al*, 2012); differential methylation of obesity-associated genes following gastric bypass (Benton *et al*, 2015); differential methylation associated with weight loss response (Moleres *et al*, 2013) |

Following the association of methylation perturbation to neoplasia, hallmark experiments demonstrated the growth suppressive effects of 5'azadeoxycytidine (5'azadC or decitabine), a nucleoside analog known to inhibit DNA methyltransferases or DNMTs (Santi *et al*, 1983). Other nucleoside inhibitors of DNA methylation include 5'fluro-2'-deoxycytidine and zebularine (Cheng *et al*, 2004), both currently in Phase I-III clinical trials

(clinicaltrials.gov). Non-nucleoside DNA methylation inhibitors include procainamide and derivatives as well as the turmeric-derived curcumin (Villar-Garea *et al*, 2003; Liu *et al*, 2009); the discovery and development of therapeutic DNA methylation inhibitors remains an area of active research.

The fact that DNMTs themselves do not recognize specific genetic sequences also limits the specificity of DNMT inhibitor therapy. However, DNA methylation *in vivo* is targeted to specific genomic sequences in a time- and cell-specific manner. How this is regulated is the central focus of this thesis.

In Section I.A and I.B, the *trans*-acting factors required for the establishment, maintenance, and removal of DNA methylation, as well as effectors of histone modifications and chromatin remodeling, are discussed. Section I.C discusses the dynamic changes in methylation during early embryonic development. Section I.D expands on two classes of genes that are specifically targeted for methylation, imprinted genes and transposable elements, as well as the specialized mechanisms and factors known to target and maintain methylation at these loci. Finally, in Section IE, findings from *Rasgrf1,* a model system for the mechanisms underlying DNA methylation, are discussed.

### I.A.1 Establishing, removing, and maintaining epigenetic state

In this section, the factors that establish and remove DNA methylation are discussed; namely, the DNA methyltransferases (DNMTs), the Ten-Eleven-Translocation (TET), and Activation Induced Cytosine Demininase (AID)/Apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like

(APOBEC) enzymes. Enzymes required for histone modifications and chromatin state and how they interact with DNA methylation are also described.

### I.A.1.i Establishment of DNA methylation: the *de novo* DNA methyltransferases Dnmta, Dnmt3b, and Dnmt3c

Dnmt3a and Dnmt3b, the canonical mammalian de novo DNA methyltransferases, are essential for *de novo* methylation genome-wide (Okano *et al,* 1999). Their roles in the establishment of DNA methylation at imprinted loci, however, are distinct. Imprinting and imprinted loci will be discussed at length in Section I.D.

*Dnmt3a* null mice are normal at birth, but fail to thrive and die by post-partum day 28 (p28). Offspring of germline-conditional *Dnmt3a* knockout females die *in utero*, displaying hypomethylation and loss of methylation at the maternally methylated *Snrpn, Igf2r, and Peg1* DMRs. In comparison, germline-conditional Dnmt3a knockout males are sterile, with hypomethylation of the DMRs for *H19*/*IGF2* and *Dlk1-Dio3* loci (Kaneda *et al,* 2004).

*Dnmt3b* knockout animals are embryonic lethal by e11.5 displaying multiple developmental abnormalities. Mutations in *Dnmt3b* are associated with human Immunodeficiency, Centromeric instability, and Facial abnormality (ICF) syndrome, which includes a loss of pericentric DNA methylation (Hansen *et al,* 1999; Xu *et al,* 1999, Jin *et al,* 2007). Germline-conditional *Dnmt3b* knockouts have no overt phenotype. Investigation of paternally methylated DMRs in Dnmt3b-deficient prospermatogonia revealed minimal effects on

most paternal DMRs examined. A significant reduction in methylation is reported at the *Rasgrf1* DMR, but is incomplete (roughly 75% methylated relative to wild type), and was only observed in some biological replicates (Kato *et al,* 2007).

Dnmt3l, a germline-specific, non-catalytic isoform, increases the catalytic activities of both Dnmt3a and Dnmt3b approximately fifteen fold (Gowher *et al,* 2005). diverse roles in the female and male germlines. *Dnmt3l* knockout females are viable and fertile; however, pups from a *Dnmt3l* knockout female die *in utero*, displaying massive dysregulation of imprinted loci (Bourc'his *et al*, 2001). In comparison, *Dnmt3l* knockout males are viable, but are sterile, with a hypogonadism phenotype. Evaluation of Dnmt3l-deficient male germ cells reveal hypomethylation and reactivation of LTR and non-LTR class transposable elements (Bourc'his *et al*, 2004), mobile genetic elements that can insert themselves into other locations in the host genome (McClintock 1950; discussed in Section I.C). Dnmt3l is thought to act downstream of the piRNA pathway, as piRNAs are still detectable in Dnmt3l null males (Aravin *et al,* 2008). The piRNA pathway will be discussed in the following section.

The recently discovered Dnmt3c isoform is male germline-specific and is thought target certain classes of evolutionarily young transposable elements. Whole genome bisulfite sequencing (WGBS) of *Dnmt3c* knockout animals revealed a minimal change in whole genome methylation; however, significantly hypomethylated regions consistently overlapped with annotated LINEs and ERV elements (Barau *et al,* 2016; Jain *et al,* 2017).

Perhaps of greatest relevance to this thesis, WGBS revealed that the *Rasgrf1* DMR was consistently hypomethylated in *Dnmt3c* knockouts. This is consistent with prior reports that knockout of neither *Dnmt3a* nor *Dnmt3b* were sufficient to ablate methylation at the *Rasgrf1* DMR. As will be discussed in later sections and chapters, how Dnmt3c is targeted to *Rasgrf1* likely involves other known mechanisms at play at *Rasgrf1* including the piRNA pathway, specific histone marks, or proteins recruited by specific chromatin topologies.

### I.A.1.ii Maintenance of DNA methylation: Dnmt1 and isoforms

The ubiquitous Dnmt1 binds and methylates hemimethylated DNA, which typically arises after replication of methylated DNA (Bestor 1992). Dnmt1 requires *UHRF1* to recognize and bind hemimethylated DNA (Bostick *et al*, 2007; Sharif *et al,* 2007; Fang *et al*, 2016), then methylates the opposing unmethylated strand (Bashtrykov *et al*, 2012). As such, Dnmt1 is indispensable for the maintenance of methylation during DNA replication. *Dnmt1* null animals are embryonic lethal by e10.5, with multiple developmental abnormalities and genome-wide hypomethylation (Li *et al,* 1992).

Mouse oocytes and preimplantation embryos lack full length Dnmt1, but express a truncated isoform, Dnmt1o. Dnmt1o is largely sequestered in the ooplasm, but is rapidly trafficked to the nucleus during the oocyte growth phase, where maternal imprints are established, and during the eight-cell stage of embryogenesis. While Dnmt1o knockout animals are viable and fertile; their oocytes exhibit appropriate maternal genomic imprinting. However, embryos arising from these oocytes display morphological abnormalities,

defective genomic imprinting, and placental defects, often dying in the third trimester (Cirio *et al*, 2008; Howell *et al*, 2001; McGraw *et al*, 2013). Methylation analysis of Dnmt1o null oocytes indicate that maternal imprints are appropriately established; however, imprinted methylation is lost postzygotically (Howell *et al,* 2001).

An additional isoform, Dnmt1p, is expressed only in pachytene spermatocytes, but is not thought to produce a catalytically active protein (Ko *et al*, 2005). Its function, if any, is unknown.

Many proteins are known to bind methylated DNA and aid in the maintenance of a repressive chromatin state. The methyl-CpG binding protein (MBD) family composes six known family members, with MeCP2 being the first characterized family member (Meehan *et al*, 1989; Lewis *et al*, 1992). MeCP2 maintains a transcriptionally repressed chromatin state through association with repressive histone complexes (Nan *et al*, 1998; Jones *et al*, 1998); mutations in MeCP2 have been associated with the neurologic condition Rett Syndrome (Amir *et al*, 1999). MBD2 and MBD3 also co-purify with the Nucleosome Remodeling and histone Deacetylation (NuRD) complex (Zhang *et al*, 1999), a chromatin remodeling complex that aids in maintaining transcriptional repression (dicussed in Section I.B).

#### I.A.1.iii Active and passive mechanisms for DNA demethylation

As will be described in Section I.C, methylation of the embryonic genome is rapidly removed post-fertilization. Demethylation of the embryonic genome could feasibly be achieved via two mechanisms: First, through active

replacement of 5mC or derivatives with an unmodified cytosine, known as active demethylation, and second, through failure to maintain 5mC marks through DNA replication (also known as passive or replication-coupled demethylation).

*Passive DNA Demethylation*: Passive DNA methylation involves the absence or inhibition of Dnmt1 activity, which is thought to occur by a variety of mechanisms. In pre-implantation embryos, Dnmt1 is largely restricted to the cytoplasm (Ratnam *et al*, 2002), which likely assists in the gradual dilution of 5mC in the embryo proper. UHRF1 also binds 5hmC at 10% the efficiency of 5mC and DNMT1 activity is reduced at sites of 5hmC compared to 5mC (Hashimoto *et al*, 2012) and is downregulated in PGCs (Kagiwada *et al*, 2012).

*Active DNA Demethylation*: In mammals, no double-stranded DNA demethylases are characterized. Rather, two mechanisms for active DNA demethylation have been described. First, oxidation of the methylated cytosine (5mC) by the TET enzymes followed by base excision repair, and second, cytosine deamination by the AID/APOBEC proteins followed by base excision repair.

The TET proteins oxidize 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC), 5-formycytosine (5fC), and 5-carboxylcytosine (5caC) (Ito *et al*, 2011). Three proteins compose the TET family, TET1, TET2, and TET3. TET3 is expressed at its highest levels in the zygote and is required for demethylation of the paternal pronucleus (Gu *et al*, 2011, Wossidlo *et al*, 2011). TET1 is highly expressed in the preimplantation

embryo, and is required for demethylation of imprinted loci in the embryonic germline (Yamaguchi *et al*, 2012; Yamaguchi *et al*, 2013). Depletion of TET1 in ES cells skews the developmental potential of preimplantation embryos and ES cells (Ito *et al*, 2010; Koh *et al,* 2011). Oxidized derivatives of 5mC are targets by the base-excision repair pathway (Tahiliani *et al*, 2009; reviewed in Pastor *et al*, 2013); namely, the enzyme Thymidine Glycosylase (TDG) recognizes and excises 5fC and 5caC (reviewed in Kohli and Zhang, 2013). TDG knockout or inhibition lead to an embryonic lethal phenotype with multiple developmental abnormalities; methylation analysis revealed hypermethylation at CpG islands (Cortellino *et al*, 2011).

An additional role for the TET enzymes was recently characterized with the discovery of a somatically expressed TET1 isoform, TET1s. TET1s is preferentially expressed in adult somatic cells compared to the full length TET1e isoform, which as previously stated is expressed in embryonic stem cells and primordial germ cells. Tet1s displays a significantly reduced global chromatin binding profile compared to Tet1e, and is insufficient to erase imprints in PGCs (Zhang *et al*, 2016). Preferential expression of Tet1s in the soma likely also contributes to somatic maintenance of DNA methylation, as a reduced inclination to oxidize 5mC to the excisable 5hmC and other derivatives would in theory increase the likelihood of existing 5mC. Other maintenance mechanisms for DNA methylation were discussed in Section I.A.1.ii.

The Activation-induced cytosine deaminase (AID)/APOBEC enzymes

were discovered prior to the TET enzymes, and catalyze the deamination of cytidine to uridine. This was first shown to be coupled to genomic demethylation in zebrafish (Rai *et al*, 2008). AID is implicated in somatic mechanisms including antibody diversification (Murumatsu *et al*, 1999; Murumatsu *et al*, 2000) and cancer metastasis (Munoz *et al*, 2013). AID/APOBECs have significantly less affinity for 5mC compared to unmodified cytosine (Nabel *et al*, 2012) and deaminates 5mC at a faster rate than 5hmC, 5fC, and 5caC, which could argue against a significant role for AID/APOBEC in active demethylation. Nevertheless, AID-deficient PGCs are hypermethylated three-fold over wild type (Popp *et al*, 2010), and reprogramming of somatic cells for induced pluripotent stem cells requires AID (Bhutani *et al*, 2009); as such, AID activity could certainly contribute to demethylation of the early embryo and the germline.

## I.B Epigenetic regulations beyond DNA methylation

While DNA methylation is the focus of this thesis, this epigenetic modification is inextricably intertwined with other forms of epigenetic regulation. As such, a discussion of known histone modifications, chromatin remodeling complexes, and chromatin architecture is warranted. Where possible, known interactions with DNA methylation are included.

### I.B.1 Histone modifications and effectors

Histones represent a family of proteins around which DNA is wrapped for packaging into a basic unit, the nucleosome, in nearly all cells with the exception of male germ cells, which exchange nucleosomes for protamines

late in spermiogenesis (reviewed in Balhorn, 2007). The canonical nucleosome is composed of DNA wrapped around a histone octamer containing two each of Histone H3, Histone H4, Histone H2A, and Histone H2B; as DNA exits the nucleosome, it is stabilized by Histone H1 (reviewed in Cutter and Hayes, 2015). Post-translational modification of histone tails can modify transcriptional activity of local chromatin, composing a “Histone Code” (Strahl and Allis, 2000; Jenuwein and Allis, 2001). While hundreds of histone modifications exist, among them phosphorylation, ubiquitination, ribosylation, acetylation, and methylation, the latter two have been best characterized by researchers will be focused on here.

I.B.1.i *Histone modifications marking active transcription*

H3K4me3, deposited by the SET and MLL enzymes, is a canonical histone marker connoting active promoters. SET2-mediated deposition of H3K36Me2 and 3 marks active gene bodies (Bannister *et al*, 2005; Edmunds *et al*, 2008). DOT1L is the only known H3K79 methyltranferase (Steger *et al*, 2008); H3K79 methylation associates with active transcription and elongation of gene bodies (Steger *et al*, 2008, Wang *et al*, 2008; Kouskouti *et al*, 2005), though other data point to the requirement for DOT1L for the appropriate formation of heterochromatin (Jones *et al*, 2008).

An opposing affinity for DNA methylation and activating histone marks has been characterized: For example, Dnmt3L has increased binding affinity for unmethylated H3K4 residues (Ooi *et al,* 2007). Loss of the H3K4 demethylase KDM1 leads to loss of maternal imprinting in oocytes (Ciccone *et*

*al,* 2009); *KDM1*-null embryonic stem cells display a progressive loss of methylation due to reduced Dnmt1 stability (Wang *et al,* 2008).

I.B.1.ii *Histone modifications marking heterochromatin*

Heterochromatin is distinct from euchromatin in its degree of chromatin compaction, first evidenced by increased intensity of staining by DNA binding dyes (Arrighi and Hsu, 1971), and reduced transcriptional activity (Milot *et al*, 1996). Constitutive heterochromatin, which remain condensed throughout the cell cycle and development, is enriched for H3K9Me2 and 3 (Peters *et al*, 2002) deposited by the histone methyltransferases G9a, Suv39H1/2, SETDB1 (Tachibana *et al,* 2002; Rea *et al*, 2000; Schultz *et al*, 2002). Spreading of the repressive H3K9Me2 and 3 is mediated by Heterochromatin Protein 1 (HP1), a structural component of heterochromatin (James and Elgin, 1986; Bannister *et al*, 2001). HP1 recognizes H3K9Me2/3 and recruits SUV39H1, which deposits more H3K9Me2/3 in a positive feedback loop (Lachner *et al*, 2001). H4K20Me3, deposited by Suv4-20H, is also associated with constitutive heterochromatin and is spread by HP1 binding (Schotta *et al*, 2004; Kourmouli *et al*, 2004; Schotta *et al*, 2008).

Several links between repressive histone marks and DNA methylation are characterized. UHRF1, the cofactor of DNMT1, recognizes and binds H3K9Me3 (Liu *et al*, 2013), acting as a link between H3K9Me3 and maintenance of DNA methylation. In addition, DNMT3a associates with G9a via MPP8 (Chang *et al*, 2011).

H3K27Me3 is typical marker of constitutive heterochromatin;

established by the EZH1 or EZH2 component of the PRC2 complex (Kuzmichev *et al*, 2002). PRC2 is also thought to exclude DNMT activity; H3K27Me distribution is impacted by local DNA methylation in *Arabidopsis*, mouse embryonic stem cells (Hagarman *et al*, 2013), embryonic fibroblasts (Reddington *et al*, 2013; Lindroth *et al*, 2008), and neural stem cells (Wu *et al*, 2010).

I.B.1.iii *Histone modifications marking enhancers*

Two main histone modifications mark enhancers. Mono or di-methylation of H3K4 (H3K4me1/2), deposited by the histone methyltransferases MLL3 and 4, is a general marker for enhancers (Heintzman *et al*, 2009; Visel *et al*, 2009), whereas H3K27Ac is thought to specifically mark active enhancers (Creyghton *et al*, 2010).

I.B.1.iv Histone mark cooperation, antagonism, and bivalency

It is also important to note that while histone modifications are discussed as separate entities (and indeed they are separate modifications), they often co-occur and cooperate functionally. For example, recent data also demonstrate a cooperative relationship between PRC2, H3K9Me2/3 and H3K27Me3 to maintain HP1 at sites of heterochromatin (Boros *et al*, 2014). Conversely, H3K4Me3 impairs H3K9Me, which helps to maintain a euchromatic state (Binda *et al*, 2010).

Genes involved in differentiation and lineage control are often bivalently marked for H3K27Me3 and H3K4Me in embryonic stem cells (Azuara *et al*, 2006; Bernstein *et al*, 2006) and zebrafish blastomeres (Vastenhouw *et al*,

2010). Bivalently marked chromatin domains can further be separated by occupancy by PRC2 with or without co-occupancy by PRC1; PRC1-co-occupied domains are more enriched for promoters of developmental regulating genes (Ku *et al*, 2008). Other data support histone residue monomethylation as a mark for maintaining some degree of activation potential—for example, the enhancers of genes involved in differentiation are marked with the monomethylation of H3K4, H3K9 and H3K27. Other genes marked with H3K27Me3 prior to differentiation rapidly gain H3K4Me1, H3K49Me1, and K4K20Me1 upon differentiation (Cui *et al,* 2008).

I.B.1.v Removal of histone marks

Enzymes that catalyze the removal of histone modifications, notably the histone deacetylases (HDACs) and histone demethylases, have been extensively reviewed by others (Upadhyay and Cheng, 2010; Klose and Zhang, 2007; Haberland *et al*, 2009) and is not the focus of this review. Of note, however, is the interaction of DNA methylation with histone mark erasers, most commonly erasers of activating histone marks. Methylated DNA and bound MeCP2 recruits histone deacetylase 1 or HDAC1 (Nan *et al*, 1998; Jones *et al*, 1998) as does Dnmt1 itself (Fuks *et al*, 2001).

I.B.2 Chromatin remodeling complexes

ATP-dependent chromatin remodeling complexes (CRCs) utilize energy to disrupt or reposition nucleosomes, thereby affecting numerous cellular processes including transcription, chromatin structure, and DNA repair (reviewed in Smith and Peterson, 2004 and Clapier *et al*, 2017). While CRCs

vary widely in size and composition, all contain an ATPase subunit, by which they are divided into the subfamilies SWI/SNF, ISWI, Mi2, and INO80; general biological functions of these subfamilies also appear to differ. The SWI/SNF subfamily are thought to function as positive regulators of transcription via nucleosome repositioning, ejection, or histone dimer eviction; further, SWI/SNF binding at promoters seems to directly influence DNA methylation loss (Banine *et al*, 2005). In comparison, the Mi-2 subfamily of CRCs, which includes the NuRD complex, typically act as transcriptional repressors. Most Mi-2 complexes contain a histone deacetylase (Kehle *et al*, 1998; Xue *et al*, 1998; Zhang *et al*, 1998, Reynolds *et al*, 2012) and some, as previously stated, associate with methyl binding proteins (Wade *et al*, 1999). The ISWI subfamily of CRCs primarily mediates nucleosome assembly and spacing (Clapier *et al*, 2001; Langst *et al*, 1999). Finally, the INO80 family is thought to function primarily in nucleosome editing by which it can affect transcription as well as DNA repair and replication (Papamichos-Chronakis *et al*, 2010).

### I.B.3 Spatial organization of the genome: Chromatin domains

#### I.B.3.i *Topologically Associated Domains (TADs)*

Interphase chromatin has long been known to occupy the nucleus in a nonrandom pattern (Stack *et al*, 1977; Kurz *et al*, 1996; Croft *et al*, 1999; Sun *et al*, 2000; Bolzer *et al*, 2005), with a postulated effect on coordinated gene expression (Spilianakis *et al*, 2005). Recent advances have made possible the evaluation of genome-wide long range chromatin interactions and transcriptional synchronization of topologically associated domains, or TADs

(Lieberman-Aiden *et al*, 2009; Dixon *et al*, 2012; Rao *et al*, 2014). TADs form distinct units that are functionally synced, sharing similar transcriptional activity (Hou *et al*, 2012; Sexton *et al*, 2012) and replication timing (Ryba *et al*, 2010; Pope *et al,* 2015). TAD interactions change upon differentiation of embryonic stem cells, allowing for widespread changes in transcriptional programming (Dixon *et al*, 2015); changing TAD interactions have also been implicated in driving enhancer switching at the *Hox* loci during mouse development (Andrey *et al*, 2013, Noordermeer *et al*, 2014) and have been used to predict genomic regions associated with disease states (Javierre *et al*, 2016). Targeted disruption of TADs via CRISPR/Cas9 mediated inversion of TAD sequence, or disruption of TAD partitioning elements such as CTCF binding sites (described below) lead to inappropriate enhancer-promoter interactions and clear effects on phenotype (Lupianez *et al*, 2015).

I.B.3.ii TAD partitioning: CTCF and cohesin

How TADs are partitioned, maintained, and reorganized has been an area of active study. Inter-TAD DNA is enriched for ubiquitous “house-keeping” genes as well as the proteins CTCF and cohesin (Dixon *et al*, 2012).

CCCTC-binding factor, or CTCF, was originally identified in heterologous reporter assays as a transcription factor that could both activate and repress gene expression (Baniahmad *et al*, 1990; Lobanenkov *et al*, 1990). CTCF occupancy of DNA is methylation-sensitive (Rodriguez *et al*, 2010; Lai *et al*, 2010; Chang *et al*, 2010; Wang *et al*, 2012); CTCF also helps to maintain its binding sites in an unmethylated state by forming a complex

with DNMT1 and poly(ADP-ribose) polymerase 1, which then inactivates DNMT1 (Zampieri *et al*, 2012). In addition to its role as an insulator protein that can disrupt enhancer-promoter interactions (Bell *et al*, 1999; Varma *et al*, 2015), CTCF also acts as a mediator of chromatin looping: CRISPR/Cas9-mediated inversion of CTCF sites reconfigures chromatin looping and impacts transcription patterns (Guo *et al*, 2015; de Wit *et al*, 2015). CTCF also acts a boundary element delineating TADs. Deletion of CTCF sites within the X inactivation center leads to increased interactions between separate TADs (Nora *et al*, 2012). Finally, CTCF can contribute to chromatin anchoring, as with the inactive X in concert with the lncRNA *Firre* (Yang *et al*, 2015).

CTCF is thought to act in concert with the multiprotein complex cohesin, a ring-shaped multiprotein complex originally described as a protein that prevents premature separation of sister chromatids during mitosis and meiosis (Michaelis *et al*, 1997). Depletion of both CTCF and cohesin leads to loss of TAD looping (Nora *et al,* 2017 and Schwarzer *et al,* 2016 respectively) though different effects on domain structure and interactions occur with loss of either protein suggesting differential activities (Zuin *et al*, 2013). Many lines of evidence support a cooperative relationship between CTCF and cohesin. ChIP-Chip experiments demonstrated a correlation of cohesin components to CTCF binding sites (Rubio *et al*, 2008), suggesting chromatin co-occupancy; these results were confirmed genome-wide via ChIP-Seq for both proteins, which show a significant overlap in peaks (Hansen *et al*, 2017). Co-immunoprecipitation of CTCF and cohesin subunits Rad21, Smc1, and Smc3

revealed a physical interaction between the two (Hansen *et al*, 2017). A series of cell knockouts for CTCF and cohesin's known interactor Wapl (Tedeschi *et al*, 2013) indicate that cohesin position on chromatin is dependent on CTCF and Wapl as well as local transcription (Busslinger *et al*, 2017).

## I.C Dynamics of epigenetic reprogramming during embryogenesis

The embryonic epigenome changes dramatically between fertilization and implantation, as will be described in this Section and depicted in **Fig I.1**. Especial attention is paid to the embryonic methylome; however, histone modifications and other changes are also discussed.

Upon fertilization at embryonic day 0 (e0), the male and female pronuclei undergo massive epigenetic reprogramming that is asynchronous between the male and female pronuclei. The male pronucleus loses methylation within six to eight hours post-fertilization. Loss of methylation in the female pronucleus occurs over a number of cell divisions, reaching completion by e3.5 (Oswald *et al*, 2000; Mayer *et al*, 2000). Previously thought to be exclusively the result of active and passive demethylation mechanisms respectively, recent work has revealed that the male and female pronuclei both undergo some degree of active and passive demethylation (Wang *et al* 2014; Guo *et al*, 2014). Stella prevents TET3-mediated demethylation of the maternal genome and paternally imprinted loci through binding of H3K9Me2—as such, these remain methylated in a parent-of-origin specific manner in the soma throughout development, though they will later be demethylated in the embryonic germline (discussed in the next paragraph). Certain classes of

transposable elements, namely intracisternal A particles (IAPs), a highly active class of retrotransposon (Dewannieux *et al*, 2004), also resist demethylation in this initial wave of demethylation.

The active histone mark H3K4Me3 is depleted in the early zygote, accruing rapidly by the two-cell stage (e1.5), accompanying the major wave of zygotic genome activation (ZGA). Interestingly, oocytes and the maternal pronucleus carry a noncanonical version of H3K4Me3 which is only replaced by the canonical version by the late two-cell stage (Zheng *et al*, 2016). In comparison, the repressive histone mark H3K27Me3 marks transcriptionally silent and hypomethylated regions of chromatin in oocytes; upon fertilization, H3K27Me3 is globally lost from the paternal pronucleus and developmental genes in the maternal pronucleus, but persists at other regions, leading to an asymmetric distribution of H3K27Me3 that is corrected by the late two-cell stage, leading to the classic bivalent histone mark of H3K4Me, H3K27Me3 of developmental genes in the early embryo (Liu *et al*, 2016).

Chromatin accessibility also mirrors transcriptional activation of the zygotic genome. At the time of ZGA, at roughly e3, the chromatin of the maternal and paternal pronuclei, though asymmetrically methylated, are substantially more permissive as measured by ATAC-Seq (Wu *et al*, 2016), presumably due to the requirement of pluripotency. As such, cells of the eight- and sixteen-cell morula, and the early 32-cell blastocyst, inherit a hypomethylated, genetically accessible genome; early pluripotent genomes are also marked with H3KMe (Torres-Padilla *et al*, 2007). Throughout early

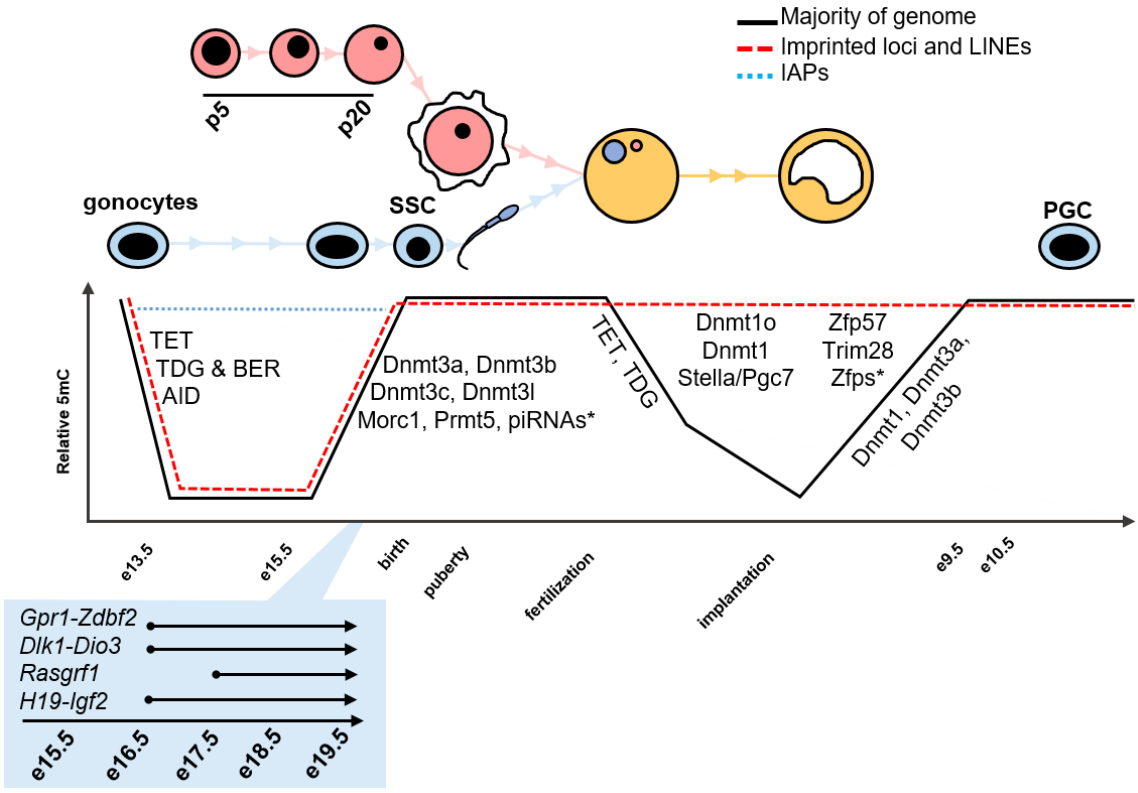
cleavage stages, imprinted loci are bound by ZFP57 and Trim28 to resist Tet3-mediated demethylation; methylation is maintained across cell divisions by Dnmt1.

By the late blastocyst stage (e3.5), the maternal pronucleus is hypomethylated (Smallwood *et al*, 2011) and *de novo* methylation is initiated. From the blastocyst to the epiblast stage, DNA methylation is targeted to the promoters of lineage specific genes (Borgel *et al*, 2010) and is accompanied by dynamic changes in histone mark distribution in lineage-specific patterns (Rugg-Gunn *et al*, 2010). By implantation at mouse embryonic day 4.5, the nascent embryonic genome is fully remethylated.

Specification of primordial germ cells (PGCs) occurs at roughly day e6.5 in the proximal epiblast, where they can easily be identified by robust expression of alkaline phosphatase (MacGregor *et al*, 1995). At e7.5, PGCs migrate briefly to the endoderm; by e8.5 they migrate towards the genital ridge by way of the hindgut endoderm and mesentery. Once arrived, roughly at e10.5, PGCs expand rapidly, and undergo a global wave of demethylation of virtually all methylation marks, including imprinted loci and some classes of transposable elements. Demethylation occurs via both passive (Ohno *et al*, 2013) and TET1, TET2, and TET3-mediated active (Peat *et al*, 2014; Hackett *et al*, 2013) processes; notably, certain classes of transposable elements, namely IAPs and ERV-K elements, as well as certain single copy genes, resist demethylation in the PGC as well (Seisenberger *et al*, 2012). Following complete demethylation by e13.5, *de novo* methylation of paternal imprints

occurs in the mitotically arrested gonocytes starting at e14.5, with some variation in the exact timing of establishment amongst paternal DMRs (Ueda *et al,* 2000, Davis *et al,* 2002, Hiura *et al,* 2010; **Fig I.1**). In females, reestablishment of maternal imprints occurs after birth in developing oocytes, between p5 and p20 (Li *et al,* 2004, Lucifero *et al,* 2004, Hiura *et al,* 2006).

**Fig I.1. Timing of and *trans*-acting factors for epigenetic reprogramming during embryogenesis and germ cell development.** Adapted from Tomizawa and Sasaki 2012. Timing and involvement of factors described in the previous section from fertilization (e0) to birth. **Outset**: Establishment of paternally imprinted loci begins asynchronously at different imprinted loci; methylation is thought to be complete by birth. *denote factors specifically involved in transposable element methylation. Maternal imprints are established in growing oocytes after birth. SSC: Spermatogonial stem cell; PGC: primordial germ cell; p: postnatal day; e: embryonic day.



## I.D Models for the study of DNA methylation mechanisms: Imprinted loci and transposable elements

Two classes of genes have proven especially utile in elucidating the mechanisms by which DNA is targeted to specific genomic regions. As previously stated, while the majority of the genome undergoes epigenetic reprogramming immediately post-fertilization, some classes of genes resist this first wave of demethylation and instead remain methylated in germline patterns throughout early embryogenesis. Imprinted loci, therefore, retain methylation in a parent-of-origin specific manner. These are subjected to demethylation and remethylation in the embryonic germline, allowing for sex-specific reprogramming.

### I.D.1 Imprinted loci

Diploid organisms harbor two complete copies of their genome: One inherited from the mother, and one inherited from the father. While the vast majority of genes express biallelically, from both the maternal and paternal alleles, a small subset of genes express monoallelically in a parent-of-origin specific manner. These are referred to as imprinted genes. First coined to describe parent-of-origin-dependent differences in chromosome behavior during embryogenesis and meiosis of *Sciara* species (Crouse 1960), the existence of imprinted genes was first hypothesized as the reason for the inviability of embryos derived from two male or two female pronuclei (McGrath and Solter 1984; Surani *et al,* 1984). The case for imprinted genes was

furthered by a series of translocation experiments which pinpointed certain chromosomal regions that differed based on maternal or paternal origin (Cattanach 1986). In this century, viable bimaternal embryos have been generated by using nuclei that each carried a mutation at two different imprinted loci (Kono *et al,* 2004; Kawahara *et al,* 2007).

To some extent, the results of early transplantation experiments would also predict the failure of embryos derived from somatic cell nuclear transplantation. To date, somatic cell nuclear transfer (SCNT) has also been successfully performed for a number of domestic species and is performed somewhat regularly in some production animal species (Campbell *et al*, 1996; Lee *et al*, 2005; Baguisi *et al*, 1999; Kato *et al*, 1998; Galli *et al*, 2003; Woods *et al*, 2003; Shin *et al*, 2002). However, the low success rate and common developmental and placental abnormalities associated with SCNT have also been attributed in part to incomplete epigenetic reprogramming, including at imprinted loci (Blelloch *et al*, 2004; Beaujean *et al*, 2004; Zhang *et al*, 2004; Chen *et al*, 2014).

Though many hypotheses for the evolution of imprinted loci exist (reviewed in Wilkins and Haig 2003; Spencer and Clark 2014), perhaps the best known is the kinship theory or “conflict hypothesis,” which assumes an intrinsic evolutionary conflict between the sexes. Simply put, in species where postzygotic care of the fetus falls predominantly on the mother, a conflict arises between use of maternal resources and the growth of the fetus (which serves to perpetuate of both maternal and paternal genetic material to the next

generation). While the mother optimizes her survival and perpetuation of genetic material through conservation of resources for herself and all of her offspring, the greatest benefit to the father is to optimize the growth and survival of his offspring. This is especially plausible in species where multiple paternity litters are common, or in species where females might be raising offspring from a previous mating while pregnant. Under these expectations, imprinted genes that promote fetal growth would be maternally silent, whereas imprinted genes that inhibit fetal growth would be paternally silent. Compellingly, the vast majority of imprinted genes play important roles for body growth and metabolism, with maternally imprinted genes restricting body growth, and paternally imprinted genes encouraging growth (Smith *et al,* 2006). Further, all known imprinting disorders have dramatic impacts on body growth and metabolism (**Table I.2**). Finally, in animal hybrids, directionality of cross greatly affects offspring phenotype. Hybrid offspring are often much larger or smaller than either parental species with other distinct differences in appearances and behavior (Troyer 2006, McKinnell and Wessel 2012).

#### I.D.1.i.1 Features of imprinted loci

While the hallmark feature of all imprinted genes is monoallelic, parent-of-origin-specific gene expression, it should be noted that monoallelic expression is often tissue- and developmental stage-specific, achieved by tissue-specific DMRs (Song *et al*, 2004) and tissue-specific histone marks and chromatin structure (reviewed extensively in Prickett and Oakey 2012). However, generally speaking, methylated DMRs are transcriptionally

repressed and carry repressive histone marks such as H3K9me3, H3K27Me2, and H4K20me2, whereas their unmethylated homologues are transcriptionally active and are marked by H3K4me3 and H3/H4 acetylation (Dindot *et al,* 2009; Kacem and Feil 2009). One notable exception to this involves the *Rasgrf1* locus, where the paternal allele is both expressed and methylated. At *Rasgrf1*, DNA methylation antagonizes the deposition of H3K27Me3 on the paternal allele; H3K27Me3 accrues instead on the maternal allele (Lindroth *et al,* 2008).

Imprinted genes are generally located in 1 Mb-long clusters throughout the genome. While all genes in a cluster are typically imprinted, not all are imprinted from the same parental allele.

**Table I.2. Disorders of imprinting in human and mouse**. Affected genomic locus, perturbed gene, relevant phenotypes, and primary literature are included. ✋ and 🐁 precede human and mouse chromosomal locations.

| Disorder Name | Imprinted locus | Gene perturbation | Relevant phenotypes | Citations |
|---|---|---|---|---|
| Prader-Willi Syndrome | ✋15q11-13<br>🐁 7B5 | *Snrpn,* paternal repression | Neonatal hypotonia, gonadal hypoplasia, dysphagia; Adult hyperphagia and obesity | Buiting *et al,* 1997; Yang *et al,* 1998 |
| Angelman Syndrome | ✋15q11-13<br>🐁 7C | *Ube3a,* maternal repression | Mental retardation, aphasia, ataxia | Meng *et al,* 2014 |
| Beckwith-Wiedemann Syndrome | ✋11p15<br>🐁7F5 | IC1, maternal repression of *Cdkn1c*; IC2, maternal derepression of *Igf2*) | Macrosomia, macroglossia, embryonal-origin cancers, and other abnormalities | Algar *et al,* 2000; Andrews *et al,* 2007; Tunster *et al,* 2011 |
| Silver-Russell Syndrome | ✋11p15<br>🐁7F5 | Hypomethylation of ICR1, repression of *Igf2* | Pre- and postnatal growth retardation | Monk *et al,* 2002; Gicquel *et al,* 2005 |
| Temple Syndrome | ✋14q32<br>🐁12F1 | Repression of *Dlk1* and *Rtl1* | Low birthweight, short stature, hypotonia, early puberty | Kagami *et al,* 2008 Ioannides *et al*, 2014 |
| Paternal UPD14 Syndrome | ✋ 14q32<br>🐁12F1 | Depression of *Rtl1* | Placentomegaly, mental retardation, skeletal abnormalities | Kurosawa *et al*, 2004 |
| Pseudohyperparathyroidism 1b | ✋20q<br>🐁2H4 | *Gnas1* loss of function | Hyperparathyroid hormone resistance (maternal only); calcium phosphate imbalances | Juppner *et al,* 1998; Bastepe 2008; Lemos and Thakker 2015 |
| Transient neonatal diabetes mellitus | ✋6q24<br>🐁10A2 | Derepression of *Hyma1* and *Zac* | Growth retardation, increased chance of Type II diabetes in adulthood | Gardner *et al,* 2000; Arima *et al,* 2001; Mackay *et al,* 2002 |

#### I.D.1.i.2 Regulation of imprinted loci

Imprinted expression is dictated by DNA methylation at differentially methylated regions (Li *et al*, 1993). Differentially methylated regions (DMRs) comprise CpG-rich stretches of DNA, in the case of imprinted genes, vary in methylation between the maternal and paternal allele. In quick succession, DMRs were identified at two well-known imprinted genes, *Igf2r* (Stoger *et al*, 1993; Wutz *et al*, 1997) and *H19* (Ferguson-Smith *et al*, 1993). Others have subsequently been identified at nearly every gene (reviewed in Bartolomei and Ferguson-Smith 2011). Indeed, DMRs have been used for the discovery of novel imprinted genes in mouse and human tissues (Smith *et al,* 2003; Stelzer *et al,* 2013; Yuen *et al,* 2011). Another common approach has been the use of RNASeq for transcriptome profiling of polymorphic strains to identify monoallelically expressed genes (Wang *et al*, 2008; Wang *et al*, 2011). Other methods of imprinting regulation are non-mutually exclusive to differential methylation; first, binding of the insulating factor CTCF; and second, lncRNA-mediated mechanisms. Described in Section IB, CTCF has been demonstrated to regulated expression at the *H19*/*Igf2* locus, binding the maternal Imprinting Control Region (ICR) of *H19*/*Igf2* and preventing Igf2 from accessing upstream enhancers (Bell *et al*, 2000). At *Rasgrf1,* binding of CTCF to the maternal, unmethylated DMR prevents the ability of an upstream enhancer to promote *Rasgrf1* expression; methylation of the paternal DMR precludes CTCF binding, permitting *Rasgrf1* expression from the paternal allele (Yoon *et al*, 2005).

In addition, lncRNAs have been characterized at nearly every imprinted locus and are expressed antisense to imprinted genes. lncRNA expression is thought to augment gene expression in *cis* and sometimes in *trans* through a number of mechanisms, which are discussed in Chapter IV.

I.D.1.i.3 Directing DNA methylation to DMRs: The discovery of *cis*-elements

Much effort has been put into defining sequences and states that could direct DNA methylation to DMRs. Targeted and genome-wide computational approaches have delineated some sequence features of DMRs. For example, transposable element composition of flanking regions, notably long repetitive elements, has been proposed to be a hallmark feature of imprinted loci (Walter *et al,* 2006). Indeed, the flanking regions of imprinted loci are relatively devoid of SINEs and CpG islands (Greally 2012, Ke *et al,* 2002) but are enriched for LINEs (Allen *et al,* 2003, Luedi *et al,* 2006).

Others have sought to identify specific *cis*-elements that might “flag” DMRs for methylation, with varying success. Nevertheless, most of the *cis*-acting elements that have been identified have utilized imprinted loci. For example, a series of transgenic constructs at the *Snrpn* locus identified 5 *cis*-elements required for DNA methylation (Kantor *et al,* 2004). At the *Gnas* locus, truncations or microdeletions of the distal portion of the *Nesp* gene led to loss of methylation at the *Gnas* ICR (Chotalia *et al,* 2008; Frohlich *et al*, 2010). A negative-effect *cis*-element was discovered at *H19* that requires CTCF binding for maternal repression of the *Igf2* (Thorvaldsen *et al*, 1998; Schoenherr *et al*, 2003; Pant *et al,* 2003, Pant *et al,* 2004; Fedoriw *et al*, 2004). Finally, a series

of GC-rich tandem repeats are required for methylation at the *Rasgrf1* DMR (Yoon *et al,* 2001).

### I.D.1.ii Resisting demethylation: Imprinted loci

Methylation patterns at imprinted loci are retained through the first wave of demethylation in the preimplantation embryo. As such, imprinted loci remain differentially methylated in a parent-of-origin specific manner in the soma.

The proteins Stella, Zfp57 and its cofactor Trim28 protect imprinted loci from demethylation. Stella or PGC7 is a maternal effect germ-cell specific protein essential for early development (Sato *et al,* 2002; Payer *et al,* 2003). Stella protects maternal and paternal imprints from the conversion of 5mC to 5hmC (Nakamura *et al,* 2007) through the binding of heterochromatic histone mark H3K9Me2 (Nakamura *et al,* 2012) in the maternal and paternal pronuclei.

Similarly, the maternal effect protein Zfp57 is necessary for the maintenance of both maternal and paternal imprints. A KRAB-domain zinc finger protein (ZFP), Zfp57 directly recognizes methylated CpGs (Quenneville *et al,* 2011). *Zfp57* null animals are embryonic lethal, while loss of zygotic *Zfp57* only causes incompletely penetrant neonatal lethality (Li *et al,* 2008).

KRAB-domain ZFPs are thought to effect chromatin changes largely through their cofactor Trim28 (also known as KAP1), first identified as a KRAB-ZFP interacting partner by several independent groups, as a known interactor recruitor of histone modifiers and nucleosome remodeling complexes (reviewed in Iyengar and Farnham 2011). Later, Trim28 was also identified as the affected gene in two nearly simultaneous ENU mutagenesis

screens. First, a screen for recessive mutations that arrest development in mouse mid-gestation characterized the *chatwo* mutant as a Trim28 hypomorph; *chatwo* mutants arrest by e9 with severe developmental defects (Shibata *et al,* 2011). Second, a screen for dominant mutations that modify epigenetic state identified the *Trim28*$^{Momme9}$ allele (Ashe *et al,* 2008). *Trim28*$^{MommeD9}$ heterozygotes are viable and fertile, but display incompletely penetrant metabolic and behavioral perturbations (Whitelaw *et al,* 2010).

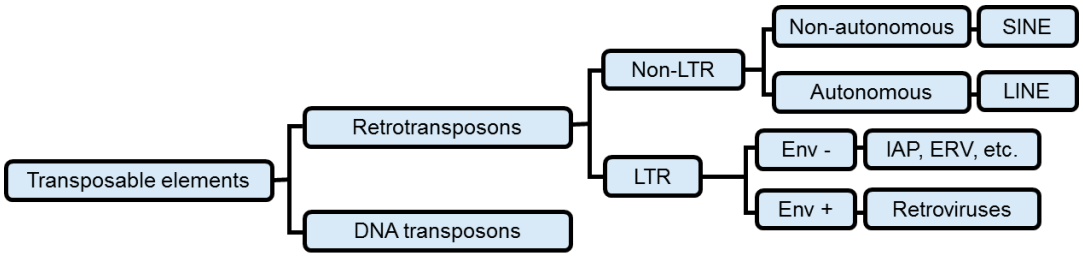
The role of Trim28 at germline imprinted loci was later characterized with the discovery that *Trim28* null embryos showed evidence of epigenetic dysregulation at imprinted loci (Messerschmidt *et al,* 2012); moreover, it has been shown to play a role in the establishment and regulation of secondary or somatic imprints, which arise from germline imprints and spread in the soma (Alexander *et al,* 2015). Methylation analysis of Trim28 has also recently been implicated in maintaining DNA methylation at transposable element sequence, namely ERVs, in neural progenitor cells (Brattas *et al,* 2017).

### I.D.2 Transposable elements

Transposable elements (TEs) represent another class of genes targeted for DNA methylation; some classes also resist demethylation, as will be discussed in later paragraphs. First described as a driver of hypervariable phenotype in maize (McClintock 1950), in eukaryotes, TEs occupy over 50% of the genomic space and can be divided broadly into two categories; Class I, retrotransposons; and Class II, DNA transposons; based on the origin of their integration in the genome. Class I TEs, or retrotransposons, are originally

RNA-based and require reverse transcriptase to first copy the genome into DNA, whereas Class II TEs, or DNA transposons, serve as their own templates for transposition. Class II transposable elements can be further categorized based on their structural features and ability to insert themselves elsewhere in the genome (autonomy) (reviewed in Chenais *et al,* 2012, **Fig I.2**).

**Fig I.2**. **Lineages of transposable elements.** Adapted from Koito and Ikeda 2013. Transposable elements can be divided first by their mechanism of transposition. DNA transposons provide their own template for transposition; retrotransposons require an RNA intermediate and reverse transcriptase. Retrotransposons are further divided by the presence or absence of flanking long terminal repeats (LTRs); non-LTR elements can be divided into autonomous and non-autonomous. Non-autonomous elements require activity of an autonomous element for transposition; autonomous elements encode necessary activities themselves. LTR elements are divided on presence or absence of envelope protein (env) coding sequence. For the purposes of this thesis, it is important to note that LINE and ERV methylation and suppression, as will be discussed in Section I.A.3, appear to be controlled by Dnmt3c (Barau *et al*, 2016).




In all host species, TEs are potent drivers of evolutionary change. TE amplification is associated with increases in overall genomic size (reviewed in Canapa *et al,* 2015) through the expansion of gene families and novel gene formation (DeBarry *et al,* 2006). TE-induced disruption of genes or gene processing has contributed to phenotype in many species (Morgan *et al*, 1999; Clark *et al*, 2006; Marchant *et al*, 2017). Coding sequence aside, derepression

of retrotransposons often activates nearby genes, and in the early embryo and developing oocyte, transient TE derepression may even be required for appropriate development (Peaston *et al*, 2004; Macfarlan *et al*, 2012). TEs have also contributed to the birth and expansion of transcriptional regulatory sequences (Thornburg *et al,* 2006; Sundaram *et al,* 2014). In addition, over 30% of annotated lncRNAs harbor TE sequence (Kapusta and Feschotte 2014), which in some cases could contribute to their ability to act in *trans* as with the lncRNA *ANRIL* and Alu, a family of human LTR-type transposable element (Holdt *et al,* 2013).

While their abilities to excise and insert themselves in novel parts of the genome have contributed greatly to the richness of the host genome, this virtue doubles as a constant threat to genomic integrity. TEs have forced an "evolutionary arms race," where there exists a coevolution of TEs and the silencing mechanisms by which their deleterious effects can be minimized (reviewed in Slotkin and Martienssen 2007). The resulting mechanisms typically involve DNA methylation: TEs are, similar to imprinted loci, protected during the genome-wide demethylation of the early embryo. Pathways and factors required for TE methylation are divergent from but often overlap with those required for imprinted loci, as will be discussed below. A further layer of regulation involves the transposable element and its silencing factors working together to regulate local gene expression (Ecco *et al,* 2016; Sundarem *et al,* 2017).

Different classes of transposable elements display some diversity in

methylation and remethylation patterns. LINE1 elements are subject to extensive demethylation and remethylation at roughly the same period as imprinted loci, whereas IAP elements resist demethylation in the preimplantation embryo and primordial germ cell development (Lane *et al*, 2003). Why some classes of TEs resist demethylation and others do not is unknown—however, a plausible reason could involve varying local chromatin accessibility, as transposable element classes are also differentially marked with repressive histone modifications (Walter *et al*, 2016), or potentially differential enrichment of consensus sequences for proteins that recruit the TET proteins to methylated DNA (Okashita *et al*, 2014; Hassan *et al*, 2017). In the event of TE demethylation, a pathway for *de novo* methylation of TEs has been characterized in plants and animals, the piRNA pathway, is highly relevant to our model system and will be discussed in the following section.

### I.D.2.i The piRNA pathway

PIWI-interacting small RNAs, or piRNAs, are a germline-specific class of small noncoding RNAs whose function is to target and silence transposable elements via *de novo* methylation. Their name derives from their associated proteins—the P-element induced wimpy testes or “Piwi” proteins first characterized in *Drosophila*. True to its name, knockout of *piwi* leads to hypogonadism and sterility in male flies (Lin and Spradling 1997). Piwi homologues have since been characterized in *C. elegans*, *Arabidopsis*, mouse, and human (Cox *et al,* 1998; Kuramochi-Miyagawa *et al,* 2001; Deng and Lin 2002; Qiao *et al,* 2002) as well as many other species.

The three murine piwi homologues, named Miwi, Mili, and Miwi2, differ based on their expression patterns in the male germline as well as their binding affinities for 1) lengths of piRNAs and 2) protein partners. Miwi2 is expressed transiently between e15.5 and p3 and binds 28nt piRNAs (Aravin *et al*, 2006); whereas Miwi is restricted to the post-meiotic spermatid and binds 30nt long piRNAs, and is most pertinent to the activity of pachytene piRNAs; Mili is expressed as early as e12.5 and persists through sexual maturity, and binds 26nt piRNAs (Aravin *et al*, 2008). Null animals for all three family members are viable but sterile (Carmell *et al,* 2007, Reuter *et al,* 2009, Kuramochi-Miyagawa *et al*, 2004); further, knockout of MILI and MIWI2 is associated with severe hypomethylation of LINE1 and IAP elements, supporting an essential role for MILI and MIWI2 in establishing *de novo* methylation at some classes of retrotransposons (Kuramochi-Miyagawa *et al*, 2008). More recently, the roles of MIWI2 and MILI have been further differentiated, as MILI knockout has more widespread effects on the methylation of LINE1 elements (Manakov *et al*, 2015).

Mammalian piRNAs were discovered almost simultaneously by a number of groups in 2006. First, by total RNA gel separation, which revealed a population of small RNAs larger than miRNAs at 26-30 nt (Girard *et al,* 2006); second, by RNA sequencing of Piwi-bound RNAs (Grivna *et al,* 2006); and third, by RNA sequencing a small RNA fraction arising from a screen for candidate transcriptional gene silencing complexes (Lau *et al,* 2006).

Two classes of piRNAs are characterized, pre-pachytene and

pachytene, based on the timing of their expression in the male germline. This thesis focuses on the former; though pachytene piRNAs remain an area of active research based on their key role in maintaining genomic integrity and mRNA degradation in the terminally differentiating male germ cell (Zheng and Wang 2012; Gou *et al,* 2014; Goh *et al*, 2015).

Pre-pachytene piRNAs are located in 20-90kb intergenic clusters throughout the genome and are expressed in the male germline from e12.5 through meiosis post-natally (Aravin *et al*, 2007). piRNA precursor molecules are transcribed as single long transcripts and are trafficked out of the nucleus for primary piRNA processing. There, they are cleaved by the mitochondrial protein MitoPLD (*Drosophila* homologue Zucchini) (Pane *et al*, 2007; Olivieri *et al*, 2010; Ipsaro *et al*, 2012; Nishimasu *et al*, 2012), a process that is mediated by the RNA helicase MOV10L1 (Vourekas *et al*, 2015) and GASZ, which targets piRNAs to the mitochondria (Ma *et al*, 2009; Zhang *et al*, 2016). Fragmented piRNA precursors are then loaded onto Mili or Miwi2. They are further modified by Henmt1 to carry 3’ 2’-O-methylation (Kirino *et al,* 2007; Simon *et al,* 2011; Lim *et al,* 2015); this is tightly coupled with trimming of the 3’ end of primary piRNA intermediates (Kawaoka *et al*, 2011). Trimming is performed by an unknown exonuclease in mouse, though in *Bombyx*, PNLDC1 has been identified as the responsible exonuclease (Izumi *et al*, 2016); in *C. elegans,* the RNase PARN-1 has similarly been identified (Tang *et al*, 2016). The Tudor protein TDRKH (other Tudor proteins will be discussed in later paragraphs) is known to be required for primary piRNA processing and is

thought to promote 3’ trimming (Saxe *et al*, 2013).

Processing of secondary piRNAs involves the “Ping-Pong” cycle, which leads to piRNA amplification. MILI-bound primary piRNAs guide MILI to transposable element transcripts, which slices the tenth nucleotide of the transcript complimentary to the primary piRNA (De Fazio *et al*, 2011). This process is dependent upon ATPase activity of Mouse Vasa Homologue or MVH (Tanaka *et al*, 2000; Kuramochi-Miyagawa *et al*, 2010; Wenda *et al*, 2017). Secondary piRNAs are loaded onto MIWI2, which, complexed with EXD1 and TDRD12, translocates to the nucleus and targets primary piRNA precursor transcripts (Yang *et al*, 2016). *Zucchini*, the *Drosophila* homologue of MitoPLD, is also thought to play a role in diversifying the piRNA pool via 3’ piRNA processing and promoting 3’-end-directed phased piRNA biogenesis (Han *et al,* 2015; Mohn *et al,* 2015).

Many members of the Tudor protein family are required for the piRNA pathway, among those already mentioned are TDRKH (also known as TDRD2), the RNA helicase TDRD9. Tudor proteins act as molecular adaptors, mediating protein-protein interactions and therefore represent a promising scaffold upon which a Piwi-piRNA complex can be built. The Tudor proteins associate with the Piwi proteins through methylated arginine residues, a modification catalyzed by PRMT5 (Kirino *et al*, 2009; Kirino *et al*, 2010; Gonsalvez *et al*, 2006; Anne *et al*, 2007) and often also harbor helicase, rna-binding, and zinc-finger domains (Handler *et al*, 2011), which almost certainly lends to their role as molecular adaptors. Piwi proteins display distinct

preferences for Tudor proteins—in mouse, Mili interacts exclusively with TDRD1 (Reuter *et al*, 2009; Vagin *et al*, 2009; Wang *et al*, 2009), whereas Miwi interacts with several Tudor family members and Miwi2 a smaller subset (Kojima *et al*, 2009; Shoji *et al*, 2009; Vagin *et al*, 2009; Wang *et al*, 2009; Vasileva *et al*, 2009). Loss of Tudor proteins TDRD1, TDRD6, TDRD9 and TDRD12 lead to sterility phenotypes though with diverse impacts on the piRNA pathway (Chuma *et al*, 2006; Vasileva *et al*, 2009; Wenda *et al*, 2017; Pandey *et al*, 2013). For example, loss of TDRD1 leads to reduced levels of MIWI2-bound piRNAs, reduced methylation of transposable elements, LINE1 element derepression and Miwi2 mislocalization. In comparison, loss of TDRD9 does not impact piRNA biogenesis but is required for transposon methylation (Wenda *et al*, 2017).

Ultimately, piRNAs lead to *de novo* methylation of transposable element sequence (in mammals). The exact mechanism by which piRNAs guide effector proteins to target loci remains unclear. In theory, piRNA-Piwi protein complexes to target their genomic targets for methylation. This has been demonstrated for the *Drosophila* Piwi (Brennecke *et al*, 2007; Darricarrère *et al*, 2013) but not for Mili; Miwi2 does display partial localization to the nucleus, though this is ablated in mice deficient for other components of the piRNA pathway including Mili and MVH-null mice (Aravin *et al,* 2009; Kuramochi-Miyagawa *et al*, 2010), suggesting that Miwi2 must be bound to piRNAs prior to translocating to the nucleus. Targeting Miwi2 to a regulatory region of a class of LINE-1 elements via zinc finger protein fusion on a Mili-null

background did result in *de novo* methylation of LINE-1 elements; the ZP-Miwi2 fusion also associated with Dnmt3a2 and Dnmt3l (Kojima-Kita *et al*, 2016).

How Miwi2 might target and associate with DNA *in vivo* remains in question. Certainly, an interplay between the Piwi protein complexes, perhaps by way of Tudor protein-mediated interactions, and other protein complexes or chromatin modifications is plausible. *Drosophila* Piwi associates directly with HP1 (Brower-Toland *et al*, 2007), though this association has not been demonstrated in animals; piRNAs in *Drosophila* are also known to induce chromatin changes (Le Thomas *et al*, 2014). Mice lacking the GHKL ATPase Morc1 show a loss of DNA methylation at specific transposon classes (Pastor *et al*, 2014); Morc1 known to mediate chromatin condensation in animals and plants (Moissiard *et al*, 2012; Harris *et al*, 2016) and in *C. elegans* acts downstream of RNA interference pathways to maintain repressive chromatin state (Weiser *et al*, 2017). Knockouts of the histone demethylase LSD1 lose methylation at transposable elements (Wang *et al*, 2009). As a specific demethylase targeting mono- or dimethylated H3K4 and H3K9, a potential pathway could involve accumulation of H3K4Me1 or 2, which reduces Dnmt binding (Ooi *et al*, 2007; Otani *et al*, 2009; Li *et al*, 2011; Guo *et al*, 2015). Certainly, piRNAs are known to recruit the repressive H3K9Me3 themselves, supported in mammals (Pezic *et al*, 2014) and extensively characterized in *Drosophila* (Klenov *et al*, 2011; Sienski *et al*, 2012; Le Thomas *et al*, 2013; Rozhkov *et al*, 2013). Deposition of H3K9Me3 could also recruit DNA

methyltransferases, as DNMT1's cofactor UHRF1 associates with H3K9Me3 (Rothbart *et al*, 2013) and helps to maintain repressed chromatin in a methylated state.

### I.D.2.ii Resisting demethylation: Transposable elements

Methylation at TEs must similarly be maintained and regulated throughout embryogenesis. In addition to the activity of Zfp57 at imprinted loci in the early embryo, other Zfps are thought to aid in the maintenance of methylated state at transposable element sequence. The list of interacting KRAB-Zfps is extensive and TE family and subfamily-specific; most are thought to maintain a repressive chromatin state via Trim28-mediated recruitment of histone modifying complexes and DNA methylation (Wolf and Goff, 2008; Rowe *et al*, 2013; Turelli *et al*, 2014; reviewed in Gifford *et al,* 2013). Recently, the role of ZFP-mediated TE regulation has been shown to extend beyond the embryo, potentially playing a role in the regulation of tissue-specific expression patterns (Ecco *et al,* 2016).

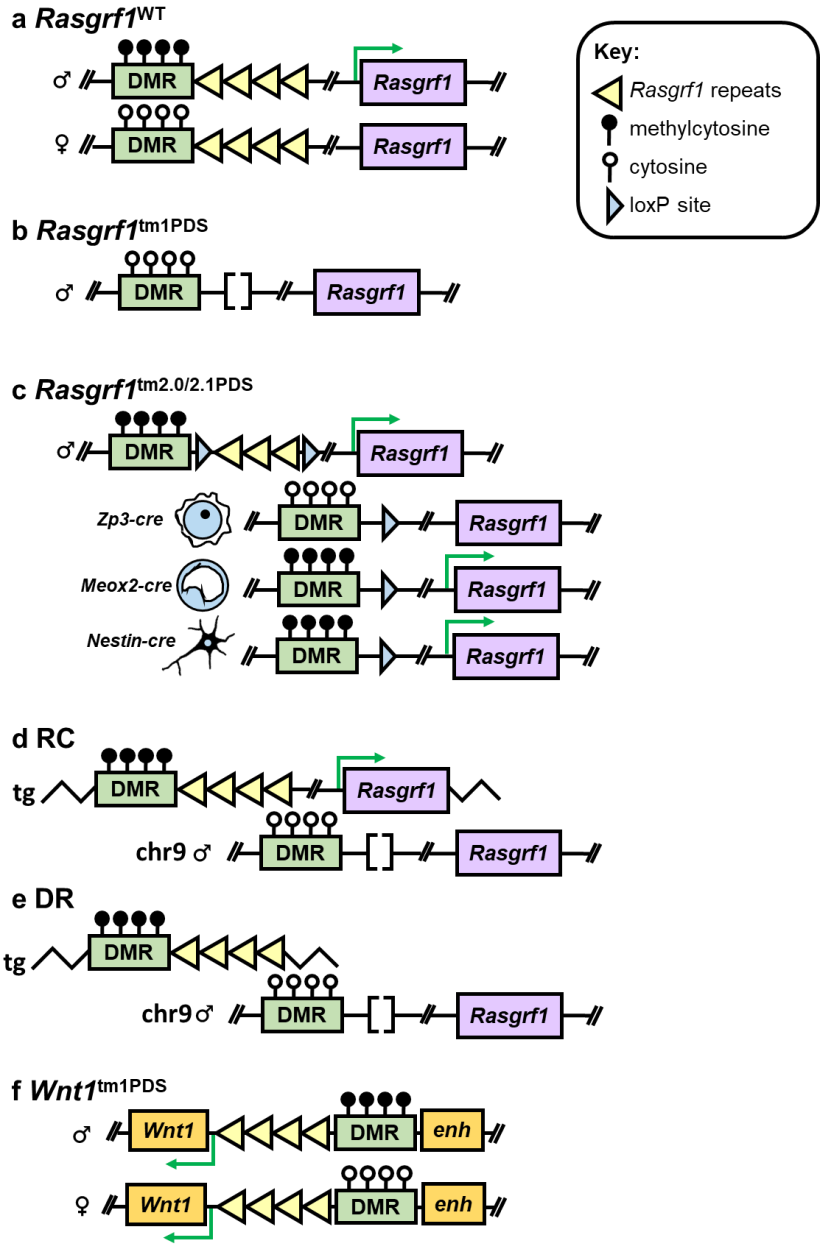
The arginine methyltransferase PRMT5 is also necessary for the maintaining a repressive histone state at transposable element sequence, namely IAP and LINE1 elements (Kim *et al*, 2014). PRMT5 catalyzes the repressive histone modifies arginine residues on histones as well as the Piwi proteins (Kirino *et al*, 2009) which, as a component of the piRNA pathway, act to methylate transposable element sequence *de novo*, as has been discussed previously.

## I.E *Rasgrf1* as a model system for the study of DNA methylation mechanisms.

*Ras-activating guanine nucleotide releasing factor* (*Rasgrf1*) is paternally imprinted in the neonatal mouse brain (Plass *et al,* 1996). The *Rasgrf1* ICR lies 30kb upstream of the *Rasgrf1* coding sequence. An adjacent series of tandem repeats (Pearsall *et al*, 1999) represent a *cis*-element required for establishment of DMR methylation in the male germline; deletion of the repeats leads to loss of methylation and reduced expression of *Rasgrf1* (Yoon *et al,* 2001). Animals deficient for the repeats recapitulate *Rasgrf1* knockout animals (Itier *et al,* 1998; Font de Mora *et al,* 2003; Clapcott *et al*, 2003; Giese *et al*, 2001), being significantly smaller than wild type littermates with deficits in olfactory learning and memory (Drake *et al*, 2009), but remain viable and fertile. Necessity for the repeats to maintain somatic DMR methylation is restricted to the preimplantation embryo (Holmes *et al,* 2005). As such, *Rasgrf1* has proven to be a useful system for the study of imprinted loci, yielding many insights into the mechanisms of imprinting (**Fig 1.3**).

**Fig I.3. Schematics of genetic manipulations at, or with components of, the *Rasgrf1* ICR. a)** The wild-type (WT) *Rasgrf1* ICR. Paternally, the *Rasgrf1* differentially methylated region (DMR) is paternally methylated; *Rasgrf1* is expressed from the paternal allele in neonatal brain. The maternal allele is unmethylated at the DMR and *Rasgrf1* is not expressed. **b)** *Rasgrf1*$^{tm1}$(Yoon *et al*, 2001) defines the *Rasgrf1* repeats as a *cis*-element that are required for the establishment of methylation at the paternal DMR. In the absence of DMR methylation, *Rasgrf1* is not expressed. **c)** *Rasgrf1*$^{tm2.0}$ (Holmes *et al*, 2005) permits *Cre*-mediated recombination of the *Rasgrf1* repeats. Deletion of the repeats at different points in development defines a requirement for the repeats to maintain somatic methylation of the DMR through the epiblast stage (Meox2-cre); recombination and subsequent loss of the repeats is dispensable in nervous tissue post-e11 (*Nestin-cre*). The transgenic constructs **d)** RC and

**e)** DR (Park *et al*, 2011) place either the entire *Rasgrf1* locus (RC) or the *Rasgrf1* ICR (DR) at a locus unlinked to *Rasgrf1*. Endogenous locus labelled "chr9;" transgene labelled "tg." At RC, the DMR methylation is sufficient to effect *Rasgrf1* expression; at DR, the repeats are sufficient to impart methylation to the DR DMR. **f)** Insertion of the *Rasgrf1* repeats and DMR between the normally non-imprinted *Wnt1* coding sequence and its annotated enhancer (Taylor *et al*, 2016) demonstrates that the repeats impart methylation at its associated DMR; however, this was insufficient to provoke imprinted expression of *Wnt1*.



### I.E.1 A lncRNA is expressed antisense to the *Rasgrf1* DMR and is targeted by the piRNA pathway.

Our lab has previously identified a lncRNA transcribed antisense to the DMR using the *Rasgrf1* repeats as a promoter. lncRNA expression patterns coincide with the epigenetic reprogramming of the embryonic male gonad, reaching its highest levels at embryonic day 16.5.

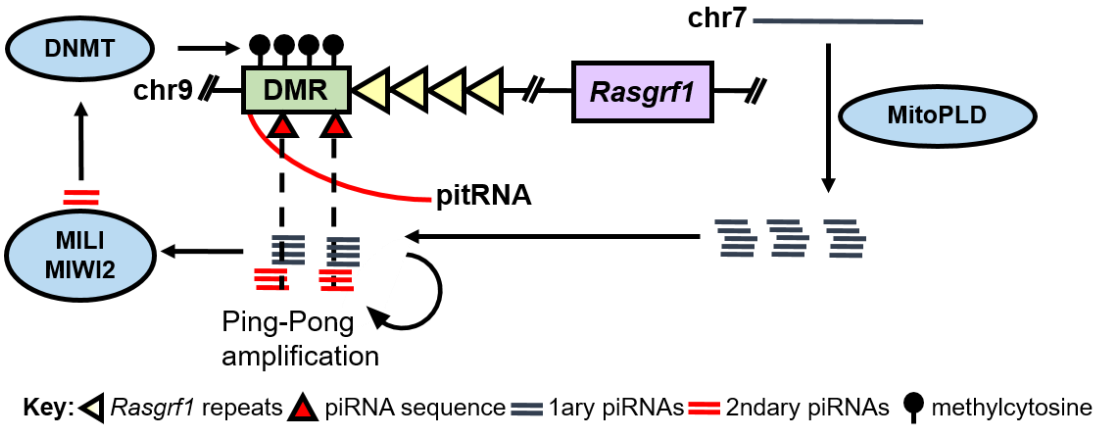
5’ RACE maps the transcriptional start site of this lncRNA to the repeats. Consistent with the repeats acting as a promoter for the lncRNA, repeats-null e16.5 testes express the lncRNA at less than 10% of wild type. Additional characterization of the lncRNA revealed that it is at least 1kb long and is unspliced.

In addition, the lncRNA sequence includes an RMER4B sequence, an LTR-type retrotransposon. In rat and mouse, the two rodent species where *Rasgrf1* is imprinted (Pearsall *et al*, 1999), the RMER4B sequence is conserved at both the DMR and a piRNA cluster on chr7. Small RNA sequencing revealed piRNAs that map to both locations; moreover, these small RNAs display a 10 nt overlap between chr7 and *Rasgrf1* DMR piRNAs, a signature characteristic of the piRNA Ping-Pong cycle (described earlier). Thus we termed this the “piRNA-targeted RNA,” or “pitRNA”.

Further evidence for the piRNA involvement for methylation at *Rasgrf1* comes from piRNA component knockouts. *Rasgrf1* DMR methylation is reduced to 23.3% in the spermatogonia of *Mili* knockouts, and 16.8% in *MitoPLD* knockouts. Moreover, *MitoPLD* knockout e16.5 testes, the pitRNA is

upregulated approximately five fold, and pitRNA cleavage fragments disappeared, suggesting failure of pitRNA targeting and cleavage in *MitoPLD* mutants. Notably, loss of *Mili* and *MitoPLD* did not lead to complete ablation of methylation at the DMR, suggesting that a secondary piRNA-independent mechanism might contribute to methylation at the DMR.

**Fig I.4. Model for regulation of DNA methylation at *Rasgrf1.*** The paternal allele is depicted. A series of repeats at *Rasgrf1* act as a promoter for a non-coding RNA (pitRNA) transcribed antisense to the differentially methylated region (DMR). A cluster of piRNAs mapping to mouse chr7 have homology to the pitRNA. These are targeted to the pitRNA to generate secondary piRNAs. piRNAs are loaded onto MILI and MIWI2, which then mediate DNA methyltransferase activity at the DMR.




Transgenic experiments demonstrate that the under the control of the repeats, the pitRNA-mediated mechanism is restricted to acting in *cis*, at its allele of origin (Park *et al,* 2011). However, as will be discussed in the following section, other data support a potential ability for the pitRNA to act trans-allelically and to impart transgenerational effects.

### I.E.2 Paramutation, an additional facet of the *Rasgrf1* system

Paramutation describes the occurrence of heritable epigenetic changes

unlinked to a genetic variant. Paramutation involves trans-homologue effects, whereby one allele (the paramutagenic allele) can exerts epigenetic changes on its homologous allele (the paramutable allele); these changes are maintained through meiosis and heritable even in the absence of the original paramutagenic allele.

First described in peas (Bateson and Pellew 1915) and crocus (Catcheside 1947), the term “paramutation” was coined to describe a phenomenon at the *r1* locus in maize, where the R-mb allele, which produces maize ears with marbled or stippled purple and yellow kernels, exerts effects on the dominant R-r allele, which produces uniformly purple seeds. Crossing heterozygous R-r/R-mb maize strains to r plants, which produce uniformly yellow seeds, results in r/R-r strains that inevitably exhibit a stippled phenotype, though they had not inherited the R-mb allele (Brink 1956).

Paramutation and paramutation-like phenomena have since been characterized in other plant species, for example, the *sulfurea* locus in tomato species (Hagemann and Berg 1971), the hygromycin phosphotransferase transgene in *Arabidposis* (Scheid *et al*, 2003), as well as in animals including *Drosophila* (Hermant *et al*, 2012; de Vanssay *et al*, 2012; Capovilla *et al*, 2017) and *C. elegans* (Shirayama *et al*, 2012; Sapetschnig *et al*, 2015). In mouse, paramutation-like events have been observed at the *kit* locus (Rassoulzzadegan *et al,* 2006), the *Phactr3* locus (Worch *et al,* 2008), and the *Rasgrf1* locus (Herman *et al,* 2003). In human, a potential paramutation-like effect has been described at the *VNTR* insulin effect locus (*Bennett et al*,

1997).

### I.E.2.i Characteristics of paramutable loci and alleles

In maize, *Drosophila*, and potentially in mammals, paramutation events are associated with repetitive sequence. Characterization of paramutation typically involves 1) recognition and monitoring of the phenotypic change induced by the paramutagenic allele, such as the stippled phenotype at the aforementioned R locus in maize or the white-tail phenotype in mouse, and 2) tracking the heritable alterations of the paramutable or paramutated allele. A recurrent theme of paramutable loci include enrichment for repetitive sequences are present at many loci that undergo paramutation (Kermicle *et al*, 1995; Sidorenko and Peterson, 2001; Stam *et al*, 2002), though not all (Rassoulzadegan *et al*, 2002; Qin and Arnim 2002). Indeed, some data support a potential role of tandem repeats in mediating paramutation. In a series of transgenic experiments in maize, the tandem repeats from the b1 locus were sufficient to induce paramutation and gene silencing (Belele *et al,* 2013); a repeat binding protein CBBP has been shown to permit paramutation at the b1 locus in maize (Brzeska *et al*, 2010).

Distinguishing between the paramutable and paramutagenic alleles is possible at the epigenetic level. The paramutant allele tend to display different epigenetic marks secondary to *trans* homologue effects by the paramutagenic allele. In many cases, the paramutagenic allele is hypermethylated relative to the paramutable allele (Walker 1998; Sidorenko *et al*, 2001; Mittelstein Scheid *et al*, 2003, Rassoulzadegan *et al*, 2002; Meyer *et al*, 1993). At the *b1* locus in

maize, the paramutable B-1 allele displays a more open chromatin structure as evaluated by DNAse hypersensitivity (Stam *et al*, 2002); the paramutagenic B' allele is consistently marked with repressive histone modifications H3K9Me and H3K27Me (Haring *et al*, 2010).

### I.E.2.ii Potential mechanisms for *trans* homologue effects

An RNA-based mechanism, where transcription from one allele targets the homologous allele for epigenetic modifications is supported in all species in which paramutation has been observed. First clues come from the known *trans* factors required for paramutation. In maize, the screens for factors that impacted paramutation identified *mop1*, an RNA-dependent RNA polymerase (Dorweiler *et al*, 2000; Alleman *et al*, 2006), *mop2*, a subunit for the plant-specific RNA polymerase V; *rmr6*, a subunit of the DNA-dependent RNA polymerase IV (Hollick *et al*, 2005; Erhard *et al*, 2009); and *rmr1*, a SNF2-like ATPase (Hale *et al,* 2007). Further, paramutation at the b1 locus in maize is demonstrated to be RNA-mediated (Arteaga-Vazquez *et al*, 2010). Together, these data support the requirement for the RNA-dependent DNA methylation (RdDM) pathway (reviewed in Matzke *et al,* 2009) in effecting paramutation in plants.

In animals, an RNA-mediated mechanism is also supported (Rassoulzadegan *et al*, 2006). Microinjection of small RNAs targeting the *Cdk9* gene or fragments of the *Cdk9* coding sequence itself produced a heritable cardiomyopathy induced by overexpression of Cdk9. In this model, methylation at *Cdk9* was unchanged—perhaps not surprisingly, given that the

paramutated allele had been activated rather than repressed (Wagner *et al*, 2008). Similarly, microinjection of miR-124, an miRNA important for central nervous system development (Cao *et al*, 2007; Visvanathan *et al*, 2007; Makeyev *et al*, 2007), led to a heritable “giant” phenotype and an increase in *Sox9* expression; further, heritable chromatin changes were observed, where the affected Sox9 locus had an increased enrichment for the H3K9Me3 at a putative regulatory region upstream of Sox9 (Grandjean *et al*, 2009). These findings also support an RNA-mediated mechanism for paramutation. However, to date, an RNA-dependent RNA polymerase such as MOP1 has not been characterized in animals, which may imply a divergent mechanism in animals. Parallels to RdDM in animals do exist, notably the piRNA-mediated DNA methylation of transposable elements by piRNAs. In *Drosophila*, paramutation is linked to maternal inheritance of piRNAs (de Vanssay *et al*, 2012) and requires components of the piRNA pathway (Hermant *et al*, 2015), but the factors directly responsible for RNA-guided DNA methylation have yet to be identified. With the recent discovery of murine Dnmt3c, which is known to methylate transposable element sequence in the male germline (Barau *et al*, 2016), as well as MORC1 (Pastor *et al*, 2014), which is also required for transposable element methylation and displays DNA binding capacities, a mechanism for paramutation in animals may not be far behind.

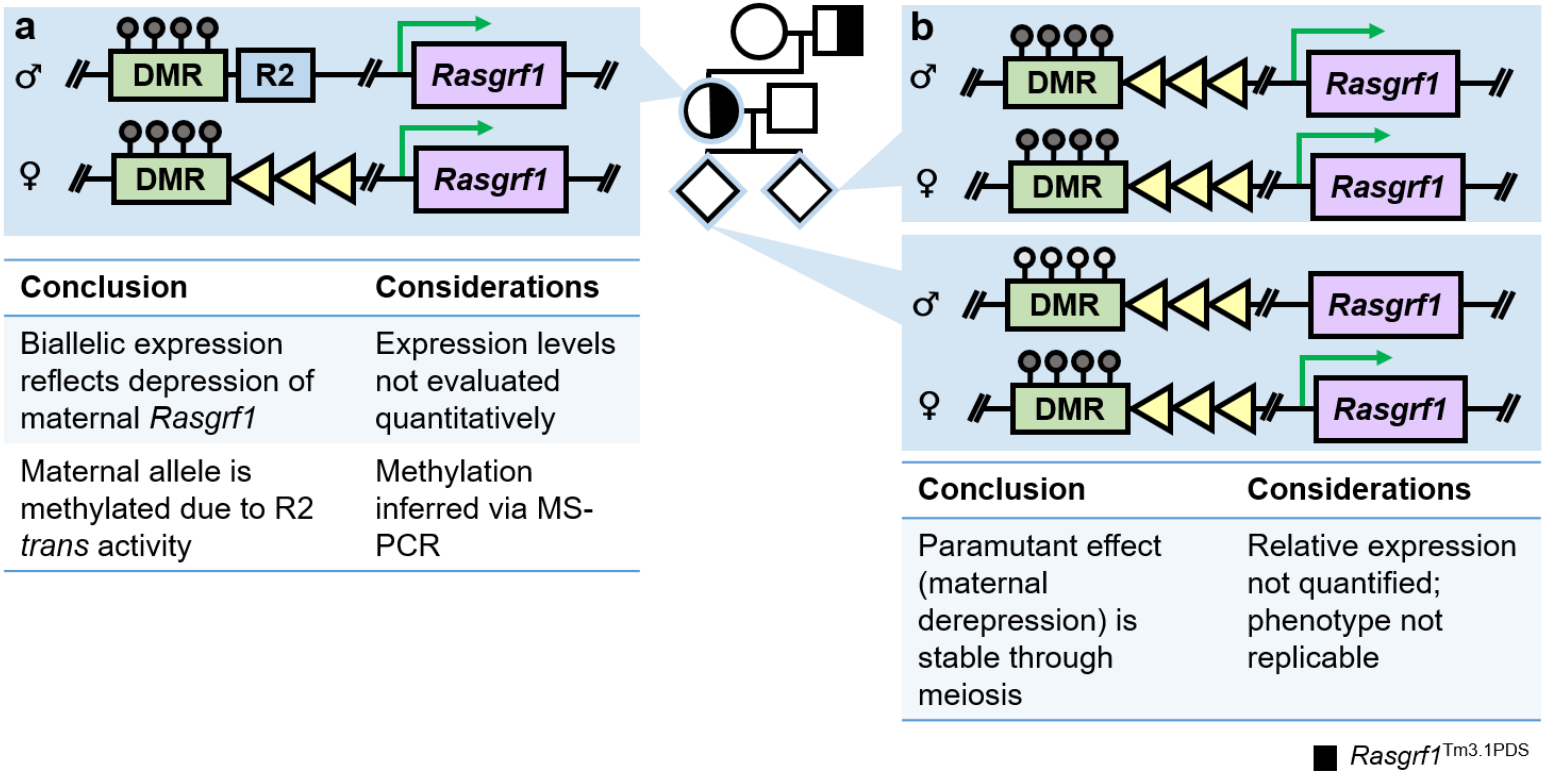
#### I.E.2.iii Paramutation at *Rasgrf1*

Paramutation with respect to the *Rasgrf1* locus was reported for a targeted mutant, *Rasgrf1*$^{tm3.1PDS}$ or tm3.1, carrying the imprinting control region

of the *Igf2r* gene, which acts as a promoter for the lncRNA *AIRN* (Wutz *et al*, 1997; Lyle *et al*, 2000). Upon analysis of *Rasgrf1* expression and methylation patterns, tm3.1 animals displayed a curious phenotype. In the F1 generation, paternally-inherited tm3.1 led to biallelic *Rasgrf1* expression, which was interpreted as derepression of the maternal allele. In the F2 generation, the genetically wild-type offspring of derepressed tm3.1 heterozygous females maintained the paramutant phenotype (**Fig I.5**). As such, the tm3.1 allele meets the two criteria for paramutation: First, that a paramutant allele can exert its effects on a paramutable allele; and second, that these effects can persist through meiosis.

With the later discovery of the pitRNA (Watanabe, 2011), the tm3.1 system could be explained by an RNA-loading effect. For example, should tm3.1 females express abnormally high levels of pitRNA in the germline, high levels of pitRNA in oocytes that lack the tm3.1 allele could exert effects on methylation and *Rasgrf1* expression in genetically wild-type offspring. However, the tm3.1 allele does not permit precise control of pitRNA levels or timing: Whether the pitRNA can itself transallelically or transgenerationally, as observed in the tm3.1 system, remains to be explored.

**Fig I.5. Paramutation-like effects observed with *Rasgrf1*$^{tm3.1PDS}$. a)** Animals with a paternally-inherited tm3.1 allele express *Rasgrf1* biallelically, as shown by the green arrows; methylation of the wild-type maternal allele was inferred by MS-PCR (gray filled lollipops). **b)** Genetically wild-type offspring of maternally methylated, derepressed R2 females retained the derepressed expression phenotype. A subset of animals also expressed monoallelically from the maternal allele. Conclusions from the genotypes examined are listed below each allele schematic, as are accompanying considerations for the methods utilized.




| Conclusion | Considerations |
|---|---|
| Biallelic expression reflects depression of maternal *Rasgrf1* | Expression levels not evaluated quantitatively |
| Maternal allele is methylated due to R2 *trans* activity | Methylation inferred via MS-PCR |

| Conclusion | Considerations |
|---|---|
| Paramutant effect (maternal derepression) is stable through meiosis | Relative expression not quantified; phenotype not replicable |

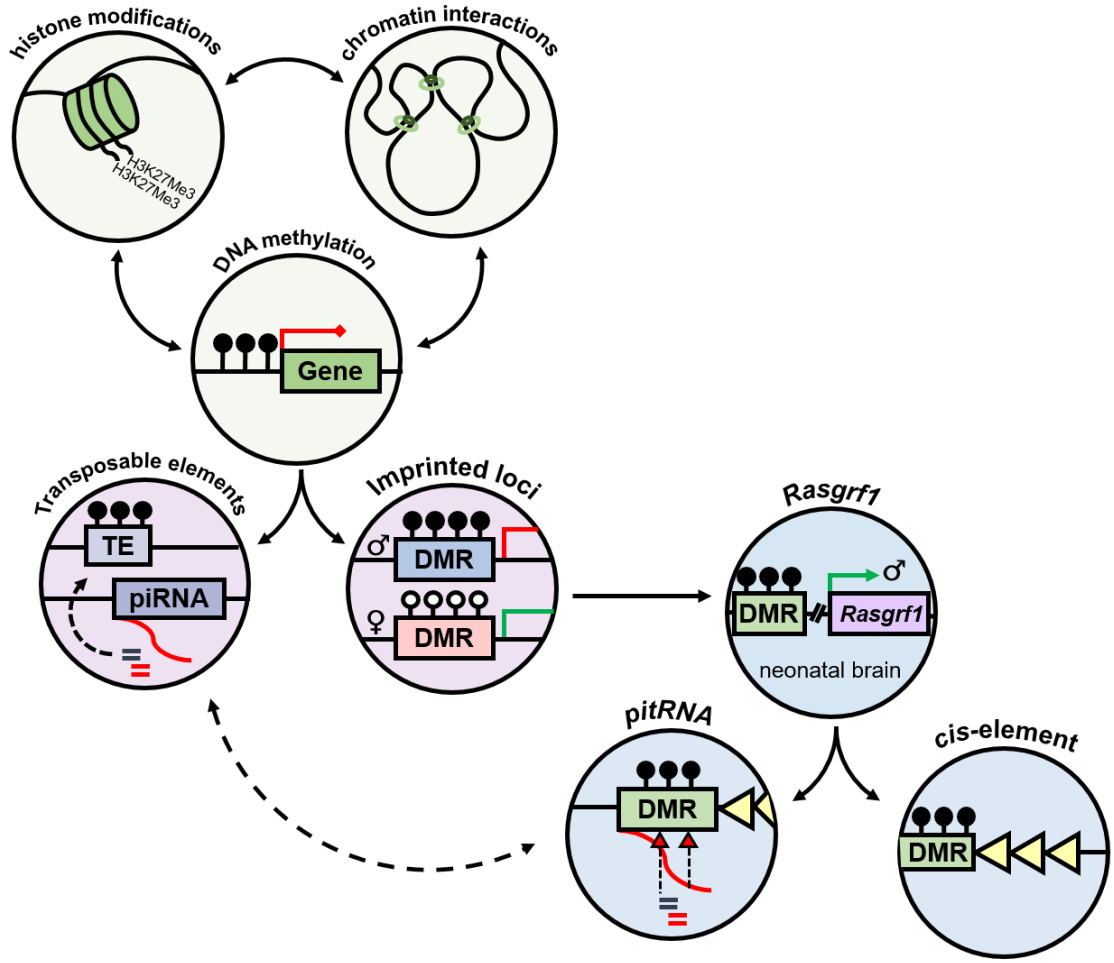
### I.E.3 Summary of Chapter I and findings at *Rasgrf1*

In Chapter I, we have defined and discussed three key epigenetic modifications: DNA methylation, histone modification, and chromatin interactions. Two systems with known mechanisms or elements for DNA methylation include imprinted loci and transposable element. At *Rasgrf1*, imprinted in neonatal brain in mouse, a *cis*-element required for DNA methylation as well as a piRNA-targeted lncRNA, the pitRNA, have been characterized (Figure 1.6).

From our existing data, the following can be concluded: First, that the repeats are required for DNA methylation at *Rasgrf1*. Second, that the repeats are required for expression of the pitRNA. Third, that the pitRNA is processed into piRNAs. Fourth, that components of the piRNA pathway are required for complete methylation at *Rasgrf1*. Fifth, that under the control of the repeats, this mechanism is restricted to acting in *cis*. What remains undefined is the sufficiency of the pitRNA alone to impart methylation at the *Rasgrf1* repeats, and whether manipulations of the pitRNA can contribute to transallelic or transgenerational effects. Further, it is unknown whether the repeats function in an additional role beyond driving pitRNA transcription. The experiments in Chapters II and III seek to define, respectively, the sufficiency of the pitRNA to impart methylation at *Rasgrf1*, and the factors that may bind the *Rasgrf1* repeats, thereby driving pitRNA expression and potentially mediating other functions.

**Figure 1.6. Summarizing figure of Chapter I. Green panels**: Three known

categories of epigenetic modifications include DNA methylation, histone modifications, and chromatin interactions. **Lavender panels**: Two classes of genes, transposable elements and imprinted loci, are targeted for DNA methylation. A know small RNA-mediated pathway, the piRNA pathway (piRNAs denoted by blue and red bars), target transposable elements for methylation and silencing. **Blue panels**: *Rasgrf1* is a model imprinted locus with a characterized *cis*-element, a set of tandem repeats (yellow triangles) adjacent to the *Rasgrf1* differentially methylated domain (DMR); a long noncoding RNA, the pitRNA, is transcribed antisense to the *Rasgrf1* DMR and is targeted by piRNAs. The characterization of the pitRNA represents previously unknown link between thhe piRNA pathway and regulation of methylation at imprinted loci (dotted arrow).



# II RNA-INDEPENDENT REGULATION OF IMPRINTING AT *RASGRF1*

This is a pre-print of a manuscript submitted for review to *PLoS Genetics*.

Chu ET, Hofstedt M, Taylor DH, and Soloway PD. RNA-independent regulation of imprinting at *Rasgrf1*.

## II.A Abstract

Long noncoding RNAs (lncRNAs) have garnered much attention as possible links between DNA sequence and the protein factors that mediate DNA methylation. However, the mechanisms by which DNA methylation is directed to specific genomic locations remain poorly understood. We previously identified a lncRNA in mouse, the pitRNA, that was implicated in the control of DNA methylation at the imprinted *Rasgrf1* locus. The pitRNA is transcribed in the developing male germline antisense to the differentially methylated region (DMR) that harbors paternal allele methylation, and is driven by a series of tandem repeats that are necessary for imprinted methylation. MitoPLD, a factor necessary for piRNA biogenesis, both processes piRNAs from the pitRNA and is necessary for complete methylation at the locus, along with piRNA binding proteins. Using two independent mouse systems where pitRNA transcription is driven by the doxycycline-inducible Tet Operator, we demonstrate that pitRNA transcription across the DMR is insufficient for imprinted methylation, and that the *Rasgrf1* repeats have additional, critical *cis*-acting roles for imparting DNA methylation to *Rasgrf1*, independently of their control of pitRNA transcription. Furthermore, pitRNA overexpression and oocyte loading of pitRNA is insufficient to induce

transallelic and transgenerational effects previously reported for *Rasgrf1*. Notably, manipulation of the pitRNA with the *TetOFF* system led to transcriptional perturbations over a broad chromosomal region surrounding the inserted Tet Operator, revealing that the effects of this regulatory tool are not localized to a single target gene.

## II.B Summary

DNA methylation is a heritable genetic modification known to impact vital biological processes. While the proteins that establish, maintain, and remove DNA methylation are well characterized, the mechanisms by which these proteins are directed to specific genetic sequences are poorly understood. We have previously demonstrated that DNA methylation at the imprinted *Rasgrf1* locus requires a DNA element with a series of tandem repeats. These repeats act as a promoter for a long noncoding RNA, the pitRNA, which is targeted by a small noncoding RNA pathway known to silence viral elements in the male germline via DNA methylation. We queried the sufficiency of the pitRNA to mediate DNA methylation at *Rasgrf1*. We show that, in the absence of the repeats, the pitRNA expression is insufficient to establish imprinted methylation. This work supports a pitRNA-independent mechanism for methylation at *Rasgrf1*, and a critical cis-acting role for the tandem repeats separate from their control of pitRNA transcription.

## II.C Introduction

DNA methylation is essential for appropriate embryonic development.

While the *trans*-acting factors required to establish (Okano *et al*, 1999; Kato *et al*, 2007; Bourc'his D *et al*, 2001; Barau *et al*, 2016; Jain *et al*, 2017), maintain (Li *et al*, 1992; Howell *et al* 2001; Nakamura *et al*, 2006; Payer *et al*, 2003; Quenneville *et al*, 2011; Shibata *et al*, 2011), and remove (Ito *et al*, 2011; Cortellino *et al*, 2011; Popp *et al*, 2010; Bhutani *et al*, 2010) DNA methylation have been identified, little is known of *cis*-acting elements that direct these *trans*-acting factors to specific genomic locations (Chotalia *et al*, 2008; Frohlich *et al*, 2010; Kantor *et al*, 2004; Pant *et al*, 2003; Pant *et al*, 2004).

One such *cis* element exists at the imprinted *Rasgrf1* locus. In mouse and rat, *Rasgrf1* is paternally methylated and expressed in the neonatal brain. Imprinted expression in mouse is controlled by a differentially methylated region (DMR) 30 kb upstream of *Rasgrf1* coding sequence, and requires a 1.6 kb stretch of tandem repeats immediately adjacent to the DMR. Targeted deletion of the *Rasgrf1* repeats (*Rasgrf1*$^{tm1PDS}$, *tm1*) leads to loss of methylation at the *tm1* DMR in the male germline, and loss of imprinted *Rasgrf1* expression (Yoon *et al*, 2002). The repeats also play a role in the maintenance and spreading of DMR methylation in the embryonic somatic lineage after fertilization, though they are dispensable past the epiblast stage (Lindroth *et al*, 2008; Holmes *et al*, 2006).

Our lab previously characterized a long noncoding RNA (lncRNA), the pitRNA, which is driven by the *Rasgrf1* repeats and is expressed in the embryonic male gonad (Watanabe *et al*, 2011). LncRNAs, with their ability to recruit and bind effector proteins (reviewed in Lee, 2012; Taylor *et al*, 2015),

represent a molecular class that could bridge the gap between the protein effectors of local epigenetic states if they recruit the effectors while being transcribed. Indeed, lncRNAs have been implicated in diverse biological processes, and have been proposed to modulate gene expression via a number of mechanisms including recruitment of histone modification complexes (Nagano *et al*, 2008; Plath *et al*, 2003), transcriptional interference (Latos *et al*, 2012), and enhancer regulation (Orom *et al*, 2010).

Using the *Rasgrf1* repeats as a promoter, the pitRNA is transcribed antisense to the DMR, spanning an RMER4B element, an LTR-type retrotransposon. The repeats and RMER4B element are conserved at the *Rasgrf1* DMR in species where *Rasgrf1* is imprinted (Pearsall *et al*, 1999). The pitRNA is processed into secondary piRNAs by the piRNA pathway, which is required for DNA methylation and transcriptional silencing of retrotransposons (Kuramochi-Miyagawa *et al*, 2008), and also for full methylation at the *Rasgrf1* DMR. Given the apparent importance of the pitRNA and piRNA pathway in controlling methylation at *Rasgrf1*, we hypothesized that aberrant expression of the pitRNA could explain transallelic and transgenerational effects previously reported at *Rasgrf1* (Herman *et al*, 2003). Indeed, aberrations in non-coding expression have been associated with such effects in other systems (Alleman *et al*, 2006; Rassoulzadegan *et al*, 2006; de Vanssay *et al*, 2012).

More recently, our lab targeted the *Wnt1* locus, inserting the *Rasgrf1* repeats and DMR between the *Wnt1* coding sequence and its annotated

enhancer (*Wnt1*$^{DR}$). We found that when paternally transmitted, the *Wnt1*$^{DR}$ allele was methylated, recapitulating patterns of imprinted methylation found at *Rasgrf1*. However, pitRNA expression from the *Wnt1*$^{DR}$ was extremely low (less than 2% of the pitRNA expressed from the endogenous locus) (Taylor DH *et al*, 2016). These data suggested that the *Rasgrf1* repeats could impart methylation to their associated DMR independent of robust pitRNA expression.

None of the systems described above have uncoupled the pitRNA from the *Rasgrf1* repeats to ascertain necessity or sufficiency of either individual element for methylation in *cis* at the endogenous *Rasgrf1* locus. Here, we directly queried the sufficiency of the pitRNA to establish methylation at *Rasgrf1* using a targeted mutation in mouse, *Rasgrf1*$^{tm5.1PDS}$ (*tm5.1*) where the *Rasgrf1* repeats were replaced by the Tet Operator. This enabled inducible control of pitRNA expression through combination of *tm5.1* with one of two transactivating proteins: *TetON*, which binds the Tet Operator and drives pitRNA expression in the presence of doxycycline (Urlinger *et al*, 2001); and *TetOFF*, which binds the Tet Operator and drives pitRNA expression in the absence of doxycycline (Gossen *et al*, 1992; Baron *et al*, 1997). We found that induction of the pitRNA at physiologic levels in male gonocytes was insufficient to impart methylation to the *tm5.1* DMR, revealing a critical role for the repeats in methylation control, independent of their regulation of pitRNA transcription, consistent with our findings with *Wnt1*$^{DR}$. Using *tm5.1* as well as a transgenic allele, TetO$\Delta^{Tg}$, we also determined that the pitRNA overexpression was

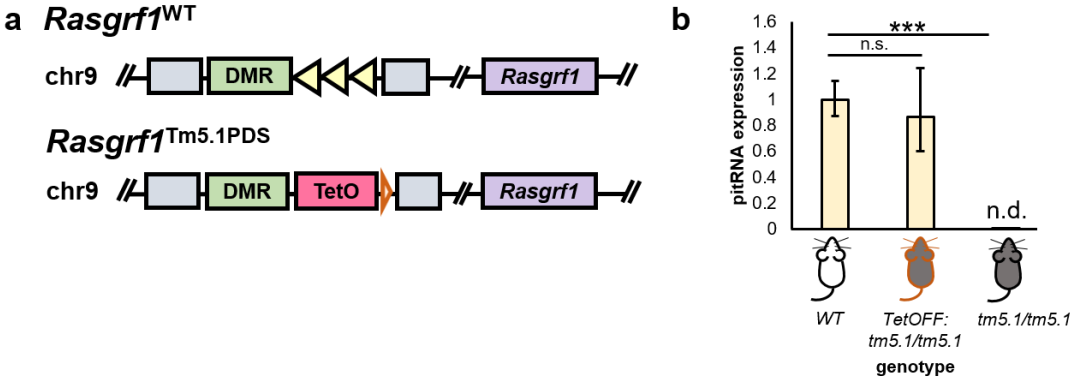
insufficient to induce transgenerational or transallelic effects on *Rasgrf1* expression or methylation. Finally, in addition to enabling control of pitRNA expression as designed, *TetOFF* transactivation of *tm5.1* activated transcription across a broad chromatin domain previously shown to exhibit interactions, and that activation was not confined to the target sequences at the DMR. Our data identify a role for *Rasgrf1* repeats as a *cis*-element directing DNA methylation, independently of the pitRNA it drives. Furthermore, we show that expression patterns controlled by engineered Tet repressor proteins can be exerted over large regions of the genome.

## II.D Results

### II.D.1 Generation of *Rasgrf1*$^{tm5.0PDS}$.

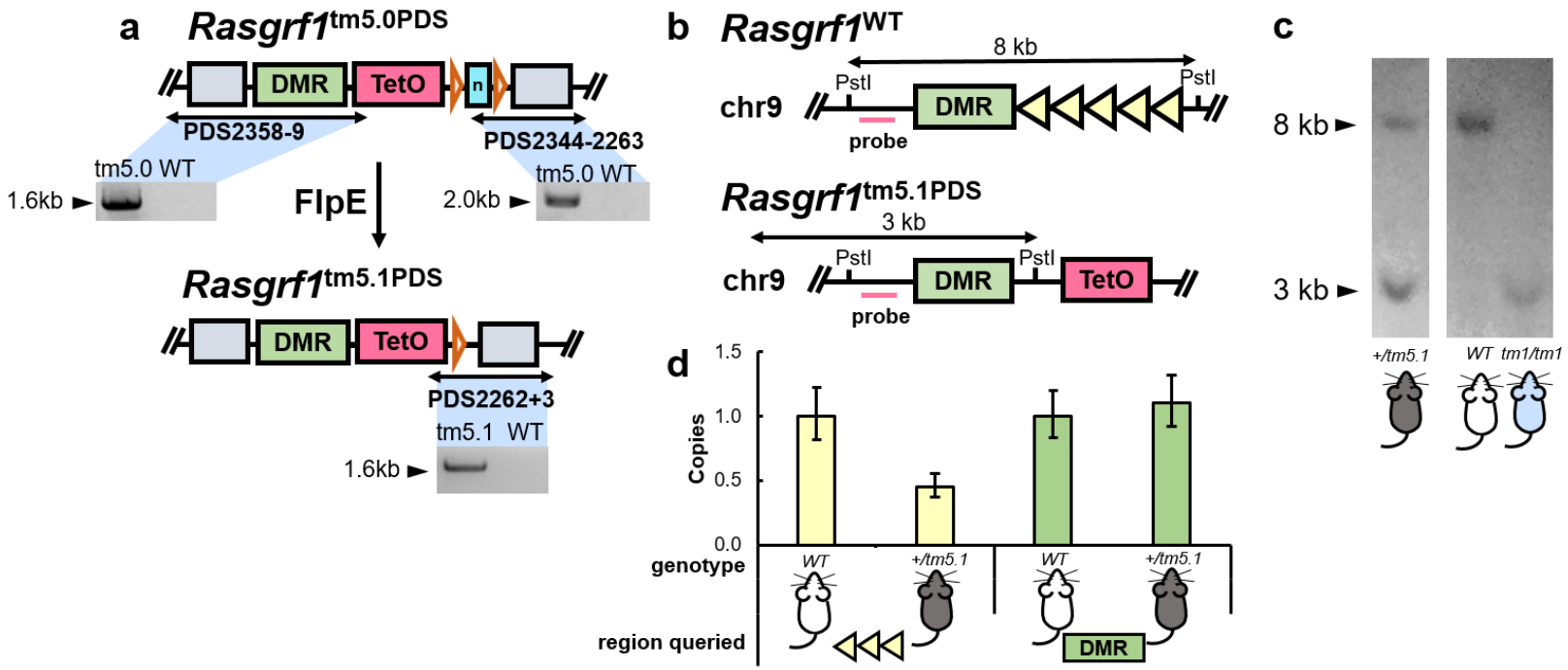
We successfully generated the targeted mutant *Rasgrf1*$^{tm5.1PDS}$ (*tm5.1*), where the endogenous repeats were replaced with the Tet Operator (Fig 1a).

**Figure 1. Schematics and pitRNA expression of the wild-type *Rasgrf1* and *Rasgrf1*$^{tm5.1PDS}$ alleles**. **a)** Schematics of the wild-type *Rasgrf1* and *Rasgrf1*$^{tm5.1PDS}$ alleles. Green box represents the *Rasgrf1* differentially methylated region (DMR). The Imprinting Control Region (ICR) lies 30 kb upstream of *Rasgrf1* coding sequence (purple box). Yellow triangles connote the *Rasgrf1* repeats. Grey blocks connote 5' and 3' homologous arms of the targeting vector pETC6, described in Methods. Orange triangle downstream of *Rasgrf1*$^{tm5.1PDS}$ connote residual frt site. Figures are not drawn to scale. **b)** pitRNA expression postnatal day 1 testes of wild-type, *tm5.1/tm5.1*, and *TetOFF*:*tm5.1/tm5.1* animals. *TetOFF* transactivates *tm5.1* to physiological levels; *tm5.1* homozygotes express undetectable levels of pitRNA in the absence of *TetOFF*. Error bars represent standard error across biological triplicates. ***, $p < 10\text{-}6$; n.s., not significant; n.d., not detected at 40 cycles.




Allelic structure was validated by Southern blot, and Sanger sequencing of PCR products that spanned the vector ends and the endogenous sequence at the target locus, as well as by copy number qPCR (Fig S1a-d).

**Figure S1. Validation of *Rasgrf1*$^{tm5.1PDS}$. a)** Generation of *Rasgrf1*$^{tm5.1PDS}$. From *Rasgrf1*$^{tm5.0PDS}$. ***Rasgrf1*$^{tm5.0PDS}$** (**a**, upper schematic) was generated via CRISPR-Cas9-mediated homology-directed repair in v6.5 embryonic stem cells. Targeting was confirmed by Sanger sequencing of PCR products generated with primers PDS 2359-8 and PDS 2344-2263, which respectively span the junctions of the 5' and 3' homologous arms of the pETC6 vector (grey boxes), and the flanking sequences of the *Rasgrf1*$^{tm5.0PDS}$ allele. PDS 2344 falls within the neo resistance cassette (n). PDS 2358 sequence falls partially within TetO. DNAs from Wild-type (WT) animals show no product for either PDS 2359-8 or PDS 2344-2263. ***Rasgrf1*$^{tm5.1PDS}$** (*tm5.1*, **a**, lower schematic) was generated by crossing *Rasgrf1*$^{tm5.0PDS}$ males to females constitutively expressing FlpE recombinase (FlpE). Recombination was confirmed with Sanger sequencing of PCR products generated with primers PDS 2262-2263. PDS 2262 maps to Tet Operator sequence, and the product contains the single residual frt site (orange triangle) remaining after recombination. PCR products shown below the amplicon schematics arise only when using DNAs from mutant animals. **b)** Schematic for Southern blot shown in **c)**; probe location, shown with pink line, is outside of the targeting vector; location of *Pst*I sites are indicated. Yellow triangles indicate the *Rasgrf1* repeats. Homologous arms are omitted for clarity. **c)** Southern blot of a *tm5.1* heterozygote vs. a wild-type animal and a *Rasgrf1*$^{tm1}$ homozygote (*tm1*); in these animals, the probe detects the same 3kb fragment in *tm1* as is *tm5.1*. **d)** Copy number qPCR support targeting of the endogenous allele: *tm5.1* heterozygotes have half the copy number of the repeat element as WT animals, but the same number of DMR sequences. Error bars represent standard error across biological duplicates.


a
Rasgrf1tm5.0PDS
DMR
TetO
n
PDS2358-9
PDS2344-2263
tm5.0 WT
1.6kb
FlpE
tm5.0 WT
2.0kb
Rasgrf1tm5.1PDS
DMR
TetO
PDS2262+3
tm5.1 WT
1.6kb
b
Rasgrf1WT
8 kb
chr9
PstI
PstI
DMR
probe
Rasgrf1tm5.1PDS
3 kb
chr9
PstI
PstI
DMR
TetO
probe
d
Copies
1.5
1.0
0.5
0.0
genotype
WT
+/tm5.1
WT
+/tm5.1
region queried
DMR
c
8 kb
3 kb
+/tm5.1
WT
tm1/tm1

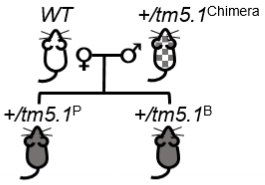
### II.D.2 The *tm5.1* allele lacks DMR methylation and *Rasgrf1* expression, like the *Rasgrf1*$^{tm1}$ repeat-deficient allele.

We first characterized *tm5.1* in the absence of a transactivator. We expected that, in the absence of a transactivating protein, *tm5.1* would neither accrue methylation at its DMR or impart imprinted expression at *Rasgrf1*, similar to the repeat-deficient allele, *Rasgrf1*$^{tm1}$ (*tm1*). The *tm5.1* DMR was hypomethylated when paternally transmitted (+/*tm5.1*, Fig S2e), leading to minimal expression of *Rasgrf1* in the brain as measured by qRT-PCR (FigS2d); +/*tm5.1* animals were on average of lower body weight than wild-type littermates (Fig S2f), consistent with findings that repeat-deficient animals are underweight (Drake *et al*, 2009).

Also consistent with *tm1*, a portion of +/*tm5.1* animals were methylated at and expressed *Rasgrf1* at wild-type levels from the *tm5.1* allele. Also as seen with the *tm1* allele, a portion of mice with the +/*tm5.1* genotype had *tm5.1* methylation and expression in the N2 and N3 generations (4 out of 14 animals in the N2 generation; 2 out of 15 animals in the N3 generation). Consistent with findings using repeat-deficient animals (Drake *et al*, 2009), this was a stochastic event, and the methylation status of *+/tm5.1* offspring was not dependent on the methylation state of their +/*tm5.1* fathers (Fig S2a-b).

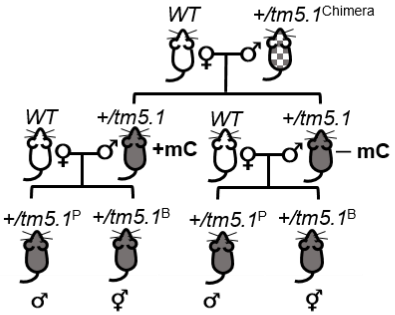
**Figure S2. The *tm5.1* allele lacks imprinted methylation and expression in the absence of transactivator.** Summary of *Rasgrf1* expression, *Rasgrf1* imprinting, and total animals measured in *tm5.1* animals of the **a)** N2 and **b)** N3 generations. The majority of animals inheriting *tm5.1* paternally (+/*tm5.1*) expressed *Rasgrf1* biallelically as assayed by endpoint PCR, but levels were only 2% those seen in wild-type (WT) mice, and the alleles were unmethylated. A subset of +/*tm5.1* animals (4 of 10 in the N2, and 2 of 15 in the N3 generation) expressed *Rasgrf1* at WT levels from the paternal *tm5.1* allele. N3 animals expressing *Rasgrf1* paternally and at wild-type levels did not arise from methylated fathers, indicating that sporadic *tm5.1* DMR methylation was not an inherited state, consistent with findings with the *Rasgrf1*$^{tm1PDS}$ (Yoon et al, 2002)**Error! Bookmark not defined.**. **c)** Allele-specific expression analysis. A male chimera prepared using C57BL/6 blastocysts, and v6.5 ES cells with the *tm5.1* allele on the 129S4 (129) background, was crossed with C57BL/6 females. Neonatal brain cDNA was subjected to endpoint RT-PCR using primers spanning SNPs from the 129 and C57BL/6 backgrounds that harbor distinct *Aci*I sites. Product digestion with *Aci*I produces allele-specific bands, reporting the expressed allele(s). The slowest and fastest migrating bands represent the 129 paternal *Tm5.1* allele; the two middle bands represent the C57BL/6 allele. WT animals, inheriting the C57BL/6 paternal allele, expressed *Rasgrf1* solely from the WT C57BL/6 allele(s). A portion of +/*tm5.1* animals express paternally from the *tm5.1* 129 allele. The majority express biallelically from the maternal C57BL/6 and paternal 129 alleles. **d)** qRT-PCR of *Rasgrf1* in wild-type and +/*tm5.1* animals. Paternally expressing +/5.1 animals express *Rasgrf1* at WT levels, whereas biallelically expressing +/5.1 animals express at 2% of WT, indicating biallelic expression detected in *tm5.1* by endpoint PCR was seen when the normally active paternal allele was silent. Error bars represent standard error across at least biological triplicate. **e)** Targeted bisulfite sequencing (bar graphs) and COBRA analyses (gel images) of the WT DMR in tail gDNA of WT animals (left), and the *tm5.1* DMR of +/*tm5.1* animals with paternal (middle) and biallelic (right) *Rasgrf1* expression. Animals with paternal expression were methylated at the *tm5.1* DMR, whereas animals with biallelic expression were hypomethylated. WT animals have two copies of the WT *Rasgrf1* DMR, one hypermethylated and one hypomethylated; bisulfite analysis of the WT DMR in soma is 50%. Bar graphs report the percentage of total reads with the levels of methylation shown in the key on the right. COBRA queries methylation at five *Bst*UI sites in both the WT and the *tm5.1* DMRs. “+” and “–” connote addition or lack of *Bst*UI. Methylated (+mC) and unmethylated –mC) sites are respectively sensitive or resistant to *Bst*UI digestion. Digestion products are indicated by black arrowheads. The PCR product of the WT DMR shows partial digestion, reporting the different methylation states of the two parental alleles; paternally expressing +/*tm5.1* show full digestion of the *tm5.1* DMR; biallelically expressing +/*tm5.1* animals show no digestion. **f)** +/*tm5.1* animals are lower in body weight compared to their WT littermates consistent with previous findings (Drake *et al*, 2009).

**a**



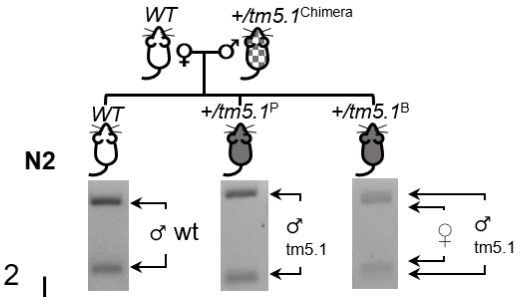

| | N2 | |
|---|---|---|
| ***Rasgrf1* expression ± s.e.** | 1 ± 0.5 | 0.02 ± 0.01 |
| **Parental allele expressed** | ♂ | ♀ |
| **tm5.1 DMD mC** | + | – |
| **n** | 4 | 10 |

**b**

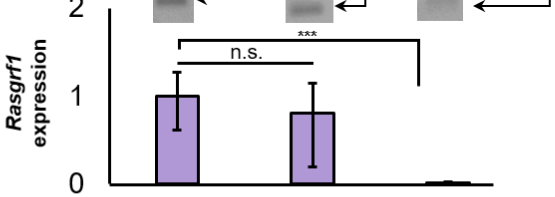



| | N3 | | | |
|---|---|---|---|---|
| **Parental allele expressed** | ♂ | ♀ | ♂ | ♀ |
| **tm5.1 DMD mC** | + | – | + | – |
| **n** | 0 | 6 | 2 | 7 |

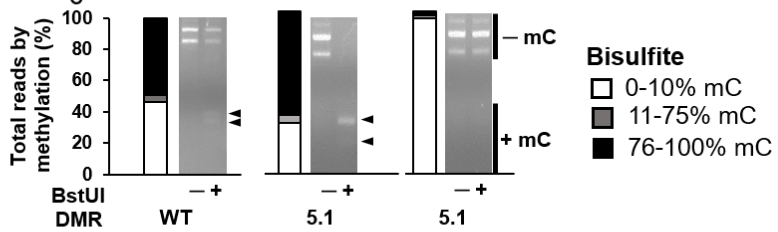
**c**



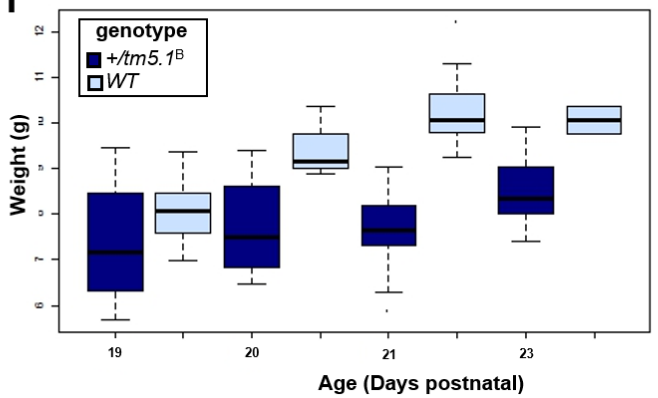

**d**




**e**



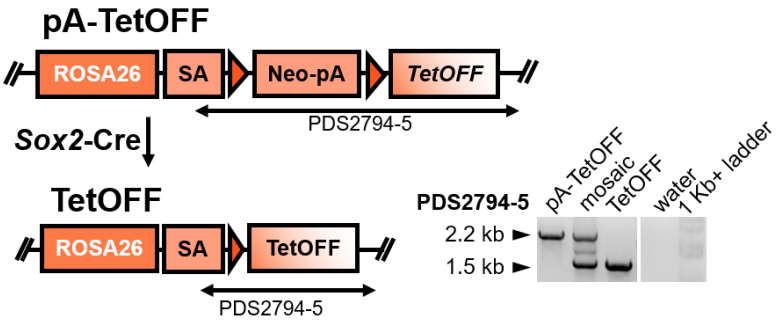

**f**



### II.D.3 Induction of pitRNA from *Rasgrf1*$^{tm5.1PDS}$ via the *TetON* and *TetOFF* systems.

To ascertain transactivator-dependent induction of pitRNA, we then generated *tm5.1* mice expressing one of two transactivating proteins, *TetON* and *TetOFF*. As discussed in Methods, the *TetOFF* allele was generated from the commercially available *pA-TetOFF* allele, where *TetOFF* is preceded by a floxed neomycin-resistance polyadenylation cassette, by crossing pA-*TetOFF* males with females carrying a *Cre* transgene driven by the *Sox2* promoter, thereby driving *Cre*-mediated recombination by embryonic day 6.5 (Hayashi *et al*, 2002). Successful Cre-mediated recombination was confirmed by Sanger sequencing of PCR products spanning the pA-cassette (Fig S3).

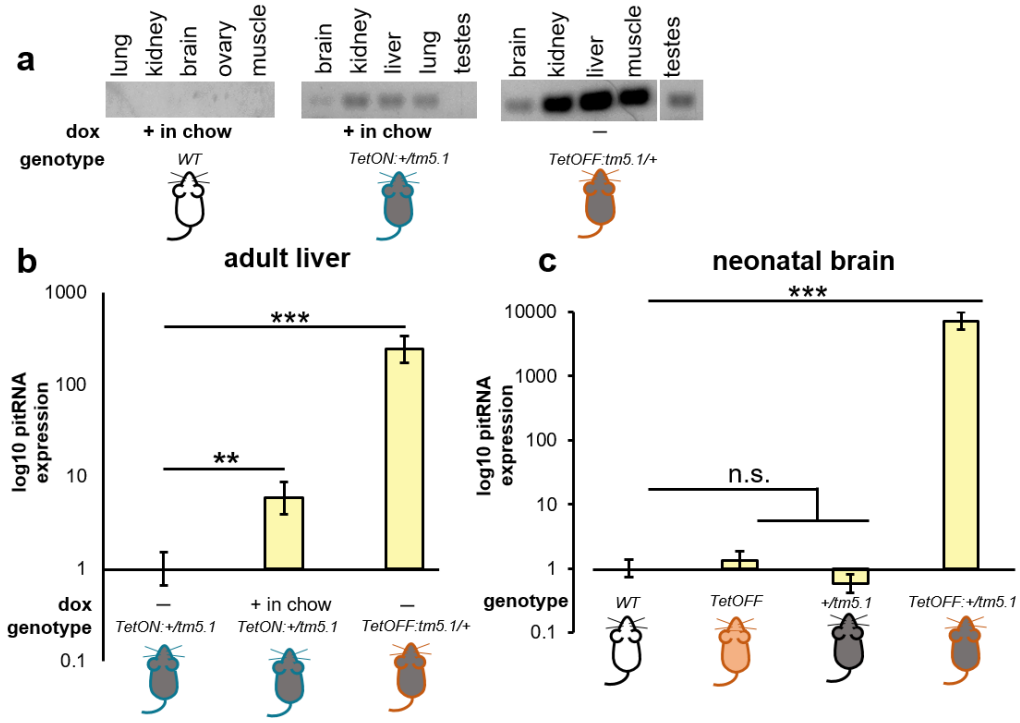
**Figure S3. Generation of the *TetOFF* allele.** *pA-TetOFF* males were crossed with females carrying a *Sox2-Cre* transgene to generate the recombined *TetOFF* allele. PDS 2794 maps to the splice acceptor site of the ROSA26 locus; PDS 2795 maps 3' of the *TetOFF* coding sequence, within wild-type ROSA26 sequence. pA-*TetOFF* animals produce a 2.2kb PCR product for PDS 2794-5 whereas *TetOFF* animals produce a 1.5kb PCR product due to loss of the floxed Neo-polyA cassette. Note that a portion of animals are mosaic for pA-*TetOFF* and *TetOFF* (mosaic); though only fully recombined mice bearing the *TetOFF* allele were analyzed in crosses with the *tm5.1* allele.




We expected pitRNA transcription from the *tm5.1* allele would depend either on the *TetOFF* transgene in the absence of tetracycline, or the *TetON*

transgene in the presence of doxycycline. We assayed pitRNA induction in several adult tissues by endpoint PCR, as well as by qPCR in the neonatal male germline, adult liver, neonatal brain, and oocytes. pitRNA induction from the *tm5.1* allele required a transactivating protein, and was expressed at physiological levels in the neonatal male embryonic germline of males (Fig 1b), and from 10 to 1000-fold wild-type levels in adult tissues (Fig S4, Fig S8) depending on the tissue assayed. In all tissues, pitRNA was not expressed in the absence of a transactivator.

**Figure S4. Successful pitRNA induction in *TetON*:*tm5.1* and *TetOFF*:*tm5.1* tissues. a)** Endpoint RT-PCR for pitRNA in the tissues of transactivated *tm5.1* animals demonstrate induction of pitRNA in several tissues of *TetON: +*/*tm5.1* and *TetOFF: +*/*tm5.1* mice. Low signals from testes are likely due to the blood testes barrier, restricting entry of doxycycline (dox). pitRNA is not detectable by endpoint PCR in wild-type (WT) animals. **b)** qPCR for pitRNA in adult liver shows 10-fold upregulation of pitRNA in *TetON*:*tm5.1* animals fed dox chow, and 500-fold upregulation of pitRNA in *TetOFF*:*tm5.1* livers relative to WT. **c)** qPCR for pitRNA in neonatal brain shows nearly 10,000 fold upregulation of pitRNA in *TetOFF*:*tm5.1* brains relative to WT, *TetOFF*:*+,* and +:*tm5.1* animals. ***, $p < 10^{-6}$; n.s., not significant.




**II.D.4. pitRNA induction in the male germline is insufficient for establishment of germline methylation at *Rasgrf1*.**

Having confirmed that pitRNA could be induced from *tm5.1* using both the *TetON* and the *TetOFF* systems, that expression in the neonatal germline was at physiologic levels, and that *tm5.1* was transcriptionally silent in the absence of a transactivator, we tested whether artificially regulated pitRNA expression was sufficient to impart methylation at the *Rasgrf1* DMR in the male germline, independently of the repeats that normally effect this regulation. To perform this analysis, we used male embryos heterozygous for the *tm5.1* allele that also carried the *TetOFF* or the *TetON* transgenes. We assayed methylation of the *tm5.1* and WT DMRs under conditions where the pitRNA is induced from the *tm5.1* allele. The use of heterozygotes enabled us to monitor methylation of both the *tm5.1* allele, with artificial regulation of the

pitRNA from the Tet Operator, and the wild-type allele, as an internal control, which has natural regulation of the pitRNA from the repeats. gDNA was prepared from the gonocyte and somatic cell fractions of developing male gonads, and assayed for DMR methylation by targeted bisulfite sequencing (Fig 2a-d) and COBRA (Fig S5a,b), using assays specific for the *tm5.1* and wild-type alleles. Both assays revealed that the *tm5.1* DMR was hypomethylated in gonocytes, despite the proper regulation of pitRNA in the germline. In contrast, the wild-type DMR from the same animals was hypermethylated, as expected. This pattern was observed in male gonocytes regardless of the parental modes of inheritance of the two alleles, or whether the *TetON* or *TetOFF* regulator was used. As a control for purity of germ cells, we performed BS-PCR and sequencing for the *Igf2r* DMR, which is methylated upon maternal transmission, and found extensive hypomethylation, as expected for male germline cells (Wutz *et al*, 1997). We also assayed methylation states of the two alleles in somatic fractions of developing gonads. As with gonocytes, the *tm5.1* allele was unmethylated regardless of mode of inheritance or transactivator. As expected, the wild-type allele was hypermethylated upon paternal transmission, and hypomethylated upon maternal transmission. Hypomethylation of the maternal DMR in somatic fractions demonstrated that pitRNA expression from the *tm5.1* allele does not act in *trans*.

We further assayed *tm5.1* and wild-type DMR methylation in mature sperm of *tm5.1* heterozygotes, where the pitRNA was regulated by *TetOFF* in

the absence of doxycycline. As with gonocytes, the *tm5.1* DMR was consistently hypomethylated, whereas the wild-type DMR was methylated (Fig 2e and Fig S5c). We concluded that in the embryonic and mature male germlines, pitRNA expression alone was not sufficient to impart methylation at the *tm5.1* DMR in *cis*, indicating that the repeats perform an additional necessary function for DMR methylation beyond controlling pitRNA expression.

**Figure 2. Induction of pitRNA in the male germline does not impart methylation *in cis*, at the *tm5.1* DMR, or in *trans*, at the *Rasgrf1* DMR**. DNAs were collected from gonocyte and somatic fractions of male embryos (a-d) and mature sperm from adult males (e). The somatic and germline fractions of one male is shown in each panel a-d; panel e shows bisulfite analysis of the sperm of two males (#1 and #2). Animals were on the C57BL/6 or FVB/n (FVB, c, d) backgrounds. Mothers were fed chow containing (+ dox), or lacking (– dox) doxycycline between mating and birth. Bisulfite PCR was done using PCR primers specific to the *tm5.1* DMR or, as controls, the DMRs from wild type (WT) *Rasgrf1* and *Igf2r*. WT and *Igf2r* are respectively expected to be hyper- and hypomethylated in male gonocytes; *Igf2r* is expected to be 50% methylated in soma. Number of CpG dinucleotides assayed from *tm5.1*, WT, *tm5.1*and *Igf2r* totaled 15, 17, and 12 respectively. MiSeq libraries prepared from PCR products were sequenced to a minimum of 12 reads per sample for the two *Rasgrf1* alleles (range: 12-28,067; median: 135). Data are reported as the percentage of reads showing the methylation levels indicated in the Bisulfite Key at lower right. *Igf2r* hypomethylation in the gonocyte fractions of each sample indicate minimal somatic contamination. Consistent hypomethylation of the *tm5.1* allele in each pedigree and sample indicates pitRNA expression in the absence of the repeats is insufficient for DMR methylation.

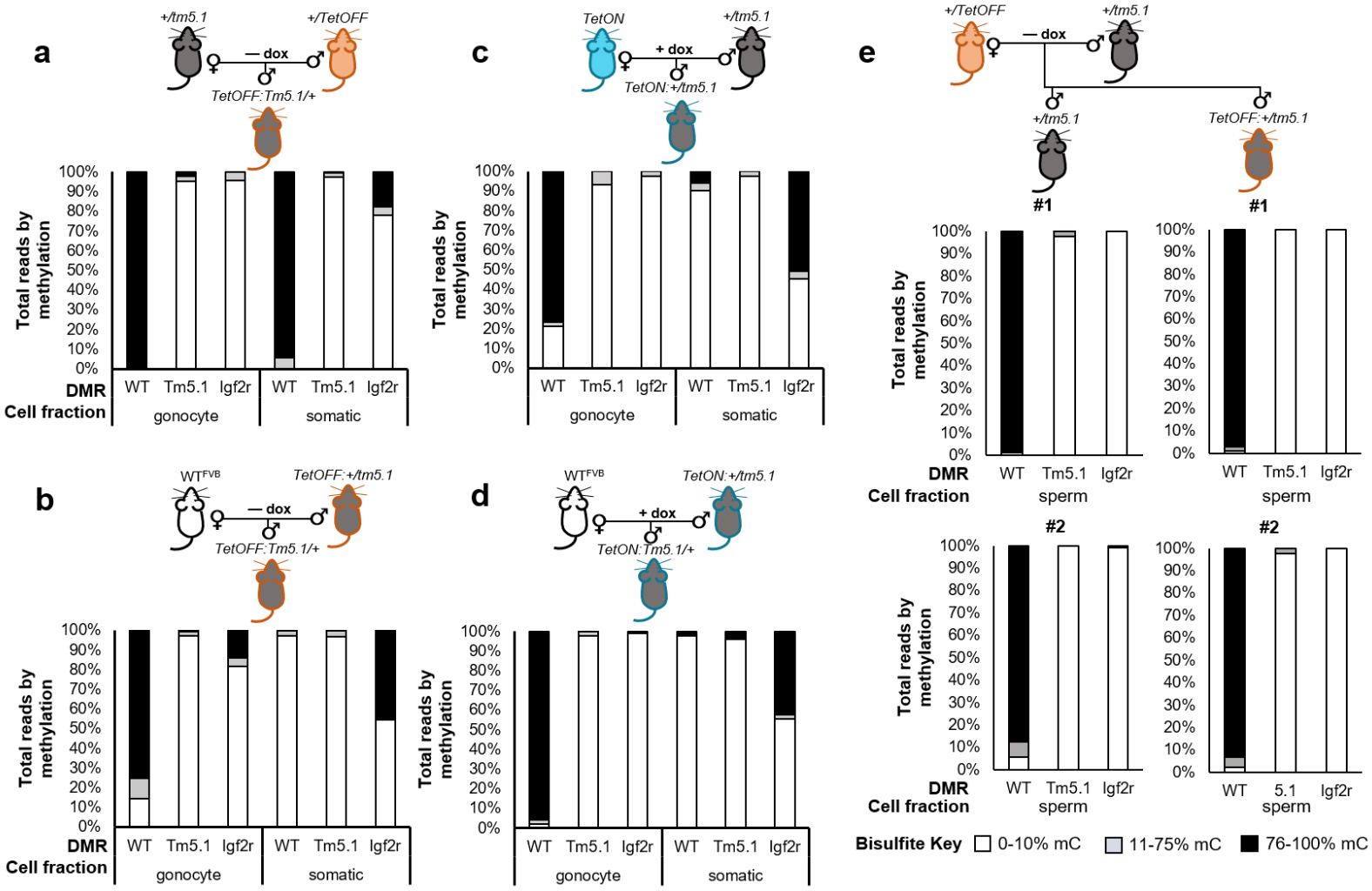
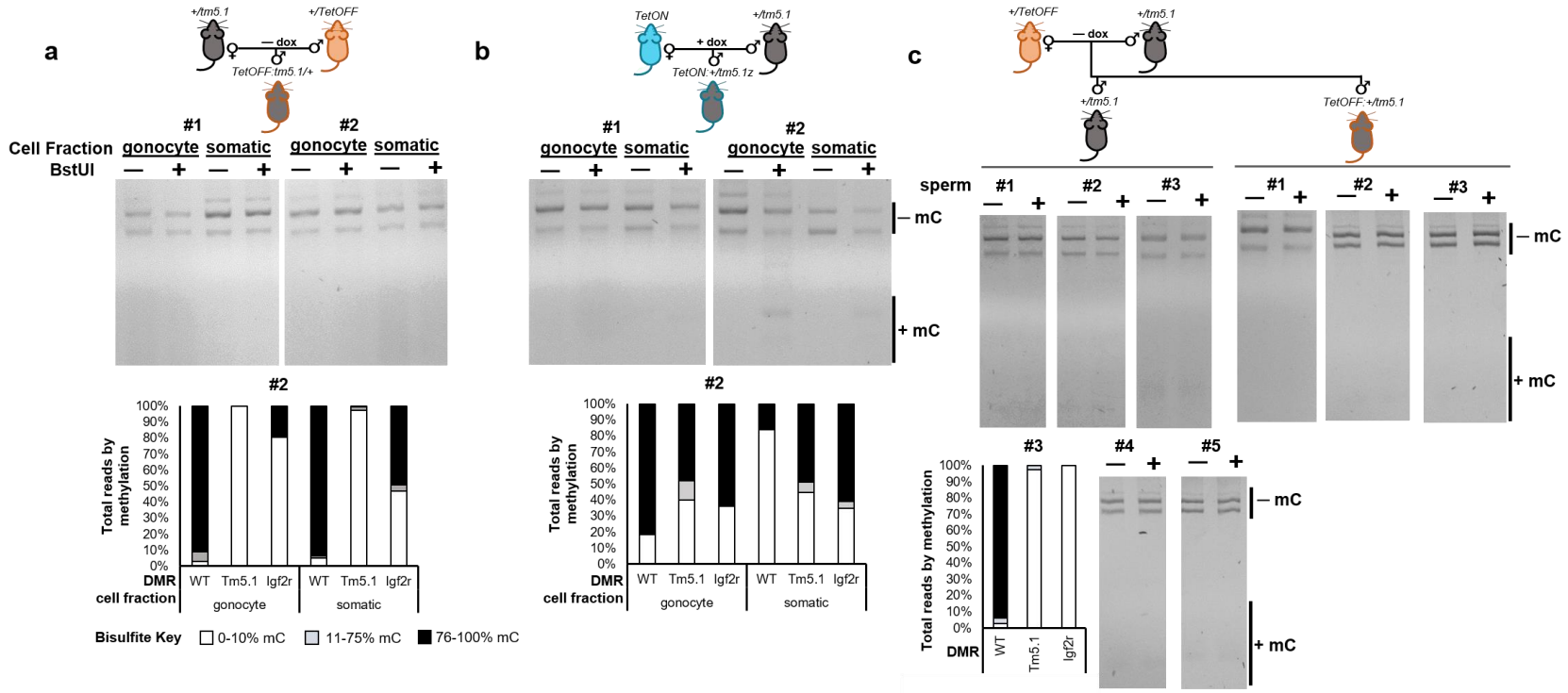

a
+/tm5.1
+/TetOFF
— dox
TetOFF:Tm5.1/+
Total reads by methylation
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
DMR
WT Tm5.1 Igf2r WT Tm5.1 Igf2r
Cell fraction
gonocyte
somatic
b
WT^FVB
TetOFF:+/tm5.1
— dox
TetOFF:Tm5.1/+
c
TetON
+/tm5.1
+ dox
TetON:+/tm5.1
d
WT^FVB
TetON:+/tm5.1
+ dox
TetON:Tm5.1/+
e
+/TetOFF
+/tm5.1
— dox
+/tm5.1
TetOFF:+/tm5.1
#1
#2
WT Tm5.1 Igf2r
WT 5.1 Igf2r
sperm
Bisulfite Key
0-10% mC
11-75% mC
76-100% mC

**Figure S5. Additional bisulfite analysis of embryonic and adult male germline. a)** COBRA of the *tm5.1* DMR for the *TetOFF/tm5.1* animal depicted in **Fig 2a** (#1) as well as a littermate of the same genotype (#2). "+" and "—" connote addition of BstUI. The *tm5.1* DMR is hypomethylated in the gonocyte and somatic fractions of both animals. Targeted bisulfite sequencing for #2 is shown below. The paternally inherited wild-type (WT) DMR is hypermethylated in both gonocyte and somatic fractions; the *tm5.1* DMR is hypomethylated in both gonocyte and somatic fractions. The Igf2r DMR is hypomethylated in gonocyte and 50% methylated in soma. **b)** COBRA for the *tm5.1* DMR for the *TetON/tm5.1* + dox animal depicted in **Fig 2c** (#1) as well as a littermate of the same genotype (#2). Bisulfite sequencing results for #2 are depicted below. Note that while the *tm5.1* DMR is hypomethylated in the gonocyte and somatic fractions of #1, it is approximately 50% methylated in #2 by bisulfite sequencing and COBRA. This is consistent with rates of stochastic *tm5.1* DMR methylation as previously described: As described in **Fig S2**, 10-25% of *+/tm5.1* show evidence of *tm5.1* methylation and expression in the soma. Rep #2 is one of six assayed (18%) transactivated *tm5.1* embryonic gonads to display partial methylation of the *tm5.1* DMR in the gonocyte and somatic fraction. **c)** COBRA for the two *+/tm5.1* and two *TetOFF/tm5.1* animals shown in **Fig 2e** (#1 and #2 of each genotype). COBRA results for an additional three *+/tm5.1* animals (#3, #4, #5) are shown; bisulfite results for #3 are below. COBRA from an additional *TetOFF/tm5.1* animal (#3) is also shown. In all samples, the *tm5.1* DMR is hypomethylated. **Bisulfite** Key at lower right.

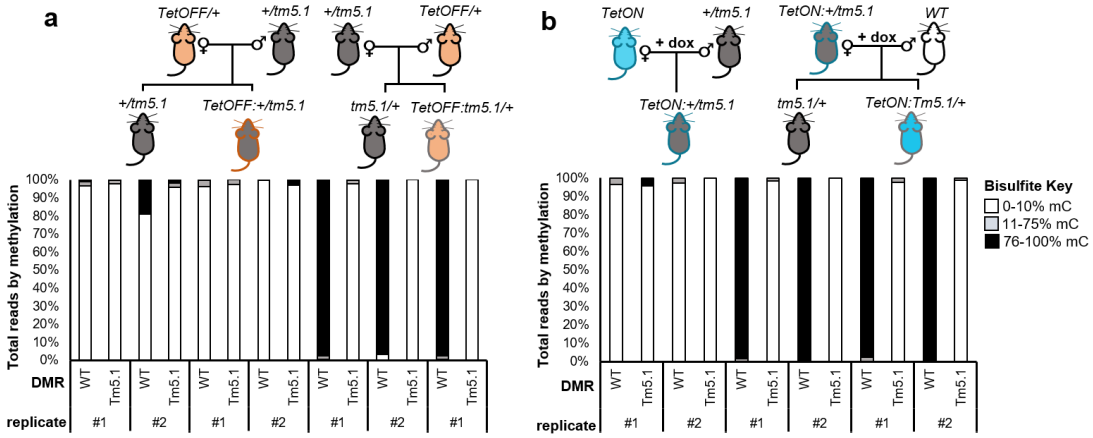

a
+/tm5.1
+/TetOFF
— dox
TetOFF:tm5.1/+
#1
#2
Cell Fraction
gonocyte
somatic
gonocyte
somatic
BstUI
— + — + — + — +
#2
Total reads by methylation
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
DMR
cell fraction
WT Tm5.1 Igf2r WT Tm5.1 Igf2r
gonocyte
somatic
Bisulfite Key
0-10% mC
11-75% mC
76-100% mC
b
TetON
+/tm5.1
+ dox
TetON:+/tm5.1z
#1
#2
gonocyte
somatic
gonocyte
somatic
— + — + — + — +
— mC
+ mC
#2
Total reads by methylation
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
DMR
cell fraction
WT Tm5.1 Igf2r WT Tm5.1 Igf2r
gonocyte
somatic
c
+/TetOFF
+/tm5.1
— dox
+/tm5.1
TetOFF:+/tm5.1
sperm
#1
#2
#3
#1
#2
#3
— + — + — + — + — + — +
— mC
+ mC
#3
#4
#5
— + — + — +
Total reads by methylation
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
DMR
WT
Tm5.1
Igf2r
— mC
+ mC

### II.D.5 pitRNA induction in the male germline is insufficient for somatic methylation at *Rasgrf1*.

In previous studies, exporting the *Rasgrf1* ICR to the *Wnt1* locus led to hypomethylation of the mutant allele in sperm, but hypermethylation in somatic tissue after fertilization (Taylor *et al*, 2016). Though sperm methylation of the modified *Wnt1* allele was higher than sperm methylation at the *tm5.1* allele, we determined if expression of pitRNA by TetO induction of the *tm5.1* allele could enable methylation in somatic tail DNA of progeny after paternal transmission. In all tail samples tested, the *tm5.1* allele remained unmethylated regardless of which transactivator was used to control pitRNA (Fig 3a-b). These findings demonstrated that like methylation in the male germline, methylation in somatic tissue of offspring after paternal transmission is not enabled by pitRNA expression alone. Instead, and consistent with findings from the *Wnt1* mutant allele, additional features of the repeats, beyond their control of pitRNA expression, are necessary for methylation.

**Figure 3.** *TetOFF*-mediated transactivation does not affect *tm5.1* DMR methylation in neonatal tail. **a**) Targeted bisulfite analysis of WT and *tm5.1* DMRs in neonatal tail gDNA; pedigree shown at top. *tm5.1* is hypomethylated regardless of parental descent and presence of *TetOFF*. The WT DMR is methylated depending on parent of origin—if inherited maternally, the wild type DMR is hypomethylated; if inherited paternally, the wild type DMR is hypermethylated. Two biological replicates for all genotypes are shown (#1 and #2) except for the *TetOFF:tm5.1/+* genotype, where one animal is shown. **b)** Targeted bisulfite analysis of 5.1/+, *tm5.1*/*TetON*, and *TetON*/*tm5.1* neonatal tail gDNA. In all genotypes, *tm5.1* DMR is hypomethylated; as expected, the wild-type DMR is methylated if inherited paternally as in *tm5.1* and *tm5.1*/*TetON* animals and hypomethylated if inherited maternally as in

*TetON*/*tm5.1* animals. Reads for two animals (#1 and #2) per genotype are shown. **, $p < 10^{-3}$; ***, $p < 10^{-6}$; n.s., not significant.




## II.D.6 *tm5.1* transactivation and pitRNA induction leads to widespread activation of neighboring genes.

We expanded our initial analysis of methylation by characterizing expression states of *Rasgrf1* and nearby loci in mice carrying the *tm5.1* allele. *Rasgrf1* expression in neonatal brain requires either methylation of the DMR, which is a methylation-sensitive enhancer blocker, or ectopic insertion of an enhancer proximal to the *Rasgrf1* promoter (Yoon *et al*, 2005). We found that *TetOFF*-mediated *tm5.1* induction led to a tenfold upregulation of *Rasgrf1* in neonatal brain relative to wild-type regardless of the parental origin of the *tm5.1* allele (Fig 4a-c). Sequencing of *Rasgrf1* RT-PCR products revealed that while *TetOFF*:+/*tm5.1* animals that inherit *tm5.1* paternally, expressed *Rasgrf1* solely from the paternal *tm5.1* allele, *TetOFF: tm5.1*/+ animals that inherit *tm5.1* maternally, expressed *Rasgrf1* from both the maternal and paternal alleles at a ratio of approximately 9:1, supporting a tenfold upregulation of

*Rasgrf1* from the transactivated *tm5.1* allele but continued *Rasgrf1* expression from the paternal wild-type allele (Fig 4c, Fig S6a). We observed similar effects with *TetON* transactivator. Though the magnitude was lower, *TetON* also activated *Rasgrf1* expression from the maternal *tm5.1* allele (Fig 4 d-f).

**Figure 4. Transactivation of *tm5.1* upregulates *Rasgrf1* expression from *tm5.1* in neonatal brain**. **a)** Pedigrees shown for **b)** and **c)**. **b)** *Rasgrf1* expression is upregulated ten-fold in the brains of both *TetOFF:+/tm5.1* and *TetOFF:tm5.1/+* neonates. *Rasgrf1* is significantly downregulated in *+/tm5.1* animals, who inherit tm5.1 paternally, but unaffected in *tm5.1/+* animals, who inherit a WT allele paternally. *TetOFF:+/+* animals express *Rasgrf1* at WT levels. Error bars represent standard error across biological triplicate at minimum. *Rasgrf1* expression levels are normalized to Rpl32. **c)** Parental allele expression in neonatal brains of animals carrying *tm5.1, TetOFF*, or WT alleles. Allelic expression key at far right. *tm5.1* was generated on a 129 background and carries 129 SNPs; *TetOFF* is carried on a C57Bl/6 background and carries C57Bl/6 SNPs. Bars represent C57Bl/6 vs. 129 reads from MiSeq libraries prepared from RT-PCR product that spans two polymorphisms within *Rasgrf1* coding sequence. Products were sequenced to a minimum of 15 reads per animal and are depicted here as a percentage of total reads averaged across biological duplicates at minimum. *Rasgrf1* expression in *TetOFF:+/tm5.1* animals, who inherit tm5.1 paternally, is 100% paternal (129) due to zero contribution from the normally silent maternal WT allele. In *TetOFF:tm5.1/+* animals, who inherit *tm5.1* maternally, *Rasgrf1* expression is largely maternal (129) due to massive transactivation of *tm5.1*. However, 10% of *Rasgrf1* expression sequence contain C57Bl/6 SNPs, indicating that the paternal WT allele continues to be expressed. **d)** Pedigree for **e)** and **f). e)** *TetON*/*tm5.1* animals express *Rasgrf1* at roughly 75% of wild-type; *tm5.1*/*TetON* animals express insignificantly different levels of *Rasgrf1* relative to wild-type. +/*tm5.1* animals express severely reduced levels of *Rasgrf1*. Animals inheriting *TetON* without *tm5.1* express *Rasgrf1* at wild-type levelsError bars represent standard error across biological triplicate at minimum. *Rasgrf1* expression levels are normalized to Rpl32. **f)** Parental allelic expression in *TetON*/*tm5.1*, *tm5.1*/*TetON*, and *tm5.1*/+ animals. *Rasgrf1* allelic expression in transactivated and non-transactivated *tm5.1* animals. Percentages reflect averages of at least biological duplicates. *TetON*/*tm5.1* animals, who inherit *tm5.1* paternally, express solely from the *tm5.1* (129) allele. *tm5.1*/*TetON* animals express biallelically, reflecting activation of both the maternal *tm5.1* (129) and paternal WT (C57Bl/6) alleles. Percentages reflect averages of at least biological duplicates. *tm5.1*/+ animals show 100% paternal expression from the WT (C57Bl/6) allele. *, $p < 0.05$; **, $p < 0.01$; ***, $p < 10^{-6}$.

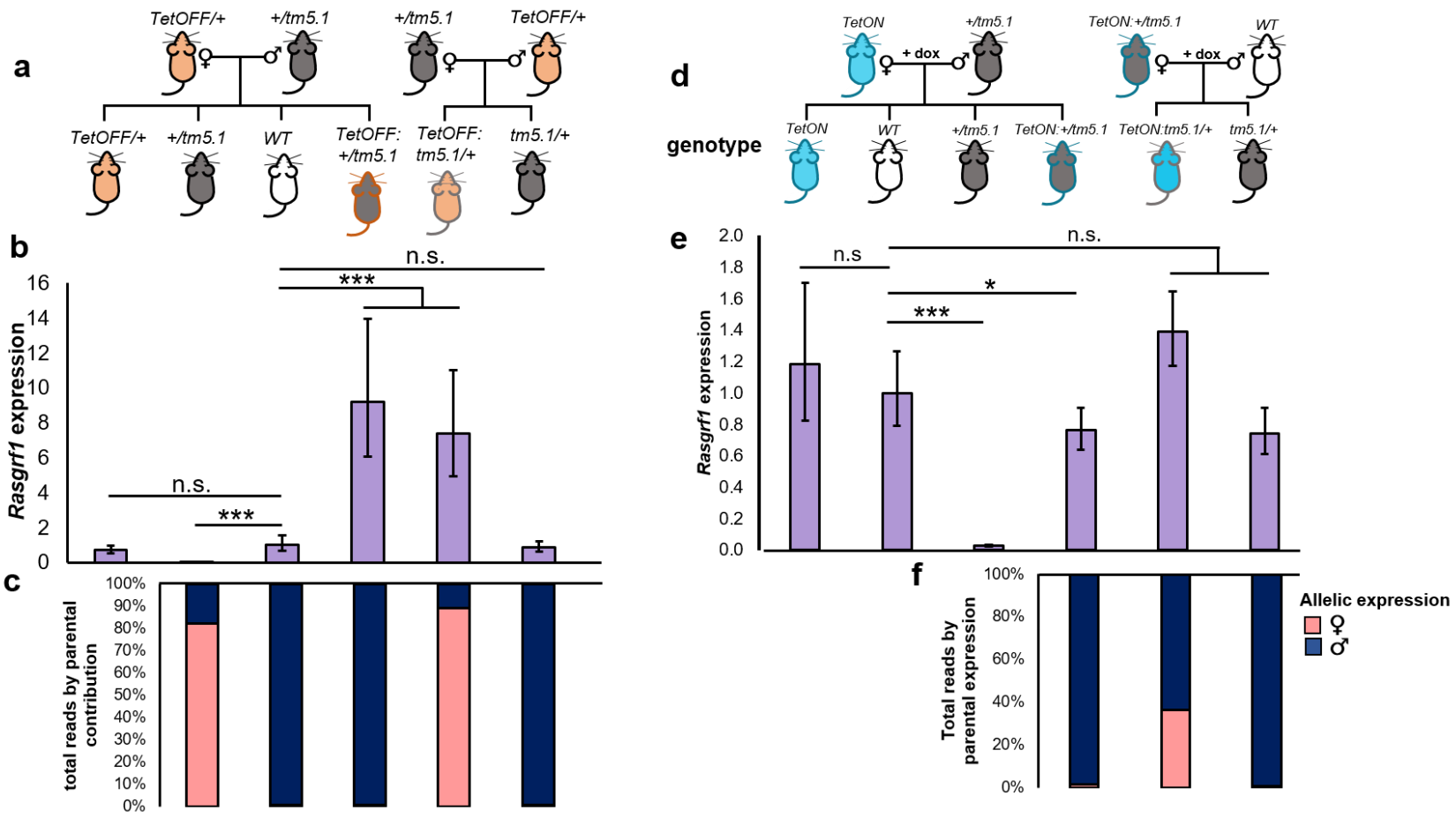
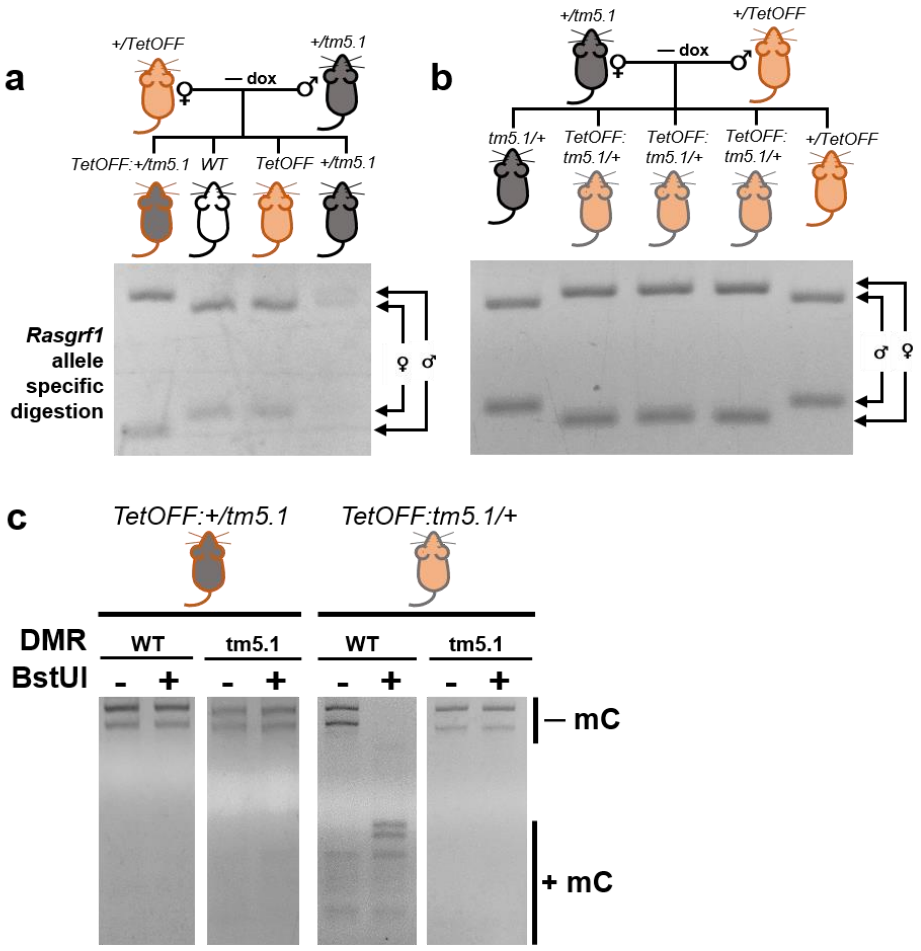

a
TetOFF/+
+/tm5.1
+/tm5.1
TetOFF/+
TetOFF/+
+/tm5.1
WT
TetOFF: +/tm5.1
TetOFF: tm5.1/+
tm5.1/+
b
Rasgrf1 expression
n.s.
***
n.s.
***
c
total reads by parental contribution
d
genotype
TetON
+/tm5.1
+ dox
TetON:+/tm5.1
WT
+ dox
TetON
WT
+/tm5.1
TetON:+/tm5.1
TetON:tm5.1/+
tm5.1/+
e
Rasgrf1 expression
n.s
n.s.
*
***
f
Total reads by parental expression
Allelic expression
♀
♂

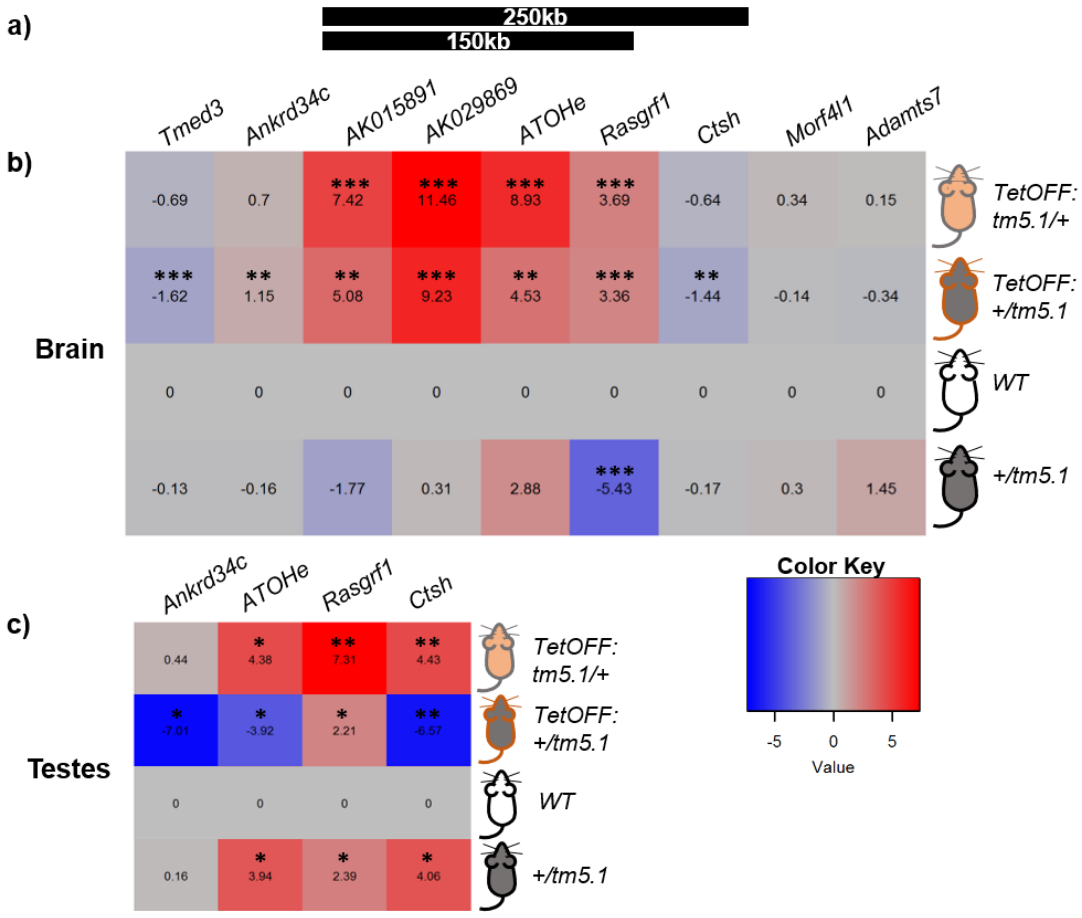
**Figure S6. Transactivation of *tm5.1* with *TetOFF* induces *Rasgrf1* expression from the *tm5.1* allele, but does not impart *tm5.1* DMR methylation or affect WT DMR methylation. a)** *Aci*I digestion of PDS 245-6 endpoint RT-PCR product in using neonatal brain RNA from progeny of *TetOFF* females x *tm5.1* males, and *tm5.1* females x *TetOFF* males. Pedigree at left: *TetOFF*: +/*tm5.1* animals express from the paternal *tm5.1* allele, whereas WT and *TetOFF*: *tm5.1*/+ animals express from the paternal WT (C57Bl/6) allele and +/*tm5.1* animals express biallelically. Pedigree at right: *AciI* digestion suggests expression solely from the maternal *tm5.1* (129) allele in *TetOFF*:*tm5.1*/+ animals, though sequencing as shown in Fig 3b define predominantly paternal expression. *tm5.1*/+ animals express normally from the paternal WT (C57Bl/6) allele. **b)** COBRA for the *tm5.1* and WT DMRs in tail DNA of *TetOFF*: +/*tm5.1* and *TetOFF*: *tm5.1*/+ animals. The WT DMR is unmethylated if inherited maternally (animal at left) but fully methylated, as evidenced by complete digestion of PCR products, if inherited paternally (animal at left). The *tm5.1* DMR is hypomethylated regardless of parental mode of inheritance.



The *Rasgrf1* locus lies within two annotated regions of chromatin interaction as shown by cohesin ChIA-PET (Dowen *et al*, 2014) (Fig 5a). To define the extent of *TetOFF*-mediated transactivation and its relationship to the bounds of known regions of interaction, we queried the effects of *TetOFF* on other transcripts within the interacting regions. The noncoding transcripts immediately adjacent to the *tm5.1* DMR and the Tet Operator, *AK029869*, *A19*, and an annotated *ATOH* binding site (*ATOHe*) were upregulated in both *TetOFF*:+/*tm5.1* and *TetOFF*:*tm5.1*/+ animals. Interestingly, while *TetOFF*: +/*tm5.1* animals displayed insignificant transcriptional changes in more distant genes *Ankrd34c* and *Ctsh* in neonatal brain, *TetOFF*: *tm5.1*/+ animals displayed significant downregulation of these transcripts (Fig 5b). This downregulation was even more dramatic in neonatal testes (Fig 5c). These findings reveal that the effects of the *TetOFF* transactivator perturbs regional transcription in a manner dependent on chromatin boundaries.

**Figure 5. Regional transcription is perturbed in *TetOFF*: +/5.1 and *TetOFF*: 5.1/+ brain and testes. a)** Two regions of chromatin interactions (black bars) are annotated at the *Rasgrf1* locus as predicted by cathepsin ChiA-PET (Dowen *et al*, 2014); locations are shown relative to the genes indicated below in **b**. Lengths of each predicted interaction are shown in kilobases (kb) within each bar. **b)** Log10 expression of nine transcripts from chr9:89,60,000-90,100,000 are shown; these are displayed left to right as they are located 5' to 3'. In brain, *TetOFF* exerts dramatic upregulation of *Rasgrf1* and nearby transcripts *AK015891*, *AK029869*, and an annotated ATOH1 binding site 3' of the *Rasgrf1* repeats (*ATOHe*) in mice with both paternally and maternally inherited *tm5.1*. However, distant transcripts *Tmed3* and *Ctsh* are downregulated modestly in brains of offspring inheriting *tm5.1* maternally. **c)** Log10 expression of four transcripts (a subset of the nine assayed in brain) in neonatal testes. While *ATOHe* and *Ctsh* are upregulated when *tm5.1* is transactivated and paternally inherited, *Ankrd34c*, *ATOHe*, and *Ctsh* are downregulated when *tm5.1* is transactivated and inherited maternally. All expression data are relative to Rpl32. **Color Key** at lower right. *, $p < 0.05$; **,

$p < 0.01$; ***, $p < 10^{-6}$.




### II.D.7 pitRNA loading of oocytes does not produce paramutation

Previously, our lab described a paramutation-like phenomenon at *Rasgrf1* associated with the *Rasgrf1*$^{tm3.1PDS}$ allele (*Tm3.1*), in which the repeats were replaced by the imprinting control region (ICR) of *Igf2r* (Herman *et al*, 2003). Progeny carrying a paternal *Tm3.1* allele exhibited a derepression of the maternal allele, as interpreted from endpoint RT-PCR; and further, some wild-type offspring of +/*Tm3.1* females demonstrated continued depression of the maternal allele, though they lacked the original paternal allele that incited the derepression. This intergenerational effect is a key feature of

paramutation. The *Igf2r* sequences in the *Tm3.1* allele harbors the promoter for *Air*, a non-coding RNA that regulates imprinted expression at the locus (Sleutels *et al*, 2002). Its presence and orientation in the *Tm3.1* allele could impart novel expression patterns of the pitRNA; accordingly, we hypothesized that the intergenerational, paramutation-like effects could involve oocyte loading of pitRNA. To test this hypothesis, we treated females carrying the *tm5.1* allele, and the TetON activator, with intraperitoneal dox for three days, which led to oocyte pitRNA levels approximately 90-fold higher than in wild-type mice (Fig S8a). Females subjected to these treatments were bred to wild-type males, and maintained on dox chow for the duration of pregnancy—as such, their *TetON:*+/*tm5.1* offspring were informative for somatic effects of pitRNA induction in oocytes as previously described (Fig 5, Fig S7).

Additionally, their wild-type, *TetON:*+/+, and +/*tm5.1* offspring were informative for determining if pitRNA loading in oocytes could induce paramutation. In none of these offspring born to mothers with pitRNA preloaded in their oocytes did we observe effects on imprinting status or expression levels of *Rasgrf1* (FigS8c, d). COBRA analysis of the maternal *tm5.1* allele transmitted by oocytes preloaded with pitRNA showed that it remained hypomethylated (Fig S8e). We concluded that preloading oocytes to nearly 100-fold levels of pitRNA relative to wild-type was insufficient to induce transgenerational effects.

**Figure S7. Transactivation of *tm5.1* with *TetON* and doxycycline induces expression of *Rasgrf1* from the *tm5.1* allele, but does not impart methylation to the *tm5.1* DMR or affect expression of methylation of the**

**WT allele. a)** *AciI* digestion of PDS 245-6 endpoint RT-PCR product in the offspring of a *TetON* x *tm5.1* cross. *TetON*: +/*tm5.1* animal express from the paternal *tm5.1* (129) allele, whereas WT and *TetON*:+/+ animals express from the paternal WT (C57BL/6) allele and +/*tm5.1* animals express biallelically. **b)** *Aci*I digestion of PDS 245-6 endpoint RT-PCR product in the offspring of a *tm5.1* x *TetON* cross. *TetON*: *tm5.1*/+ animals express *Rasgrf1* biallelically, indicating activation of the maternal *tm5.1* allele in addition to the normally active paternal WT allele. *tm5.1*/+, *TetON*: +/+ and WT animals express paternally from the WT allele. **c)** COBRA for the *tm5.1* and WT DMRs in tail DNA of *TetOFF*: +/*tm5.1* and *TetOFF*: *tm5.1*/+ animals. The *tm5.1* DMR is hypomethylated regardless of parental origin. The WT DMR is unmethylated if inherited maternally (animal at left) but fully methylated, as evidenced by complete digestion of PCR products, if inherited paternally (animal at right).

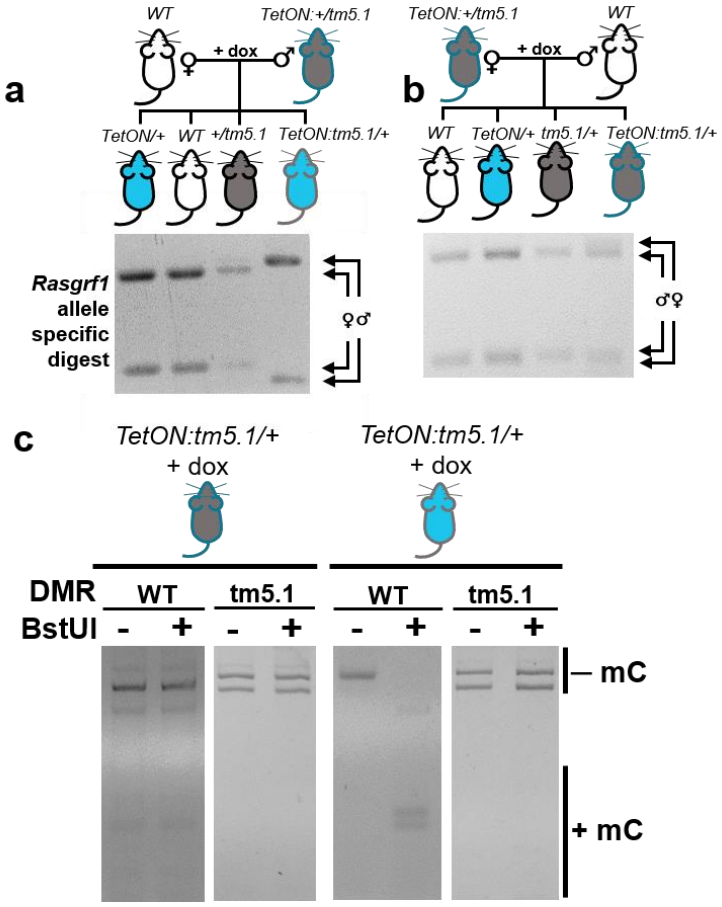
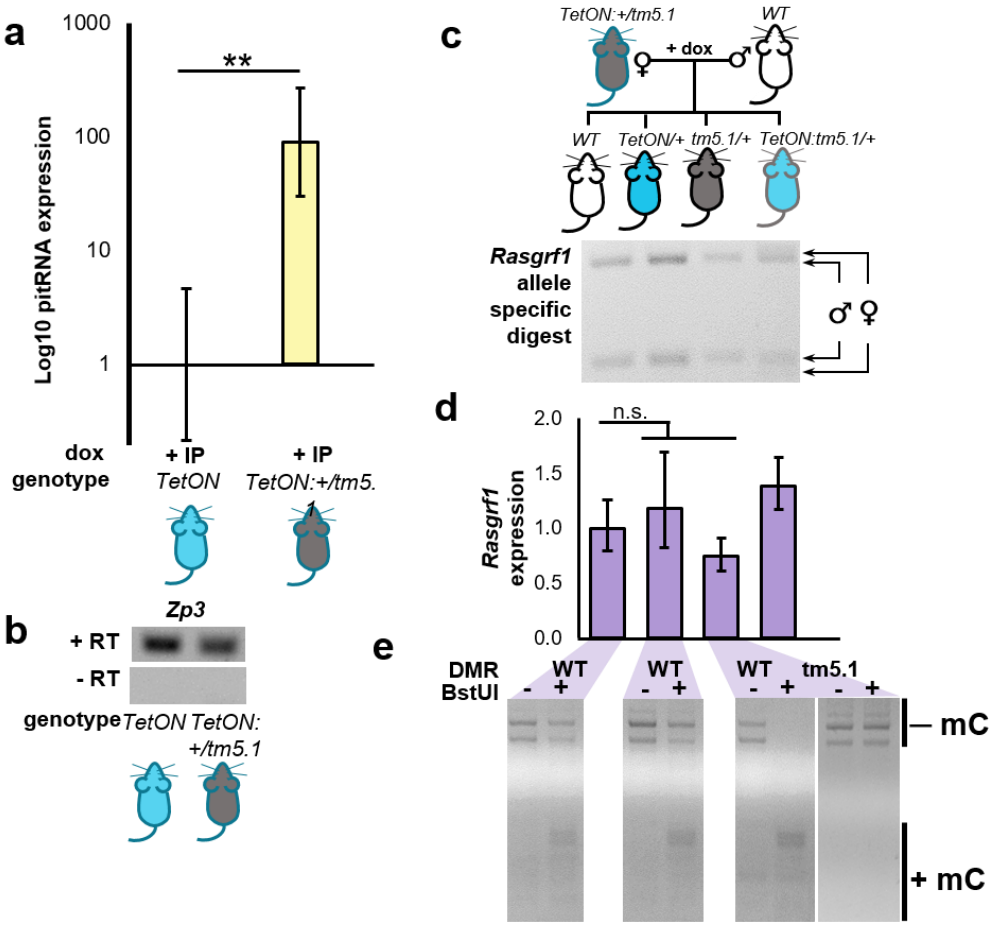



**Figure S8. Oocyte preloading of pitRNA to 90X wild type levels has no effect on *Rasgrf1* expression in wild-type offspring. a)** pitRNA levels induced by *TetON*/*tm5.1* and *TetOFF*/*tm5.1* oocytes. Error bars represent standard error across biological duplicates with the exception of the *TetOFF*/Tg condition (technical replicates of a single animal). **b)** RT-PCR for *Zp3*, an oocyte-specific transcript, was used to confirm isolation of oocytes. **c)**

The pedigree shown was used to determine if oocyte loading of pitRNA could produce intergenerational activation of the maternal *Rasgrf1* allele. pitRNA was induced in oocytes of a *TetON: +/tm5.1* female by IP injection of doxycycline for three days prior to mating; doxycycline-containing chow was provided throughout pregnancy. Activation of the maternal allele in neonatal brains of offspring was used to report intergenerational effects, with allele-specific expression assayed by *Aci*I digestion of RT-PCR products. *tm5.1tm5.1*All mice tested expressed only the paternal allele. **d)** *Rasgrf1* levels in brains of animals depicted directly above in c). Error bars represent standard error across biological duplicates at minimum. **e)** COBRA of the wild-type allele in WT, *TetON*/+, and *tm5.1*/+ tails from three representative animals from d) (purple trapezoids). The WT DMR is partially methylated in WT and *TetON: +/+* animals (having inherited two WT DMRs) and fully methylated in the *tm5.1*/+ animal (having inherited the WT DMR paternally); the *tm5.1* DMR is hypomethylated in this animal. Pale horizontal bar in the top third of each panel represents the dye front. ** p < 0.01; n.s., not significant; IP, intraperitoneal.




## II.E Discussion

### II.E.1 The pitRNA is insufficient to impart establishment of DNA methylation at *Rasgrf1*.

Many investigators have sought to identify *cis*-elements that direct DNA methylation with some notable successes (Chotalia *et al*, 2008; Frohlich *et al*, 2010; Kantor *et al*, 2004; Pant *et al*, 2003; Pant *et al*, 2004; Yoon *et al*, 2002). We previously characterized a *cis*-element at the *Rasgrf1* locus, a series of tandem repeats 30 kb upstream of the *Rasgrf1* coding sequence that is required for imprinted methylation and expression of *Rasgrf1*. More recently, we identified a lncRNA, the pitRNA, as a potential link between the *Rasgrf1* repeats and the establishment of methylation at the *Rasgrf1* DMR. We have demonstrated that pitRNA transcription is contingent upon the *Rasgrf1* repeats, that the pitRNA is targeted by the piRNA pathway, and that components of the piRNA pathway are necessary for full methylation on the paternal allele. However, it was not known if the pitRNA transcription from the *Rasgrf1* DMR was sufficient for its methylation in *cis*, or if the *Rasgrf1* repeats have any pitRNA-independent functions important for methylation.

In general, the effect of lncRNAs versus their *cis*-elements have proven difficult to uncouple: Indeed, reviews discuss the utility of full gene knockout vs. premature termination of the lncRNA transcript vs. lncRNA knockdown (Li *et al*, 2014; Goff *et al*, 2015) in parsing these two distinct elements. In the past year, many have achieved separation of the lncRNA and its *cis*-element through truncation of the lncRNA transcript while keeping the *cis* element intact (Yin *et al*, 2015; Paralkar *et al*, 2016; Amândio *et al*, 2016; Engreitz *et al*,

2016). These investigators have, with few exceptions, reported a largely insignificant effect on target gene expression and/or phenotype upon loss of the lncRNA transcript. We took a converse approach, where we removed the *cis* element and retained the pitRNA sequence, driving it with the inducible Tet Operator.

Using the *TetON* and *TetOFF* transactivators, we achieved induction of pitRNA in the male germline at physiological levels. We found that the *tm5.1* DMR was almost always hypomethylated in the gonocytes of transactivated male embryos. Of nine male embryo gonad samples analyzed, one did display partial gonocyte and somatic methylation of *tm5.1*. This frequency (roughly 11%) is consistent with the frequency of DMR methylation observed in systems lacking the repeats (*tm5.1* as discussed here, and *tm1*). pitRNA expression from the *TetO* also failed to enable methylation in mature sperm, or in somatic DNA of progeny inheriting the *tm5.1* allele from their fathers. We therefore conclude that the pitRNA is insufficient to impart DNA methylation in *cis* and that methylation requires critical functions in the repeats, separate from pitRNA regulation (Fig 6a, c, e, g).

We also examined the effects of the *tm5.1* allele transactivation on methylation state of the somatic cell lineage. We used both *TetON* and *TetOFF* in conjunction with *tm5.1* and found that *tm5.1* remained hypomethylated in neonatal tail, consistent with an intrinsic role for the repeats in the maintenance of somatic DMR methylation through the epiblast stage (Fig 6b, d, f, h).

Finally, we utilized the *tm5.1* allele to query whether previously observed intergenerational effects could be due to oocyte preloading of the pitRNA. Others have reported that gametic loading of RNAs could impart such effects to the next generation (Rassoulzadegan *et al*, 2006; de Vanssay *et al*, 2012**Error! Bookmark not defined.**). We achieved oocyte pitRNA levels up to 90-fold greater than wild-type by intraperitoneal doxycycline administration, however, this produced no effects on imprinted *Rasgrf1* expression or methylation in the wild-type offspring of females subjected to oocyte preloading of pitRNA.

In generating the *tm5.1* allele, we sought to recapitulate the *Rasgrf1* locus with a single difference: The presence or absence of the *cis*-element, retaining all other aspects of local chromatin context and ensuring availability of potential interacting factors. However, this system may not eliminate all confounding variables. The data shown here reflect a constant level of pitRNA expression from the *tm5.1* allele due to 1) constitutive expression and binding of *TetOFF* or 2) steady maternal dox administration in dox chow and therefore binding of *TetON* to TetO: As such, this system may not fully mimic pitRNA dynamics *in vivo*. Discrete temporal windows of pitRNA induction are possible using the *TetON* system, and could be a route for further investigation. However, other data from our lab using a different targeted allele indicate that the *Rasgrf1* repeats can impart methylation at its associated DMR even with minimal pitRNA expression (Taylor *et al*, 2016). Though these studies were performed at the *Wnt1* locus rather than the endogenous *Rasgrf1* locus, a

pitRNA-independent role for the repeats in imparting germline methylation at the DMR is supported in both the *Wnt1*$^{DR}$ and *Rasgrf1*$^{tm5.1PDS}$ systems.

Our data uncover another facet of the mechanism that directs methylation of *Rasgrf1*. A role for the pitRNA and the piRNA pathway for methylation at *Rasgrf1* has been defined. Our lab has previously demonstrated pitRNA is processed into secondary piRNAs, and genetic knockout of known piRNA factors *MitoPLD* and *Mili* lead to extensive, though incomplete, loss of methylation at *Rasgrf1*. The recently discovered male germline-specific Dnmt3 isoform, Dnmt3c, has been strongly implicated as the DNA methyltransferase that establishes DNA methylation at *Rasgrf1* and certain classes of transposable elements (Barau *et al*, 2016). However, from the work presented here, we conclude that the pitRNA is insufficient for targeting and methylation of the *Rasgrf1* DMR. As such, it is possible that two mechanisms, one pitRNA-dependent and one pitRNA-independent, are at play. While at this point our lab has demonstrated that the *Rasgrf1* repeats are necessary, and that the pitRNA alone is insufficient, to impart DMR methylation at *Rasgrf1*, future studies that query the necessity of the pitRNA rather than its sufficiency are merited. As demonstrated by colleagues cited previously, a viable method to query pitRNA necessity would be the insertion of a polyA cassette downstream of the pitRNA transcriptional start site but preceding the DMR, thereby resulting in truncation of the pitRNA while leaving the *Rasgrf1* repeats and DMR intact.

Considering a potential pitRNA-independent role for the repeats, the

fact that a low frequency of paternal methylation and expression persists in *tm5.1* and *tm1* animals also bears discussion. Data from *tm5.1* allele that this is a stochastic—that in 10-20% of animals lacking the repeats, the DMR accrues methylation. In light of this observation, we propose that the repeats act by increasing the probability that the DMR is targeted for methylation, analogous to the activity of the *Xite* element at the X-Inactivation Center. *Xite* is known to mediate the probability with which an X chromosome undergoes X chromosome inactivation, but does not require active transcription to influence X-chromosome choice (Ogawa Y *et al*, 2005). Rather, *Xite* is an enhancer for the lncRNA *Tsix* (Stavropoulos *et al*, 2005) and participates in spatial and transcriptional partitioning of the X inactivation center (Nora *et al*, 2012; Tsai *et al*, 2008).

The mechanism by which the repeats might act in a similar manner to *Xite* remains in question, though our lab's work in combination with many others provide some clues. The repeats are a highly repetitive, GC-rich element, which in and of itself may be sufficient to recruit methylation (Zhang *et al*, 2012; Quilez *et al*, 2016) which may then spread into the DMR, possibly in a manner dependent on pitRNA expression. In addition, the repeats harbor two canonical binding sites for the transcription factor Sp1, which besides its known role in the regulation of gene expression (O'Connor *et al*, 2016), can mediate chromatin structure through the recruitment of chromatin remodeling factors (Suzuki *et al*, 2000; Zhao *et al*, 2003) and mediating enhancer-promoter interactions (Deshane *et al*, 2010; Sumimoto *et al*, 2012).

Sp1 is known to bind the secondary DNA structure G-quadruplexes (Todd *et al*, 2008; Raiber *et al*, 2012), which the *Rasgrf1* repeats are predicted to form, as well as its canonical sequence. Recently, a G4 was characterized at the imprinted *H19* locus. Binding of Sp1, in conjunction with the G4, suppresses *H19* transcription (Fukuhara *et al*, 2017). To our knowledge, G-quadruplex formation at other imprinted loci beyond *H19* has not been investigated. However, differential G4 formation in the maternal and paternal germlines could be a platform upon which the distinct chromatin states observed at the maternal and paternal DMRs are built.

In addition, our lab has previously shown that, at the *Rasgrf1* DMR, H3K27Me and DNA methylation are mutually antagonistic (**Error! Bookmark not defined.**Lindroth *et al*, 2008); H3K27Me deposition on the maternal chromosome is mediated by YYI and PRC2, which has been demonstrated in *Drosophila* (Brown *et al*, 2005) and predicted in *silico* in mammals (Bengani *et al*, 2007) to interact with Sp1 binding sites.

The role of other factors required for parent-of-origin specific DNA methylation must also be considered. The GHKL ATPase Morc1 is a known repressor of transposable elements in the embryonic male germline; *Morc1*-null mice are hypomethylated at the *Rasgrf1* DMR (Pastor *et al*, 2014) and in *Arabidopsis* interacts with the SUVH2 and SUVH9 histone methyltransferases to effect gene silencing (Liu *et al*, 2014). While the KRAB-domain containing zinc-finger binding protein ZFP57 and its cofactor Trim28 (Quenneville *et al*, 2011) are typically thought of as an imprinting maintenance mechanism; loss

of zygotic Trim28 disrupts imprinting at *Rasgrf1* along with many other maternally and paternally imprinted loci (Alexander *et al*, 2015). Interestingly, KRAB-domain ZFP binding can also trigger *de novo* DNA methylation during mouse embryogenesis (Wiznerowicz *et al*, 2007), and in human, Trim28 is demonstrated to have broader roles in the transcriptional repression of transposable elements (Turelli *et al*, 2014). It is quite plausible that these factors could also be recruited either directly via DNA binding capacity or indirectly via histone marks to impart DNA methylation at the *Rasgrf1* DMR.

**II.E.2 Additional considerations for the study of imprinted loci.**

Our experiments also highlight an important technical consideration for the field. End-point PCR followed by allele-specific restriction digest as a measure of parent-of-origin expression must be supported with quantitative data. By end-point RT-PCR followed by allele-specific restriction digest, *TetOFF*-mediated transactivation of maternally inherited *tm5.1* appears to fully reverse imprinting, with maternal-only bands upon gel electrophoresis (Figure S7). Similar data have previously been interpreted as a silencing of the paternal allele *in trans* (Brideau *et al*, 2010). However, sequencing of *Rasgrf1* cDNA of *tm5.1/+:TetOFF* neonatal brain demonstrates that roughly 10% of total *Rasgrf1* reads are paternal in origin, suggesting that the wild-type paternal allele continues to express *Rasgrf1* at wild-type levels, and is not silenced *in trans*.

Similarly, end-point analysis of *+/tm5.1* animals lacking a transactivating allele reveals biallelic expression. While this could be

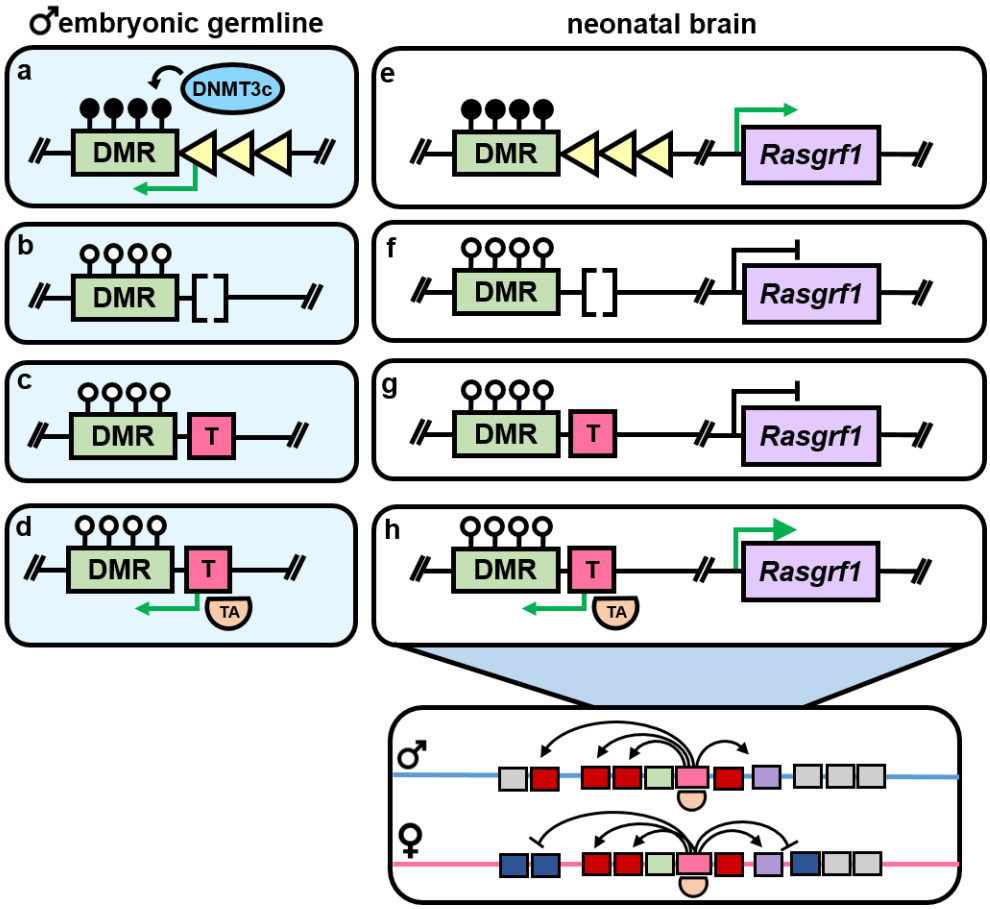
interpreted as a possible *trans* effect of the paternally inherited *tm5.1* allele exerting effects on the maternally inherited and normally transcriptionally silent WT allele, qPCR of *Rasgrf1* in these animals demonstrates severely reduced *Rasgrf1*, suggesting instead that, in the *tm5.1* system, biallelic expression reflects minimal transcription.

### II.E.2 Neighboring gene activation in response to transactivation.

An unexpected outcome of this work includes a local effect of *TetOFF*-mediated transactivation of *tm5.1*. Interestingly, effect boundaries are reflected with chromatin interaction boundaries defined by cohesin ChiA-PET (Dowen *et al*, 2014). Two regions of interaction are defined which, given the diverse effects observed depending on parental inheritance of *tm5.1*, we propose represent the maternal and paternal alleles; maternal and paternal specific-effects of *tm5.1* transactivation are observed in both neonatal brain and testes. While the *TetOFF* protein could act in a number of mechanisms, a strong possibility could involve enhancer competition, where the transactivated Tet Operator could act as an enhancer itself, hijacking transcription factors from neighboring enhancers. Indeed, *TetOFF* in conjunction with the Tet Operator has been used for just that purpose (Rose *et al*, 1997). These widespread transcriptional changes were not observed in *TetON:+/tm5.1* tissues (data not shown), suggesting that these widespread effects may be restricted to the *TetOFF* system. While this finding was not an objective of our study, it emphasizes the considerations that should be taken when interpreting transcriptional effects using potent transactivating systems such as the

*TetOFF* protein (Fig 6, outset).

**FIGURE 6. Working model for regulation of methylation at *Rasgrf1* in the male embryonic germline and neonatal brain.** For neonatal brain, only the paternal allele is depicted. **a)** At the wild type *Rasgrf1* DMR, the *Rasgrf1* repeats (yellow triangles) drive expression of the pitRNA antisense to the DMR (transcription depicted by green arrow) and increase the likelihood that the DMR is hypermethylated (filled lollipops), likely by Dnmt3c, in the male germline. **b)** *Rasgrf1*$^{tm1\text{PDS}}$ (tm1) lacks the *Rasgrf1* repeats (open brackets). pitRNA is not expressed and methylation is not established at the DMR in the germline (open lollipops). **c)** *Rasgrf1*$^{tm5.1\text{PDS}}$ (*tm5.1*) is not methylated in the male germline. **d)** *Rasgrf1*$^{tm5.1\text{PDS}}$ transactivation with *TetOFF* (orange half circle labelled TA) induces pitRNA expression but does not impart methylation to the *tm5.1* DMR. **e)** In neonatal brain, *Rasgrf1* is paternally methylated and expressed. **f, g)** *tm1* and *tm5.1* brains have severely decreased *Rasgrf1* expression. **h)** transactivation of tm5.1 causes strong upregulation of *Rasgrf1* but the *tm5.1* DMR remains hypomethylated. **Outset:** *TetOFF*-mediated transactivation of *tm5.1* affects nearby transcription differentially depending on parental inheritance (blue and pink lines). Upregulated and downregulated genes are depicted by red and blue boxes respectively; unchanged genes are depicted in gray. TetO and DMR are connoted by the pink and green boxes respectively; *TetOFF* binding is indicated by the orange half circle.




In conclusion, our data support the existence of a second, pitRNA-independent mechanism for DNA methylation at *Rasgrf1*. We propose a *cis*-acting mechanism by which the repeats increase the probability of methylation in the paternal germline. Future studies could examine the necessity of the pitRNA as well as the role of the *Rasgrf1* repeats in promoting chromatin conformation and structure.

## II.F Materials and Methods

**Primer sequences** for all analyses are listed in Table S1.

**TetO vector generation:** We modified pYP1, which carries the DMR and *Rasgrf1* repeats, and 4 kb of homologous flanking sequence (Park *et al*,

2012) to carry seven copies of the Tet Responsive Element (collectively termed TetO) in place of the *Rasgrf1* repeats as follows. The *Rasgrf1* repeats were removed *via* restriction digest with *Cla*I and *Mlu*I; sticky ends were blunted with Klenow; the plasmid backbone was gel purified and then ligated closed generating pDHT2. pDHT2 and pPX3, a vector containing the Tet Operator, were digested with *Nhe*I, and linearized pDHT2 was ligated to the TetO sequences, generating pDHT3, which was confirmed by Sanger sequencing. The 3' homologous arm of pDHT3 was then shortened to approximately 1kb by restriction digest with *Bsr*GI and *Sfi*I to generate pETC6, which was linearized with *Pci*I prior to lipofection into ES cells.

**CRISPR/Cas9-mediated generation of TetO$^{tg}$, TetOΔ$^{tg}$, *Rasgrf1*$^{tm5.0PDS}$ and *Rasgrf1*$^{tm5.1PDS}$:** pX330 (Addgene Plasmid # 42230) was modified to carry PDS 2195-6, a complementary primer pair coding for an sgRNA targeting the *Rasgrf1* repeats, following the Zhang lab protocol (Cong *et al*, 2013) to generate pX330-rep5. PDS 2195-6 was designed using the CRISPR Design Tool (http://crispr.mit.edu:8079/). To effect homology directed repair at the *Rasgrf1* locus, v6.5 embryonic stem cells (Eggan *et al*, 2001) were lipofected with linearized pETC6 and pX330-rep5 with Lipofectamine 2000, following the manufacturer's protocol. Cells were allowed to recover overnight, then treated for 10 days with 300 ug/mL G418 (Sigma A1720). Colonies were picked and genotyped with PCR 2359-8 and PDS 2344-2263, which generate a product only from the targeted allele (*Rasgrf1*$^{tm5.0PDS}$) and PDS 2757-8, an internal PCR for TetO, to detect cells harboring a randomly

inserted vector that provided the transgenic (Tg) model. tm5.0 was further confirmed by Southern blot (Yoon *et al*, 2002).

TetO$^{Tg}$ and *Rasgrf1*$^{tm5.0PDS}$ ES cells were microinjected into B6(Cg)$^{Tyrc-2J}$/J blastocysts by the Cornell University Transgenics Core. 22 chimeras were recovered. Germline transmission was confirmed by diagnostic crosses to FVB/N females and PCR with PDS 2757-8. Chimeras were crossed with mice constitutively expressing FlpE recombinase (JAX Strain 003800) to generate *Rasgrf1*$^{tm5.1PDS}$ and TetO**Δ**$^{Tg}$. Recombination was confirmed by Sanger sequencing of PDS 2262-3 PCR products, which spans the neo resistance cassette and frt sites.

**Generation of *TetOFF* mice:** To produce mice that constitutively express the tet transactivator (*TetOFF*), mice carrying a tTA transgene preceded by a floxed neomycin-polyadenylation cassette (pA-*TetOFF*, JAX Strain No 011008) were crossed with mice constitutively expressing *Sox2*-Cre (JAX Strain No 008454). Recombination and subsequent loss of the neomycin resistance cassette was confirmed by endpoint PCR with PDS 2794-5, followed by Sanger sequencing. PDS 2794-5 sequences were supplied by Bruce Morgan (Chi *et al*, 2013).

**Induction of pitRNA:** *TetOFF*: TetO**Δ** and *tm5.1* mice were crossed to *TetOFF* mice. Induction of pitRNA expression was validated in adult and neonatal tissues using PDS 2266-7.

*TetON*: Female mice were injected with 0.01 mg/g body weight of 0.01 mg/mL doxycycline hyclate (Sigma D9891) intraperitoneally every 24 hours for three

days as a preloading phase, then bred. Breeding pairs were fed 200 mg/kg doxycycline chow (BioServ S3888) for the duration of pregnancy. Induction of pitRNA expression was validated in all tissues by endpoint and qRT-PCR using PDS 2266-7 and PDS 2916-7, a primer pair specific for pitRNA produced from the *tm5.1* allele.

**Allele-specific analysis of *Rasgrf1* expression:** Neonatal brains were collected at postnatal day 2; the olfactory bulbs were visually identified under dissection microscope and discarded. Other tissues were collected by gross dissection. Tissues were snap frozen in liquid nitrogen, then submerged in 1 mL Trizol. Tissues were homogenized using a Biospec Mini-Bead Beater-8 using 1 3mm steel bead in XXTuff Microvials (BioSpec XX0TX). Total RNA for all samples was processed via the Trizol protocol (Thermo Fisher Scientific 15596018). RNA was DNAse treated, random primed, and reverse transcribed using Promega RQ1 DNAse and RQ RTase (Promega M6101 and A5003 respectively) following the manufacturer's protocol. RT-PCR was performed with PDS 245-6 (95ºC 2min, 40 cycles of 95ºC 30s, 60ºC 30s, 72ºC 30s, 72ºCC 7min using Promega GoTaq in a volume of 25 uL (Promega M3001). *For allele-specific restriction digest*: 20 uL of PCR product was digested with 2.5U *Aci*I. Digestion products were separated by electrophoresis on a 4% agarose gel. *For allele-specific read quantification*: End-point PCR samples were submitted for sequencing in MiSeq libraries as described below in. Targeted sequencing analysis. Read quantification for *Rasgrf1*-specific expression was performed on trimmed samples with the grep and wc

functions. Grep sequences are listed in Table S4. MGI SNP IDs and flanking sequences are listed in Table S5.

**qRT-PCR and heat map generation:** qRT-PCR was performed in 20 uL reactions using SYBR Green Master Mix (CAT 4367659) on a Biosystems 7500 with annealing temperature 60ºC for forty cycles followed by a dissociation stage. The following primer pairs were used: PDS 2266-7 for general pitRNA expression; PDS 2916-7 for TetO-pitRNA expression; PDS 2877-8 for *Rasgrf1* expression; PDS 72-3 for *Rpl32* expression; PDS 2719-8 for *ATOHe* expression; PDS 3211-2 for *Ctsh* expression; PDS 2178-9 for *Ankrd34c* expression. Heat maps were generated in R (Gentleman *et al*, 2004).

**Gonocyte collection**: Females were checked for plugs, weighed to confirm pregnancy, and sacrificed at gestational day 16.5. Gonocytes from male embryos were collected as previously described (Watanabe *et al*, 2011) with some modifications. Briefly, embryonic testes were collected and incubated in 50 uL 0.25% trypsin for 10 minutes. Samples were then triturated and visually inspected for tissue disaggregation. Incubation and trituration were repeated up to two times until full disaggregation was achieved; any remaining clumps were manually removed. Samples were then transferred to McCoy's basic medium supplemented with FBS and pre-plated for 1.5 hours. Suspended germ cells were harvested, pelleted at 300 x *g* for 8 minutes, then processed for RNA and DNA. We validated the purity of gonocytes by bisulfite sequencing of *Igf2r*, which is expected to be extensively hypomethylated in the

male germline (Wutz *et al*, 1997**Error! Bookmark not defined.**).

**Oocyte collection**: 28 to 42 day old females were superovulated with 5 IU human chorionic gonadotropin (Millipore 367222) followed by pregnant mare serum gonadotropin (Millipore 230734) 48 hours later. Oocytes were collected the next morning via standard methods (Duselis *et al*, 2007) and processed for total RNA via Trizol. Oocyte recovery was evaluated using RT-PCR primers for Zp3 (PDS 2212-3).

**Genomic DNA extraction from tails and cells**: All DNA was collected via overnight incubation at 55ºC in Laird's Lysis Buffer (Laird *et al*, 1991) and 20 ug/mL Proteinase K followed by isopropanol precipitation and resuspension in TE. Scant gDNA samples, such as those from gonocytes, were co-precipitated with 20 ug glycogen (Thermo Fisher R0551) and spun at 20,817 *x g* for 15 minutes.

**DNA extraction from sperm:** The caudal epididymis and vas deferens of adult male mice were harvested and placed in PBS for 1 hr at 37ºC. Large tissue chunks were manually removed and the remaining sperm were pelleted at 400 x g for 12 minutes at 4ºC. Supernatant was discarded; the pellet was resuspended in the remaining supernatant and incubated with 1 mL somatic cell lysis buffer (Goodrich *et al*, 2007) for 1 hr on ice. Lysed samples were spun at 20,817 x g for 3 minutes at 4ºC and the supernatant was discarded. The sperm pellet was resuspended in remaining supernatant; 2 uL was mixed with Trypan Blue and examined microscopically for remaining somatic contamination. The remaining suspension was mixed with 500 uL Buffer RLT

(Qiagen 79216) and 150 mM DTT, then homogenized with 2 2mm steel beads following the protocol of Wu *et al,* 2014. DNA extraction then proceeded as described above.

**Bisulfite conversion, BS-PCR, and COBRA:** Bisulfite conversion was performed using the Zymo Research EZ DNA Methylation-Lightning Kit (Zymo Research D5031). BS-PCR was performed with PDS 271-272 for the wild-type DMR, PDS 272-2627 for the *tm5.1* or Tg DMR, PDS 2934-5 for the *Igf2r* DMR, and 271-287 for the *tm1* DMR using NEB Epimark HotStart *Taq* DNA Polymerase (NEB M0590) following the following cycling parameters: 95ºC for 30s, 40X (95ºC 15s, 55ºC 30s, 68ºC 30s) with the exception of PDS 271-287, where the annealing temperature used was 58ºC. COBRA was performed with primers PDS 271-272 after bisulfite treatment via direction addition of 5U of *Bst*UI and digestion at 60ºC for 1 hour, followed by electrophoresis on a 4% agarose gel. In this assay, digestion products arise from methylated DNA; unmethylated DNA resists digestion.

**Targeted sequencing analysis:** PCR products were pooled, column cleaned with the BioBasic EZ-10 DNA Columns (BioBasic BS427), eluted in 30 uL of Tris-HCl pH 8.0, and quantified via Nanodrop. NEBNext Universal Adaptors were ligated using T4 DNA Ligase in Quick Ligase Buffer in a total volume of 22 uL at 25ºC for 15 minutes. Self-ligated adaptors (adaptor dimer) was excluded with a 0.8X (2.55 uL) Agencourt AMPure XP bead cleanup (Beckman Coulter A63880) followed by 1X PEG-NACl buffer (25% PEG, 2.5 M NaCl in DEPC water) cleanup to exclude self-ligated adaptors. Barcodes were

added via PCR amplification for 20 cycles with Phusion HF (NEB M0530L) using NEBNext Multiplex Oligos for Illumina® (NEB E7335). Adaptor dimer was again excluded via 0.8X AMPure bead cleanup followed by a 1X PEG Buffer-NaCl cleanup using the same beads. Libraries were quantified using Qubit, and sequenced on a MiSeq 2000 Paired End 2 x 250bp at the Cornell University Genomics Core. Samples were evaluated for quality and trimmed to 200bp with Trim Galore! (www.bioinformatics.babraham.ac.uk). Trimmed samples were probed for WT, *tm5.1*, or Igf2r DMR-specific sequences using the grep function; these reads were compiled into separate files and analyzed using QUMA (Kumaki *et al*, 2008). Animals with a minimum of ten reads were included for analysis. Total reads per amplicon per sample are listed in Table S2. Grep sequences and reference sequences for QUMA are listed in Table S3.

**Table S1. Primer sequences for all analyses described.**

| **Primer Name** | **5' – 3'** | **Sequence** | **Purpose (Name)** | **Notes** |
|---|---|---|---|---|
| PDS 2091 | F | caccGCAGCGGCAGCGCTAGCGAC | sgRNA pair targeting the *Rasgrf1* repeats | |
| PDS 2092 | R | aaacGTCGCTAGCGCTGCCGCTGC | | |
| PDS 2778 | F | TCGAGTGAAGACGAAAGGGC | *tm5.1*/TetO∆ internal PCR | |
| PDS 2779 | R | TATCCACTGTCCTCCACCCC | | |
| PDS 2757 | F | TGCTGCTGCCGCTAAAGATA | Tm5.0/TetO internal PCR | |

| | | | | |
|---|---|---|---|---|
| PDS 2758 | R | TGTACGGTGGGAGGCCTATA | | |
| PDS 2262 | F | CGAAAGGGCCTCGTGATACGC | 3' junctional PCR for Tm5.0/1 | Phusion High Fidelity *Taq* (NEB M0530) in a 20 uL volume with High Fidelity Buffer; Ta = 70ºC, te = 30s, 35 cycles. |
| PDS 2263 | R | CAGAGCTTGCCCGCCACATT | | |
| PDS 2344 | F | CGGAGAACCTGCGTGCAATCCA | 3' junctional PCR for Tm5.0; with 2262 | |
| PDS 2358 | R | TCGGTACCCGGGTCGAGGTAGG | 5' junctional PCR for *tm5.1* | Phusion High Fidelity *Taq* in a 20 uL volume with High GC Buffer and 3% DMSO; Ta = 68ºC, te = 30s, 35 cycles |
| PDS 2359 | F | CAGGGTCCTTCCTCAAGGTTTTAGCA | | |
| PDS 3052 | F | GTGGTCGAACAGCTCGATG | PCR for *TetOFF* | |
| PDS 3053 | R | ACCAGCTGGCCTTCCTGT | | |
| PDS 2574 | F | gaaggcctgacgacaaggaa | PCR for *TetON* | |
| PDS 2575 | R | gtcgcgatgtgagaggagag | | |
| PDS 2794 | F | GTCCAGGGTTTCCTTGATGA | Recombined *TetOFF* allele, spans neomycin resistance cassette. | te = 2 min |
| PDS 2795 | R | ggagcgggagaaatggatatg | | |

| | | | | |
|---|---|---|---|---|
| PDS 245 | F | ggctcatgatgaatgccttt | *Rasgrf1* cDNA (spans 129/C57 SNPs) | |
| PDS 246 | R | tacagaagcttggcgttgtg | | |
| PDS 249 | F | acattcgcttcagcaaaacc | *Rasgrf1* gDNA (spans 129/C57 SNPs) | |
| PDS 251 | R | attggtgaagacgcgatagg | | |
| PDS 2266 | F | ATCCGTGGCTACCGCTATTG | qPCR - pitRNA general | |
| PDS 2267 | R | AGTCGTGGTAGTTGTAGCGC | | |
| PDS 2916 | F | CATAGAAGACACCGGGACCG | qPCR - *tm5.1*-specific pitRNA | |
| PDS 2917 | R | CTGCCGCGCTAAAGCTGC | | |
| PDS 72 | F | CATGCACACAAGCCATCTACTCA | qPCR - Rpl32 | |
| PDS 73 | R | TGCTCACAATGTGTCCTCTAAGAAC | | |
| PDS 2877 | F | TGTTTTCCAGCAGCCACAAC | qPCR - *Rasgrf1* | |
| PDS 2878 | R | GATGACGTCCCGATGGCC | | |
| PDS 3179 | F | CTGCAAAGCCAAAGGTAAGGA | qPCR - *Ankrd34c* | |
| PDS 3180 | R | TGGTGGTGACTGTCTATCTTCC | | |
| PDS 3211 | F | ACCGTGAACGCCATAGAAAAG | qPCR - *Ctsh* | |
| PDS 3212 | R | TGAGCAATTCTGAGGCTCTGA | | |
| PDS 3277 | F | TTGCCTGAGTTGTGGATGGG | qPCR – *AK029869* | |
| PDS 3278 | R | CATGGCTACCGGTGACACTT | | |
| PDS 3283 | F | GAGCTGAGACCATTTGCCCT | qPCR - *A19* | |

| | | | | |
|---|---|---|---|---|
| PDS 3284 | R | GGCCTTACCTTCAGAACGCT | | |
| PDS 3275 | F | AAAAGGTCCCCAGCCTTAAA | qPCR - *ATOHe* | |
| PDS 3276 | R | AGCACTGTCCCTGAACATCC | | |
| PDS 3375 | F | GCTATGTGGAAGACCCCAGG | qPCR - *Tmed3* | |
| PDS 3376 | R | GAAGCAAAACCGGTAGACGC | | |
| PDS 3381 | F | tgggaaacggcctttcatca | qPCR - *Adamts7* | |
| PDS 3382 | R | gtcatctaagcacaggcccc | | |
| PDS 3379 | F | CGGTGCTGTGAGGTCAGG | qPCR - *Morf4l1* | |
| PDS 3380 | R | CCCTACTCTCTACCCCACCC | | |
| PDS 3397 | F | TTCACGGAGACGGAGTTACC | *Ctsh* Transcript (Spans 129/C57 SNP) | |
| PDS 3398 | R | CACATCTCGGCGCTCACT | | |
| PDS 3400 | F | TCTTAAGGAAATATGCTCCTCTCC | *Ctsh* Transcript (Spans 129/C57 SNP) | |
| PDS 3401 | R | TGCAGGCATCTGTAATCTGG | | |
| PDS 3402 | F | CAACGTTGGTCTCTGTAGCAA | Ctsh Transcript (Spans 129/C57 SNP) | |
| PDS 3403 | R | CATCAGCAGGGCTAGGAAAT | | |
| PDS 2212 | F | GGGTGATGCTGACATCTGTG | *Zp3* Expression | |
| PDS 2213 | R | TCAGCTTCATCGGTCACG | | |
| PDS 2238 | F | AGATCTCGGACAGCACCAAG | *c-Kit* Expression | |

| | | | | |
|---|---|---|---|---|
| PDS 2239 | R | GAGTTGACCCTCACGGAATG | | |
| PDS 271 | F | GGAATTTTGGGGATTTTTTAGAGAGTTTATAAAGT | BS-PCR, *Rasgrf1* (WT) DMR | Epimark HS *Taq* (NEB M0490), Ta = 55ºC, te = 30s, 40 cycles |
| PDS 272 | R | CAAAAACAACAATAATAACAAAAACAAAAACAATAT | | |
| PDS 2627 | R | CCCRAAAATCCTCTAATCAACTAAC | BS-PCR, *tm5.1* DMR; use with PDS 271 | Epimark HS *Taq*, Ta = 55ºC, te = 30s, 40 cycles |
| PDS 2934 | F | TTAAGGGTGAAAAGTTGTATAAGGAG | BS-PCR for *Igf2r* | Epimark HS *Taq*, Ta = 55ºC, te = 30s, 40 cycles |
| PDS 2935 | R | CTAATAAAACACCTTCATTTACATAACCAA | | |
| PDS 287 | R | CTATATTAAATCCTTTTATCCACTATCCTCCACCC | BS-PCR, *tm1* DMR; use with PDS 271 | Epimark HS *Taq*, Ta = 58ºC, te = 30s, 40 cycles |

**Table S2. Total reads per DMR by Sample ID.**

| | | **Total Reads** | | |
|---|---|---|---|---|
| | | **DMR: Amplicon, Size** | | |
| **Sample ID** | **Fig** | **WT:** PDS271-272, 390-420bp | ***tm5.1*/Tg:** PDS271-2627, 350-390bp | **Igf2r:** PDS2394-5, 610 bp, 650bp |
| **Embryonic Gonads: Gonocyte (G) and Somatic Fractions (S)** | | | | |
| *TetON*/*tm5.1* + dox #1 G | 2a | 25 | 29 | 11 |
| *TetON*/*tm5.1* + dox #1 S | | 27 | 5675 | 178 |
| *TetON*/*tm5.1* + dox #2 G | S6 | 81 | 19 | 851 |
| *TetON*/*tm5.1*+ dox #2 S | | 617 | 7727 | 55 |
| *TetON*/*tm5.1* + dox #3 G | n.i. | 28067 | 133 | 0 |
| *TetON*/*tm5.1* + dox #4 G | 2b | 48 | 4476 | 168 |
| *TetON*/*tm5.1* + dox #4 G | | 42 | 25 | 47 |
| *TetOFF*/*tm5.1* #1 G | 2c | 103 | 6706 | 45 |
| *TetOFF*/*tm5.1* #1 S | | 74 | 4398 | 114 |
| *TetOFF*/*tm5.1* #2 G | S6 | 340 | 72 | 31 |

| | | | | |
|---|---|---|---|---|
| *TetOFF*/*tm5.1* #2 S | | 136 | 5870 | 147 |
| *TetOFF*/*tm5.1* #3 G | 2d | 1726 | 706 | 279 |
| *TetOFF*/*tm5.1* #3 S | | 771 | 715 | 11 |
| *TetON*/*tm5.1* – dox #1 G | | 26 | 7732 | 47 |
| *TetON*/*tm5.1* – dox #1 S | | 100 | 92 | 41 |
| **Tail (Soma)** | | | | |
| +/*TetOFF* #1 | n.i. | n.a. | 167 | n.m. |
| +/*TetOFF* #2 | | n.a. | 775 | n.m. |
| *tm5.1*/+ #1 | 3c | 248 | 4804 | n.m. |
| *tm5.1*/+ #2 | | 329 | 20 | n.m. |
| *tm5.1*/*TetOFF* #1 | | 351 | 16 | n.m. |
| *tm5.1*/*TetOFF* #2 | | 0 | 12 | n.m. |
| *TetOFF*/*tm5.1* | | 289 | 205 | n.m. |
| *TetOFF*/*tm5.1* | n.i. | 6 | 6087 | n.m. |
| *TetOFF*/*tm5.1* | 3c | 369 | 399 | n.m. |
| +/*tm5.1* (Biallelic) #3 | S1 | | | n.m. |
| +/*tm5.1* (Paternal) | e | 182 | 13183 | n.m. |
| +/*tm5.1* #1 | 3c | 148 | 17388 | n.m. |
| +/*tm5.1* #2 | | 653 | 406 | n.m. |
| +/*TetON* #1 | n.i. | n.a | 97 | n.m. |
| +/*TetON* #2 | | n.a | 122 | n.m. |
| *tm5.1*/+ #1 | 5c | 74 | 7425 | n.m. |
| *tm5.1*/+ #2 | | 135 | 35 | n.m. |
| *tm5.1*/*TetON* #1 | | 43 | 11161 | n.m. |
| *tm5.1*/*TetON* #2 | | 42 | 10416 | n.m. |
| *TetON*/*tm5.1* #1 | | 239 | 23 | n.m. |
| *TetON*/*tm5.1* #2 | | 217 | 13 | n.m. |
| | | **WT:** PDS271-272 | **Tg:** PDS271-2627 | ***tm1*:** PDS271-287, 461bp |
| +/Tg #1 | S1 1b | 2097 | 4180 | n.m. |
| +/Tg #2 | | 0 | 4498 | n.m. |
| Tg/+ #1 | | 4579 | 5455 | n.m. |
| Tg/+ #2 | | 4534 | 8842 | n.m. |
| Tg:*TetOFF*/+ #1 | S1 | 1249 | 0 | n.m. |
| Tg:*TetOFF*/+ #2 | 3d | 649 | 41 | n.m. |
| Tg/+ #1 | | 588 | 25 | n.m. |
| Tg/+ #2 | | 354 | 24 | n.m. |
| *TetOFF*/*tm1*/Tg #1 | S1 | 287 | 966 | 19 |
| *TetOFF*/ *tm1*/Tg #2 | 2d | 436 | 117 | 6 |
| *TetOFF*/ *tm1*/Tg #3 | | 510 | 71 | 3 |
| *tm1*/Tg #1 | S1 | 279 | 879 | 17 |
| *tm1*/Tg #2 | 2d | 188 | 1192 | 3 |
| **Sperm** | | | | |
| +/5.1 #1 | 2e | 97 | 11333 | 18 |
| +/5.1 #2 | | 588 | 10 | 116 |
| +/5.1 #3 | S6 d | 201 | 40 | 21 |
| *TetOFF*/5.1 #1 | 2e | 100 | 15 | 25 |
| *TetOFF*/5.1 #2 | | 45 | 7798 | 65 |

**Table S3. Sequences used for QUMA.**

| |
|---|
| **Primer Pair: PDS271-2**<br>**DMD Queried:** WT – C57Bl/6<br>**grep sequence:** cccaaaaatcctctaatcaactaac<br><br>**Reference Sequence:**<br>ccgcacttcactgttgcgctaccgctgcgctacaactaccacgactgctactgctgctgctgcactaccgttgcgctacg gctgccgcgctatcgctgctgctgccgcacttcgctgccgtgctgtccctgccccagccgctactgctgctcctgccccc ccactgcccctgccccagccactactgctgcccctgcccc |
| **Primer Pair:** PDS271-2<br>**DMR Queried:** WT – FVB<br>**Grep sequence:** ATTGTTTTTGTTTTAGTTATT<br><br>**Reference Sequence:**<br>ccgcaAGTActtcactgttgcgctaccgctgcgctacaactaccacgactgctactgctgctgctgcactaccgttgc gctacggctgccgcgctatcgctgctgctgccgcacttcgctgccgtgctgtccctgccccagccgctactgctgctcct gccccccccactgcccctgccccagccactactgctgcccctg |
| **Primer Pair:** PDS271-2627<br>**DMR queried:** *tm5.1*, Tg<br>**Grep sequence:** ATTGTTTTTGTTTTAGTTATT<br><br>**Reference Sequence:**<br>ccgcacttcactgttgcgctaccgctgcgctacaactaccacgactgctactgctgctgctgcactaccgttgcgctacg gctgccgcgctatcgctgctgctgccgcacttcgctgccgtgctgtccctgccccagccgctactgctgctcctgccccc ccactgcccctgccccagccactactgctgcccctgcccc |
| **Primer Pair:** PDS2394-5<br>**DMR queried:** Igf2r<br>**grep sequence:** GAGGGTTTAGAGGGTTC<br><br>**Reference Sequence:**<br>GGATTCGGAGGGTTTAGAGGGTTCCGCGGCACTGCGCCGCAGTGCTGCCCGA GGGTTCGGAGCAATTCCGGTTGTGCCGTGATCCTTGGTTGTGCTGAGTTGCGG TGAGGGAAAGGGAAGGGAAAGCTCAGAGGGTTCCGAGCTATCCTGAGGGTGC GAAGCTGCACAAGGGCA |

**Table S4. Total reads broken down by C57 and FVB fractions for MiSeq sequencing of PDS245-6 RT-PCR product in neonatal brain**.

| Genotype | Figure | C57 | FVB/129 | Total |
|---|---|---|---|---|
| 5.1/+ #1 | 3b | 3317 | 23 | 3340 |
| 5.1/*TetOFF* #1 | | 141 | 1247 | 1388 |
| 5.1/*TetOFF* #2 | | 200 | 1596 | 1796 |
| +/*TetOFF* #1 | | 4450 | 78 | 4528 |
| +/*TetOFF* #1 | | 2991 | 5 | 2996 |
| 5.1/+ #2 | | 2007 | 15 | 2022 |
| +/5.1:*TetOFF* #1 | | 3 | 844 | 847 |
| +/5.1:*TetOFF* #1 | | 16 | 2687 | 2703 |
| 5.1/+ #1 | 5b | 1509 | 22 | 1531 |
| 5.1/+ #2 | | 780 | 3 | 783 |
| *TetON*:5.1/+ #1 | | 1390 | 769 | 2159 |

| | | | | |
|---|---|---|---|---|
| *TetON*:5.1/+ #2 | | 1187 | 695 | 1882 |
| +/*TetON* #1 | n.i. | 2112 | 102 | 2214 |
| +/*TetON* #2 | | 14396 | 23 | 14419 |
| +/*TetON*:5.1 #1 | 5b | 37 | 3493 | 3530 |
| +/*TetON*:5.1 #2 | | 103 | 5319 | 5422 |
| FVB x B6 | n.i. | 32 | 2879 | 2911 |
| B6 x FVB | | 2959 | 17 | 2976 |
| B6 x FVB gDNA | n.i. | 1056 | 876 | 1932 |

**Table S5. SNP IDs and grep sequences for allele-specific PDS245-6 digestion and sequencing**.

| **RefSNP ID** | **C57Bl/6** | **FVB** | **grep sequence** |
|---|---|---|---|
| rs29947965 | C | T | tcttgcaaagtgcttcagat[c/t]cgct atgccagcgtggagcg |
| rs30280068 | T | C | Aacaccttcctgcattccta[t/c]cgc gtcttcaccaatgctat |

# III TRANSCRIPTIONAL CONTROL OF THE PITRNA

## III.A Introduction

Long noncoding RNAs (lncRNAs) are an active area of research as potential regulators of biological processes. While a small number of lncRNAs have been characterized extensively for decades (Brown *et al*, 1991; Borsani *et al*, 1991; Lanz *et al*, 1999), in 2009, the lncRNA repertoire was greatly expanded by a genome-wide screen for the chromatin signature of active transcription; namely, H3K4Me3 as a marker for an active promoter, followed by H3K36Me3 down the length of transcription (Guttman *et al*, 2009). These investigators go on to evaluate chromatin characteristics of lncRNA promoters and report strong enrichment for CAGE tags, suggestive of RNA Polymerase II-mediated transcription, as well as enrichment for RNA Polymerase II itself in mouse embryonic stem cells.

Other investigators have examined sequence attributes of lncRNA promoters, with the aim of gaining insight into factors that contribute to lncRNA regulation as well as lncRNA functionality. Genome-wide analysis demonstrated that lncRNA promoters are as well conserved as those of protein-coding genes, with a slight enrichment for HOX factor motifs, lending themselves to developmentally relevant function; however, the vast majority of lncRNA promoters do not contain HOX factor motifs. ChIP-Seq experiments have demonstrated that evolutionarily “older” lncRNA promoters were more enriched for SUZ12, a component of the PRC2 complex (Pasini *et al*, 2004), suggesting that ancient lncRNAs could be involved in mediating chromatin

context (Necsulea *et al*, 2014). Genome wide, lncRNA promoters are also skewed towards A/T richness and have a distinct transcription factor binding site (TFBS) repertoire relative to protein-coding genes (Alam *et al*, 2014).

Here, we sought to examine the transcriptional regulation of a specific lncRNA, the pitRNA. The pitRNA is expressed in the embryonic male germline and is targeted by piRNAs, contributing to imprinted methylation of the *Rasgrf1* differentially methylated region (DMR). Our lab has shown that transcription of the pitRNA initiates within the *Rasgrf1* repeats, a stretch of 40 tandem repeats of a 41bp GC-rich sequence; loss of the repeats ablates pitRNA expression (Watanabe *et al,* 2011). The repeats lack characteristics of a "textbook" mammalian promoter, lacking a TATA box, observed in approximately 10-20% of mammalian promoters (Cooper *et al*, 2006; Gershenzon *et al*, 2005) or an Inr element (Javahery *et al*, 1994; Weis and Reinberg, 1997). They are, as previously stated, GC-rich; in humans, approximately 50% of promoters are associated with a CpG island (Antequera and Bird, 1993), though these promoters typically drive ubiquitous genes (Saxonov *et al*, 2006). The nature by which the repeats drive the pitRNA, which is expressed in a tissue- and temporal-specific manner, then, is unknown.

We sought to elucidate the factors required for pitRNA expression, first by querying potential TFBS within the repeats *in silico*. Those TFs reported to be expressed in primordial germ cells were then queried for expression in an immortalized cell line derived from embryonic male gonads. Having identified one TF expressed in this cell line, we went on to characterize the effects of

knockdown and knockout on pitRNA expression.

### III.B Materials and Methods

Primer sequences used for all methods and analyses in Chapter III are listed in **Table III.1**.

**Generation of RST7A**: All assays concerning RST7A development and validation were performed by Roman Spektor. RST7A was derived by transfection of dissociated e16.5 male gonads with vectors carrying c-myc and H-ras (Kelekar *et al*, 1987; Thompson *et al*, 1989). Single cells were subcloned by limiting dilution and expanded. One clone, termed RST7A, constitutively expressed pitRNA as well as several male germline markers and factors required for imprinting.

**Identification of candidate transcription factors:** The *Rasgrf1* repeats sequences was scanned with TFSearch (Heinemeyer *et al*, 1998; Akiyama 1999) to identify TFs with potential TFBS within the repeats. Six TFs were chosen based on the repeats harboring sequence with greater than 85% sequence similarity to the TF's consensus binding sequence and their relative expression in primordial germ cells (Seisenberger *et al,* 2016).

**RST7A fixation for Chromatin Immunoprecipitation (ChIP):** 150mm plates of 80% confluent RST7A cells were cross-linked with 11% formaldehyde for 10 minutes at room temperature with intermittent swirling. Cross-linking was quenched with 1:20 volume of 2.5M glycine pH8.0 for 5 minutes, again with intermittent swirling. Fixed cells were collected by scraping; plates were washed with PBS Cell Wash Buffer (0.1% BSA, 2mM EDTA in calcium-free,

magnesium-free PBS). Cells were pelleted at 400 x g for five minutes at 10ºC, then resuspended with PBS Cell Wash Buffer. This step was repeated twice. Cells were resuspended in PBS Cell Wash Buffer and pelleted at 1000 x g for one minute. Supernatant was discarded, 5 uL 1000X Protease Inhibitor Cocktail (Sigma Aldrich P8340) was added, and pellets were snap frozen in liquid nitrogen.

**Chromatin Preparation**: In preparation for fragmentation, cell pellets were resuspended in Nuclei Release Buffer (50 mM HEPES pH 7.4, 150 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% NP-40; 0.25% Triton X-100) and rotated at 4ºC for ten minutes, then centrifuged at 2000 x g for one minute at 4ºC. Supernatant was discarded and 5 uL Protease Inhibitor Cocktail was added. Nuclei were resuspended in SDS Lysis Buffer (1% SDS, 10 mM EDTA, 50 mM Tris-HCl pH=8.0) and rotated at room temperature for 10 minutes, then centrifuged at 2000 x *g* for one minute at 4ºC. Supernatant was discarded and 5 uL Protease Inhibitor was added. Chromatin was resuspended in Sonication Buffer (50 mM Tris-HCl pH=8.0, 100 mM NaCl, 2 mM EDTA, 0.5% sarkosyl) and centrifuged 2000 x g for one minute at 4ºC for one minute. This was repeated three times. 5 uL Protease Inhibitor Cocktail and added and snap frozen in liquid nitrogen prior to shearing.

**Chromatin shearing:** Fixed chromatin was fragmented to 1kb on a Covaris S2 sonicator with the following parameters: 5% duty cycle, 15 cycles, 2 intensity, 200 cycles/burst, 60s cycle time for three minutes. Fragmentation size was confirmed via gel electrophoresis. Sheared chromatin was spun at

20,000 x g for ten minutes at 4ºC; supernatant was discarded, and pellet was resuspended in residual supernatant, transferred to a new low bind microfuge tube. 5% of total sheared chromatin was removed for input; the remainder was snap frozen in liquid nitrogen.

**Chromatin immunoprecipitation (ChIP):** 100 uL Protein G DynaBeads (ThermoFisher 10003D) were added to Bead Blocking Buffer (0.5% BSA in calcium-free, magnesium-free PBS) and mixed with gentle pipetting and applied to a DynaMag Spin Magnet (ThermoFisher 12320D). Beads were loaded with 10 ug Rabbit anti-mouse Sp1 (Millipore 07-645) or normal mouse IgG (Life Technologies 10400C) and rotated at 4ºC overnight. Beads were washed with Bead Blocking Buffer and resuspended in ChIP Dilution Buffer (12.5 mM Tris pH 8, 162.5 mM NaCl, 2 mM EDTA, 1.25% Triton X-100, 0.625% BSA). Sheared chromatin was added to antibody loaded beads and rotated at 4ºC overnight. ChIP samples were washed the next day with chilled Low Salt Wash Buffer (1% Triton X-100, 0.1% SDS, 2 mM EDTA, 150mM NaCl, 20 mM Tris-HCl pH 8), then with High Salt Wash Buffer (1% Triton X-100, 0.1% SDS, 2mM EDTA, 500mM NaCl, 20 mM Tris pH 8), then with LiCl2 Wash Buffer (1% NP40, 1% deoxycholate, 1mM EDTA, 250mM LiCl, 10 mM Tris pH 8), and finally with Tris-EDTA (TE). *Reverse crosslinking*: Reverse crosslinking was performed by incubating ChIP samples with SDS Lysis Buffer at 65C overnight; then with 50 ug Proteinase K (ThermoFisher 25530049). Samples were applied to the DynaMag Spin Magnet. Supernatant was collected. Beads were washed with Buffer ERC; supernatant was pooled with

previous supernatant. DNA was column purified with BioBasic EZ-10 DNA Columns (BioBasic BS427) and eluted in 30 uL of Tris-HCl pH 8.0, then quantified with Nanodrop. 5 ng DNA was used in subsequent qPCR experiments along with 5% input.

**Mithramycin assays:** RST7A cells were cultured in 0 or 400 nM Mithramycin A (MTM, Sigma M6891) for 48 hours, then harvested for total RNA and gDNA.

**CRISPR/Cas9-mediated knockout of Sp1:** 75% confluent 6-well cultures of RST7A cells were transfected with pairs of the Cas9-sgRNA vector pX330 (Addgene Plasmid 42230) that had been modified to carry different sgRNA sequences targeting the promoter (ΔP1, ΔP2) and distal 6$^{th}$ exon (ΔE1, ΔE2, ΔE3) of Sp1 via the Zhang lab protocol (Cong *et al*, 2013). Plasmids were cotransfected in the following combinations: ΔP1/ΔE1, ΔP2/ΔE2, ΔP1/ΔE3 using Lipofectamine 2000 (ThermoFisher 11668027), following the manufacturer's protocol. After recovering overnight, cultures were subjected to G418 selection at 250 ug/mL G418 for eight days. Twenty-three single cell clones were recovered and genotyped for the full Sp1 deletion, Sp1$^{\Delta}$, as well as an internal Sp1 PCR (Table III.1) to determine homozygous vs. heterozygous knockout. Three heterozygous clones were confirmed. One heterozygous clone was expanded and subjected to an additional round of CRISPR/Cas9, as was a wild type RST7A culture. For this round of CRISPR/Cas9 genome editing, we used pETC9.4, a derivative of pX330 modified to carry a hygromycin resistance cassette. pETC9.4 was modified to carry five different sgRNA sequences targeting the proximal 5$^{th}$ and distal 6$^{th}$

exons of Sp1, termed Δ5.1, Δ5.2, Δ6.1, Δ6.2, and Δ6.3. These were cotransfected in the following combinations: Δ5.1/Δ6.1; Δ5.2/Δ6.2; Δ5.2/Δ6.2/Δ6.3 and selected with 400 ug/mL hygromycin for ten days. Eighteen single cell clones were recovered and genotyped for the exon 5/6 deletion ($Sp1^{\Delta 5/6}$). One $Sp1^{\Delta/\Delta 5/6}$ and three $Sp1^{\Delta 5/6/\Delta 5/6}$ were confirmed by Sanger sequencing of diagnostic PCR products. Loss of Sp1 expression was confirmed by qRT-PCR. Western blots were not performed.

**RNA and gDNA extraction:** Total RNA extraction was performed with Trizol following the manufacturer's protocol. gDNA extraction was performed with Laird's Lysis Buffer (Laird *et al,* 1991).

**RT:** All samples were DNAse treated, random primed, and reverse transcribed with Promega RQ Dnase and GoScript Reverse Transcriptase (Promega M6101 and A5003).

**qPCR**: For all assays, qPCR was performed on random-primed cDNA using Power SYBR Green Master Mix on an Applied Biosystems 7500 using an annealing temperature of 60ºC and forty cycles of amplification. An RT- condition (no RT) was run for each sample as a negative control for gDNA contamination.

**Bisulfite conversion and analysis:** Bisulfite conversion was performed using the Zymo EZ DNA Lightning Conversion kit. Bisulfite PCR (BS-PCR) was performed using NEB Epimark *Taq* Polymerase (M0490S) with primers as listed in Table III.1. COBRA was performed on BS-PCR products by incubation with 5U BstUI (NEB R0518) at 60C for 1 hour. Digestion products

were separated via agarose gel electrophoresis (4% agarose).

**Table III.1. Primer details for all analyses in Chapter III.** Unless otherwise stated, all primers were utilized with an annealing temperature of 60ºC and an extension time (te) of 30 seconds; BS-PCR primers were utilized with an annealing temperature of 55ºC. All qPCR primers are 87 to 113% efficient.

| Primer Name | 5' – 3' | Sequence | Purpose/Notes |
|---|---|---|---|
| PDS 2458 | F | TTGGGAGGCTGGTTTTGACC | SREBP-1 expression |
| PDS 2459 | R | CTGGCTGGCCAATGGACTAC | |
| PDS 2428 | F | ATGGTGTGGGAGTGCCATC | Cdx1 expression |
| PDS 2429 | R | CAACGCCTAGAGCTGGAAAA | |
| PDS 2434 | F | GAGACAGCAGGTGGAGAAGG | Sp1 expression |
| PDS 2435 | R | AAAGCGCTTCCCACAATATG | |
| PDS 2430 | F | CCGTCTTCAAGGTGTCCAAG | GATA-1 expression |
| PDS 2431 | R | CTGGTGTCCTCACCATCAGA | |
| PDS 2436 | F | TCTGCAGGGGGTAGTGTAGC | GATA-2 expression |
| PDS 2437 | R | CCGGTTCTGTCCATTCATCT | |
| PDS 2432 | F | AATGGTTGTGCCACCTCTCC | c-Myb expression |
| PDS 2433 | R | TGGTGGAACAGAACGGAACA | |
| PDS 2660 | F | GCCGCTGCTGCAATTTCTG | *Rasgrf1* repeats |
| PDS 2661 | R | TGGGGCAAGTCTGATCTTCC | |
| PDS 2610 | F | TAACGACCATTGGCCAAAGC | Sp1 promoter |
| PDS 2611 | R | GAACGAACAGCCAATTACGC | |
| PDS 2518 | F | GTGTACAGGCCCAATCCGTA | Chr13 intergenic |
| PDS 2519 | R | GGATCTCTGAATTGCCACCT | |
| PDS 2646 | F | caccGAAAAACGCGGACGCTGACG | sgRNA Sp1 Promoter (ΔP1) |
| PDS 2647 | R | aaacCGTCAGCGTCCGCGTTTTTC | |
| PDS 2648 | F | caccCAAGCGAACCCGGACCGGAC | sgRNA Sp1 Promoter (ΔP2) |
| PDS 2649 | R | aaacGTCCGGTCCGGGTTCGCTTG | |
| PDS 2650 | F | caccGCAAGACGGGCAATACCCTC | sgRNA Sp1 proximal Exon 6 (ΔE1) |
| PDS 2651 | R | aaacGAGGGTATTGCCCGTCTTGC | |
| PDS 2652 | F | caccGGCCTCCATGGCTACCATAT | sgRNA Sp1 proximal Exon 6 (ΔE2) |
| PDS 2653 | R | aaacATATGGTAGCCATGGAGGCC | |
| PDS 2992 | F | caccATCTCCGAGCACACTTGCGC | sgRNA Sp1 Exon 5 (Δ5.1) |
| PDS 2993 | R | aaacGCGCAAGTGTGCTCGGAGAT | |
| PDS 2994 | F | caccCATATACTTTGCCGCATCCT | sgRNA Sp1 Exon 5 (Δ5.2) |
| PDS 2995 | R | aaacAGGATGCGGCAAAGTATATG | |
| PDS 2996 | F | caccATGAAACGCTTAGGGCACTC | sgRNA Sp1 Distal Exon 6 (Δ6.1) |
| PDS 2997 | R | aaacGAGTGCCCTAAGCGTTTCAT | |
| PDS 2998 | F | caccAGTGCCCTAAGCGTTTCATG | sgRNA Sp1 Distal Exon 6 (Δ6.2) |
| PDS 2999 | R | aaacCATGAAACGCTTAGGGCACT | |
| PDS 3000 | F | caccGGCCTCCATGGCTACCATAT | sgRNA Sp Distal Exon 6 (Δ6.3) |
| PDS 3001 | R | aaacATATGGTAGCCATGGAGGCC | |
| PDS 2230 | F | CTAAGGCCAACCGTGAAAAG | Actin expression |
| PDS 2231 | R | ACCAGAGGCATACAGGGACA | |
| PDS 2266 | F | ATCCGTGGCTACCGCTATTG | pitRNA expression |

| | | | |
|---|---|---|---|
| PDS 2267 | R | AGTCGTGGTAGTTGTAGCGC | |
| PDS 2817 | F | GGGTTCCAAGAAGGCTGTCA | H2Be expression |
| PDS 2818 | R | CGGGTGCACCTGCTTCAG | |
| PDS 2710 | F | GTGTCTCCCACCTTTTCAAAGT | Mecp2 expression |
| PDS 2711 | R | GAAGTCTGGCCGATCTGCTG | |
| PDS 2871 | F | CATCGTCATCTCATCCCGCC | Sp1∆ genotyping, te = 1m |
| PDS 2878 | F | TCTCTGCAAGTCCATCGTCA | Internal Sp1 |
| PDS 2787 | R | GGCGGGATGAGATGACGATG | Sp1∆ genotyping |
| PDS 3002 | F | GGTTCCCTACTTTCTAGCTCCA | Sp1∆5/6 genotyping |
| PDS 3003 | R | CCATTTTCAAGGAGTCACCAG | |
| PDS 271 | F | GGAATTTTGGGGATTTTTTAGAGA GTTTATAAAGT | BS-PCR, *Rasgrf1* DMR |
| PDS 272 | R | CAAAAACAACAATAATAACAAAAA CAAAAACAATAT | |

## III.C Results

### III.C.1 Sp1 is one of six assayed transcription factors that is expressed in RST7A.

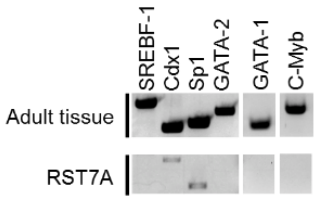
TFSearch (Akiyama 1998) revealed six TFs with predicted TFBMs in the *Rasgrf1* repeats, as follow: Cdx1, GATA-1, GATA-2, c-Myb, Sp1, and SREBP-1, and which were expressed in primordial germ cells (Seisenberger *et al*, 2016). Of these six, only Sp1 was robustly expressed in RST7A by end-point RT-PCR (**Fig III.1b**). Sp1 has two canonical binding sites at the 3' edge of the *Rasgrf1* repeats distal to the *Rasgrf1* DMR. **Fig III.1a** lists these factors as well as their expression in adult tissues, primordial germ cells, and RST7A.

**Fig III.1. Sp1 is expressed in RST7A. a)** Transcription factors with predicted binding sites in the *Rasgrf1* repeats are listed along with known tissues for expression confirmation, relative expression in PGCs, and expression in RST7A. **b)** End point RT-PCR for candidate transcription factors in RST7A. All primers are intron spanning. Adult tissues, which served as positive controls for RT-PCR reactions, are, in order, testis, colon, muscle, spleen (last three). Note that the band in RST7A-Cdx1 lane corresponds to gDNA amplification and not cDNA.

a

| TF | # Sites | Expression Pattern | | |
|---|---|---|---|---|
| | | Adult | PGC | RST7A |
| Cdx1 | 2 | Colon | ++ | No |
| GATA-1 | 9 | Spleen, endothelium | + | No |
| GATA-2 | 9 | Leukocytes, male gonads | + | No |
| c-Myb | 2 | Erythrocytes, leukocytes | ++ | No |
| **Sp1** | **2** | **Muscle, gonads, many others** | **++** | **Yes** |
| SREBP-1 | 1 | Muscle, gonads, many others | +++ | No |

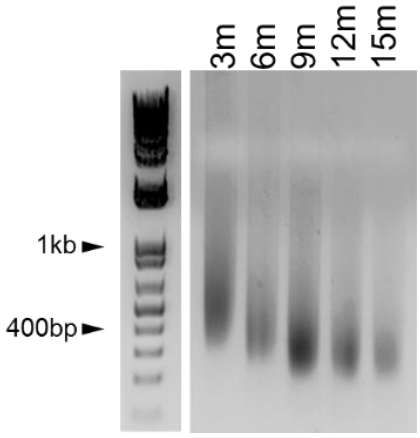
b




### III.C.2 Sp1 is enriched at the *Rasgrf1* repeats in RST7A.

We first queried whether Sp1 was enriched at the *Rasgrf1* repeats. We sheared RST7A chromatin to an average of 500bp to 1kb fragments (**Fig III.2**) as has been used for Sp1 ChIP-Seq and ChIP-PCR by others (Wang *et al*, 2013; Zu *et al*, 2009; He *et al*, 2005).

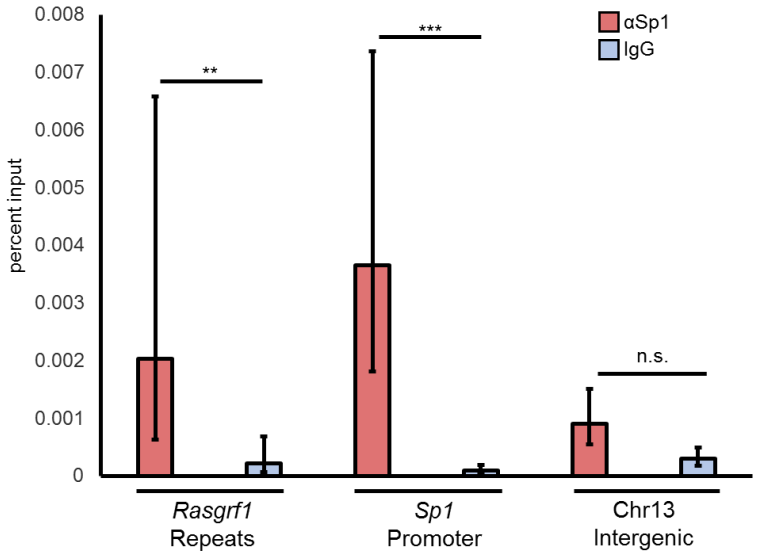
**Fig III.2**. **Distribution of chromatin fragments relative to shearing time, using RST7A chromatin**. 3m shearing time produced an average fragment size of 500-1000bp and was used for all ChIP experiments.




We performed anti-Sp1 ChIP-qPCR for the *Rasgrf1* repeats, using the promoter of Sp1, which is self-regulatory, as a positive control, and an intergenic region upstream of the *DHFR* gene on chr13 with no canonical Sp1

sites as a negative control. We found that Sp1 was enriched at the *Rasgrf1* repeats relative to normal mouse IgG (**Fig III.3**).

**Fig III.3 Sp1 is enriched at the *Rasgrf1* repeats as assayed by ChIP-qPCR.** Color key for anti-Sp1 vs. mouse IgG as a negative control at right. Sp1 is enriched at both the *Rasgrf1* repeats and the Sp1 promoter compared to an intergenic region on chr13. Enrichment is shown as percent of total input. Error bars represent standard error among biological replicates. Experiments were performed in biological duplicate at minimum. **, $p < 0.01$; ***, $p < 10^{-6}$; n.s., not significant.



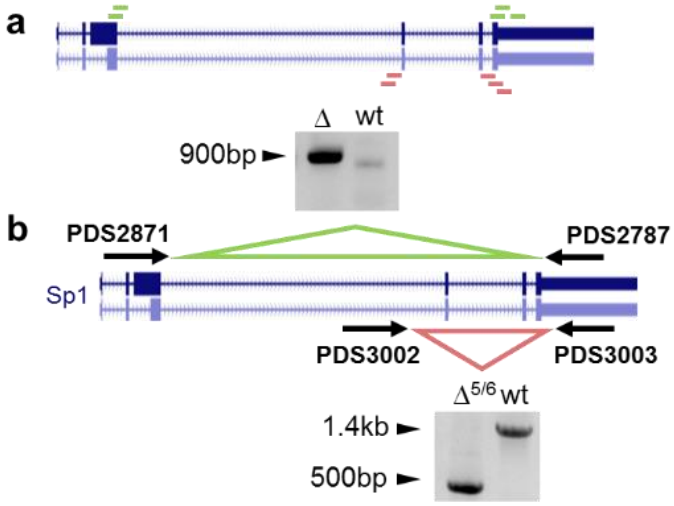

### III.C.2.i Chemical knockdown and genetic knockout of Sp1 downregulates pitRNA expression.

With confirmation that Sp1 is enriched at the repeats in an immortalized cell line that capitulates many aspects of the pitRNA system, we then hypothesized that Sp1 inhibition or knockout would reduce pitRNA expression. Mithramycin (MTM) is a known inhibitor of the Sp family of transcription factors by binding target sequences (Blume *et al,* 1991). We treated RST7A cultures with 400 nM MTM and evaluated expression of pitRNA. We found that MTM

causes an approximately 50% reduction in pitRNA expression (**Fig III.5a**).

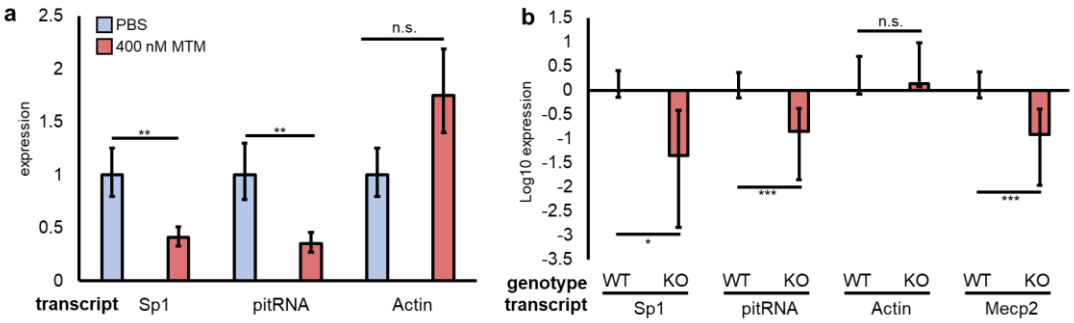
*Sp1* null animals are embryonic lethal by day 11 of gestation; however, Sp1 embryonic stem cells are viable (Marin *et al,* 1997). We generated *Sp1* knockouts using CRISPR-Cas9 genome editing. We first endeavored to generate homozygous knockouts of the entire Sp1 gene (**Fig III.4.a**, upper) ($Sp1^{\Delta/\Delta}$); only one homozygous clone was recovered. Heterozygotes ($Sp1^{\Delta/+}$) and wild type cultures were then subjected to an additional round of CRISPR/Cas9-mediated knockout of exons 5 and 6, which code for Sp1's DNA-binding domain (**Fig III.4.a**, lower). One double heterozygote ($Sp1^{\Delta/\Delta 5/6}$) and three homozygotes for the exon 5 and 6 deletion ($Sp1^{\Delta 5/6/\Delta 5/6}$) were recovered and confirmed by Sanger sequencing of deletion-spanning PCR products (**Fig III.4.b**).

**Fig III.4. Targeting scheme for CRISPR/Cas9-mediated knockout of Sp1**. **a)** Schematic of the Sp1 gene with locations of sgRNA sequences. Blue bars denote exons 1-6 from the left; exon 6 is contiguous with the 3' UTR (thinner blue bar). Upper: Green bars denote location of sgRNAs for the full Sp1 knockout ($Sp1^{\Delta}$). Lower: Red bars denote sgRNA locations for the exon 5/6 knockout ($Sp1^{\Delta 5/6}$). **b)** Diagnostic PCRs for $Sp1^{\Delta}$ and $Sp1^{\Delta 5/6}$. Note that the wild type PCR product for PDS2871-2787 is predicted to be approximately 10kb and so will not be evident with the PCR parameters used. Faint band at < 900bp in the wild type lane is nonspecific product.




As a confirmation of Sp1 knockout, Sp1 expression was assayed in $Sp1^{\Delta/\Delta5/6}$ and $Sp1^{\Delta5/6/\Delta5/6}$ via qRT-PCR. In these lines, Sp1 was reduced to approximately 10% of normal (**Fig III.5b**).

**Fig III.5 pitRNA expression is decreased with chemical knockdown or genetic knockout of Sp1. a)** Mithramycin reduces Sp1 target gene expression. Sp1 is self-regulatory and decreases with MTM treatment, as does pitRNA; actin, which is not regulated by Sp1, is not significantly affected. **b)** Genetic knockout of Sp1 leads to reduced expression of pitRNA and *Mecp2*, a known target gene for Sp1, but not actin. qPCR primers for Sp1 precede the exon 5-6 deletion, as such, some Sp1 transcription is detectable and expected in knockout cell lines. All expression levels are normalized to Histone 2BE. Error bars represent standard error across biological replicates; two wild type RST7A clones and four Sp1 knockout clones were used. *, $p < 0.05$, **, $p < 0.01$; ***, $p < 10^{-6}$; n.s., not significant. WT, wild-type; KO, Sp1 knockout.

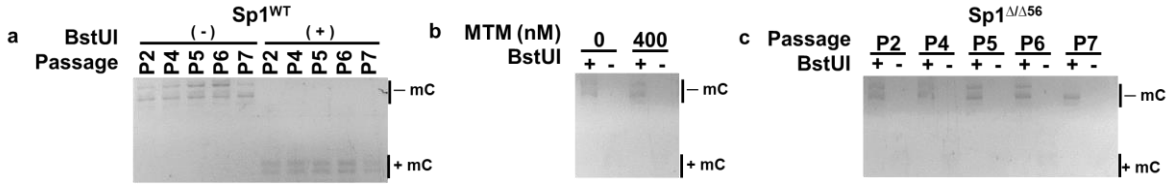



**III.C.3 Sp1 knockdown and knockout do not affect methylation of the**

***Rasgrf1* DMR in RST7A.**

To determine whether pitRNA is required for maintenance of methylation at the *Rasgrf1* DMR, we queried methylation status in MTM-treated and *Sp1*-knockout RSTA lines over several passages using COBRA. Wild type RST7A are fully methylated at the DMR (**Fig III.6.a**). The DMR remained fully methylated in both chemical knockdown cultures (**Fig III.6.b**) and Sp1 knockout lines (**Fig III.6.c**).

**Fig III.6. Chemical knockdown and genetic knockout of Sp1 do not affect methylation of the *Rasgrf1* DMR in RST7A as measured by COBRA. a)** Wild type RST7A over several passages are hypermethylated. **b)** The DMR remains hypermethylated in RST7A treated with 400 nM MTM for 48 hours compared to untreated controls. **c)** An Sp1 double heterozygote (Sp1Δ/Δ5/6) remains hypermethylated at the DMR over several passages. Note that in **b)** and **c)**, the ethidium bromide front has moved past digested products; as such they are considerably fainter than the undigested control. Complete loss of the undigested band also indicates full digestion and therefore hypermethylation.




## III.D Discussion

### III.D.1 Specific factor 1 (Sp1) binds the *Rasgrf1* repeats and drives pitRNA expression.

In a candidate transcription factor (TF) screen of RST7A, Specificity Factor 1 (Sp1) was the only TF predicted to bind the *Rasgrf1* repeats expressed at appreciable levels in RST7A, a cell line known to express the pitRNA. Specific Factor 1 or Sp1 is a ubiquitous transcription factor known to

bind the GC-rich 5'-GGGCGG-3' consensus sequence. *Sp1* null animals are embryonic lethal by day 11 of gestation, likely due to its many targets of regulation. Notably, the expression of many Sp1 targets were not significantly downregulated in *Sp1* knockouts (Marin *et al,* 1997), possibly due to redundant activity of the SP1 family, of which there are several members (reviewed in Philipsen and Suske 1999), or alternatively, residual Sp1 carried to the embryo in the ooplasm. Cooperative activity between some family members is also likely, as compound heterozygotes of Sp1 and Sp3 knockouts are not viable (Kruger *et al,* 2007). While the ubiquitous nature of Sp1 might be contrary to its potential role in a tightly temporospatially regulated process, Sp1 is known to act in a cell state- and cycle-specific manner via extensive post-translationally modifications. These can not only modulate cellular localization, but can also affect its DNA binding activity (Merchant *et al,* 1999; reviewed in Bouman and Philepsen 2002; Solomon *et al,* 2008; Chu and Ferro 2005). Moreover, Sp1 demonstrates cooperative or competitive interactions with other transcription factors (Janson and Pettersson 1990; Fischer *et al,* 1993; Billon *et al,* 1999; Nam *et al,* 2012), lending further potential specificity to its activity.

We demonstrate that Sp1 is enriched at the *Rasgrf1* repeats by Sp1 ChIP-qPCR. Further, MTM-mediated Sp1 knockdown and targeted *Sp1* knockouts downregulate pitRNA, suggesting that Sp1 drives pitRNA expression. While these data do not rule out an indirect relationship between Sp1 and pitRNA expression, binding of Sp1 at the *Rasgrf1* repeats as assayed by anti-Sp1

ChIP qPCR *in vitro* indicates that a direct relationship is likely.

### III.D.2 Sp1 knockdown and knockout does not alter methylation at the *Rasgrf1* DMR.

COBRA of early and late passages of Sp1 double knockouts indicates that methylation persists at the *Rasgrf1* repeats. From this, we conclude that the pitRNA is not required for maintenance of DMR methylation in RST7A.

A strong consideration is that all of these experiments were performed *in vitro* in an immortalized cell-based system, RST7A, and may not fully recapitulate *in vivo* mechanisms. However, other independent lines of evidence performed *in vivo* suggest that the pitRNA is not sufficient for the establishment of methylation at *Rasgrf1*. As such, further pursuit of Sp1 and its role specifically in driving pitRNA transcription may not warrant further study. However, Sp1 in the context of other data and indications at *Rasgrf1* could support a role for Sp1 beyond the control of the pitRNA.

### III.D.3 Alternative roles for Sp1 and the *Rasgrf1* repeats.

We identify one transcription factor, Sp1, that binds the repeats in this cell-based system. Chemical knockdown and genetic knockout of Sp1 leads to a downregulation of pitRNA expression. However, loss of Sp1 is not correlated with loss of methylation at the *Rasgrf1* DMR. These data support a role for Sp1 in driving pitRNA expression, though it appears that methylation, once established, is not dependent upon pitRNA expression as evidenced in our cell-based system. As we also demonstrate an RNA-independent role of the repeats for the establishment of DNA methylation at the DMR in Chapter II,

alternative roles for Sp1 binding at the *Rasgrf1* repeats remain open to speculation.

One potential hypothesis concerning the *Rasgrf1* repeats could involve differential G-quadruplex (G4) formation on the maternal vs. paternal chromosome, which the *Rasgrf1* repeats are predicted to form. While Sp1 has a canonical binding motif, it is also known to bind G4 (Raiber *et al,* 2012); should methylation of the paternal DMR by Dnmt3c modulate formation of a G4, Sp1 binding could vary between the maternal and paternal chromosome, with potential effects on local transcription or chromatin structure; G4s as a potential mode of regulation are discussed more extensively in Chapter II.

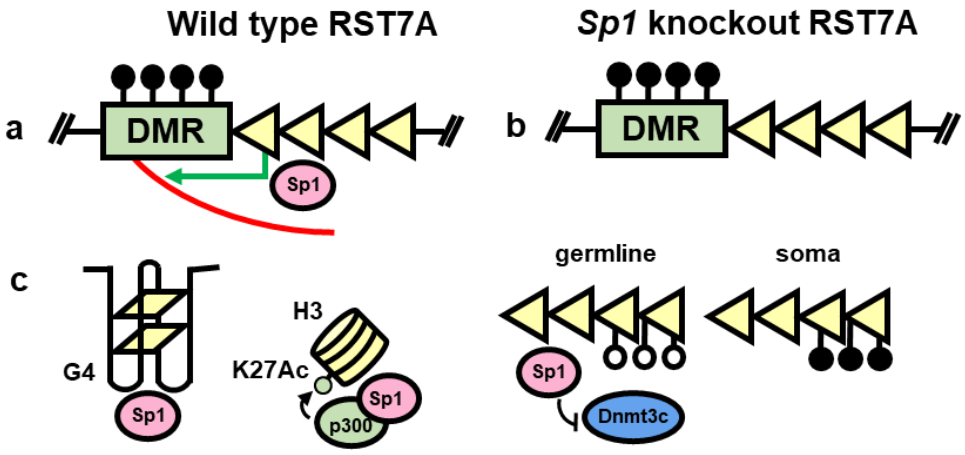
Second, Sp1 has been demonstrated to bind annotated enhancers to modulate gene expression (Isomura *et al*, 2005; Nolis *et al*, 2009; Sumimoto *et al*, 2012). It associates with the histone methyltransferase Set1 and the histoneacetyltransferase p300 and could conceivably impact histone state at the repeats. We have previously shown that both maternal and paternal alleles of the *Rasgrf1* ICR are marked with H3K9Me (Lindroth *et al*, 2008), making a concurrent H3K4 methylation mark less likely (Binda *et al*, 2010, reviewed in Eissenberg *et al*, 2010). However, marking of the repeats with the H3K27Ac mark would be consistent with an enhancer-like role (reviewed in Calo and Wysocka, 2013), which data from our lab could support (Yoon *et al*, 2005). Whether H3K27Ac is differentially enriched in the female and male germlines, and how it might impact the local chromatin environment, could be an avenue for future studies.

A final hypothesis focuses on a potential role for Sp1 mediating methylation at the repeats rather than the DMR. Sp1 is known to protect GC-rich regions from DNA methylation (Macleod *et al,* 1994; Mummaneni *et al,* 1998; reviewed in Caiafa and Zampieri, 2005). In turn, DNA methylation attenuates Sp1 binding (Mancini *et al,* 1999; Zelko *et al,* 2010; Reynard *et al,* 2014). Somatically, a portion of the repeats amplifiable by bisulfite PCR do acquire methylation on the paternal allele (Lindroth *et al*, 2008). Docking of Sp1 at the repeats in the male embryonic germline could contribute to the restriction of methylation to the DMR, with a side effect of driving pitRNA expression, leading to DNA methylation by way of the pitRNA-piRNA pathway.

In conclusion, we identify the transcription factor Sp1 that binds the repeats *in* vitro. Chemical knockdown and genetic knockout of Sp1 leads to a downregulation of pitRNA expression. However, loss of Sp1 is not correlated with loss of methylation at the *Rasgrf1* DMR. These data support a role for Sp1 in driving pitRNA expression. Alternative roles for Sp1 as a binder of the *Rasgrf1* repeats could be explored as routes of further investigation (**Fig III.7**).

**Fig III.7. Summary of Sp1 at the *Rasgrf1* repeats in RST7A cells. a)** In wild type RST7A, Sp1 is enriched at the *Rasgrf1* repeats and pitRNA is expressed (green arrow). The *Rasgrf1* DMR is fully methylated (filled lollipops). **b)** Genetic knockout and chemical knockdown of Sp1 ablates pitRNA expression; however, methylation is maintained at the *Rasgrf1* DMR. **c)** Potential alternative roles for Sp1 at the repeats. The repeats, shown in yellow, could form secondary structure such as G4s to affect local chromatin state, which Sp1 is known to bind as well as its canonical sequence. Second, Sp1 could recruit the histone acetyltransferase p300 to the repeats, thereby affecting local chromatin state and transcription. Finally, Sp1 could play a role in protecting the repeats from methylation by Dnmt3c in the male germline; lack of Sp1 binding in the soma could promote accrual of methylation at the

repeats.


Wild type RST7A
Sp1 knockout RST7A
a
DMR
Sp1
b
DMR
c
G4
Sp1
H3
K27Ac
Sp1
p300
germline
soma
Sp1
Dnmt3c

# IV LONG NON-CODING RNA REGULATION OF REPRODUCTION AND DEVELOPMENT

This Chapter was published as a comprehensive review of the roles and mechanisms for long noncoding RNAs in the regulation of reproduction and development, as cited:

Taylor DH, Chu ET, Spektor R, and Soloway PD (2015). Long non-coding RNA regulation of reproduction and development. *Molecular Reproduction and Development* 82: 932–956.

## IV.A Abstract

Non-coding RNAs (ncRNAs) have long been known to play vital roles in eukaryotic gene regulation. Many studies from more than a decade ago revealed that maturation of spliced, polyadenylated coding mRNA occurs by reactions involving snRNAs and snoRNAs; mRNA translation depends on activities mediated by tRNAs and rRNAs, subject to negative regulation by miRNAs; transcriptional competence of sex chromosomes and some imprinted genes is regulated in cis by ncRNAs that vary by species; and both siRNAs and piRNAs bound to Argonaute family proteins regulate post translational modifications on chromatin, and local gene expression states. More recently, additional gene-regulating ncRNAs have been identified. Among them are the long intergenic and long ncRNAs (collectively referred to as lncRNAs), a class totaling more than 100,000 transcripts in human, including some of the previously mentioned RNAs that regulate dosage compensation and imprinted gene expression. Here, we first provide an overview of lncRNA activities, and then review the role of lncRNAs in processes vital to reproduction, including

germ cell specification, sex determination and gonadogenesis, sex hormone responses, meiosis, gametogenesis, placentation, non-genetic inheritance, and pathologies affecting reproductive tissues. Results from a diversity of species are presented as they are informative of evolutionarily-conserved processes.

IV.B Discovery of lncRNAs

Initial efforts to characterize the mammalian transcriptome, comprehensively, revealed an abundance of RNAs that vastly exceeded what was expected from the coding genome. Many RNAs were identified as non-coding and distinct from previously known non-coding species including snRNA, snoRNA, tRNA, rRNA and Argonaute family associated small RNAs (Bertone *et al,* 2004; Carninci *et al,* 2005; Cheng *et al,* 2005; Kampa *et al,* 2004; Kapranov *et al,* 2005; Rinn *et al,* 2003; Shiraki *et al,* 2003). A very limited number of non-coding RNAs that fit into none of the known non-coding classes had previously been characterized, including *Xist*, transcribed from the inactive X-chromosome in female mammals (Borsani *et al,* 1991; Brockdorff *et al,* 1991; Brown *et al,* 1991); the imprinted transcripts Airn (Lyle *et al,* 2000), *H19* (Brannan *et al,* 1990; Sleutels *et al,* 2002), and Kcnq1ot1 (Lee *et al,* 1999); and *SRA* (Lanz *et al,* 1999; Lanz *et al,* 2002). Of these, *Xist*, Airn and Kcnq1ot1 had been respectively implicated or shown to regulate, in cis, X-inactivation (Marahrens *et al,* 1997; Penny *et al,* 1996), imprinted expression of Igf2r (Sleutels *et al,* 2002) and imprinted expression of genes in the Kcnq1 cluster (Fitzpatrick *et al,* 2002). *SRA* was shown to enhance steroid hormone

receptor responses (Lanz *et al,* 1999; Lanz *et al,* 2002). Functions of other lncRNAs remained uncharacterized, and it was not clear if indeed they had functions or were simply byproducts of transcriptional noise. In 2007, 231 additional lncRNAs from the human HOX clusters were discovered. One of these, *HOTAIR*, transcribed from the HOXC cluster was shown to regulate coding transcripts from the HOXD cluster in trans (Rinn *et al,* 2007). This finding extended the functions of the few characterized lncRNAs beyond cis-regulation of the silent X-chromosome and a few imprinted genes.

With the development of genome-wide chromatin-state maps using chromatin immunoprecipitation followed by sequencing (ChIP-Seq), it was observed that known genes actively transcribed by RNA polymerase II carried H3K4me3 at the promoter, and H3K36me3 across the transcribed region. These so-called K4-K36 domains were also found at sites not previously annotated as genes, leading to the discovery of more than 1,600 intergenic, spliced non-coding transcripts. The fact that they were found to be evolutionarily-conserved, with many exhibiting coordinated regulation, argued against the notion that they represented transcriptional noise (Guttman *et al,* 2009). Since that report, more than 100,000 lncRNAs have been described for human alone (Volders *et al,* 2015a). Accepted lncRNA properties, and practices for their identification and naming, are evolving, but lncRNAs generally exhibit the following features (Mattick and Rinn 2015): Lengths are >200 nt with a median size of ~500 nt, smaller than mRNAs, though some exceed 100 kb; 98% are spliced, with 80% having 2-4 exons, and the majority

existing as a single isoform; most are polyadenylated, though Poly(A)– forms exist more commonly than they do for mRNAs; many show nuclear enrichment and chromatin association, though cytoplasmic forms exist; coding potentials are low, as assayed by Codon Substitution Frequency (CSF) scores (Cabili *et al,* 2011; Guttman *et al,* 2009; Guttman *et al,* 2010), low ribosome association (Guttman *et al,* 2013) and lack of open reading frames >100 nt; expression levels are lower than mRNAs and expression is more tissue specific; purifying selection is common, though with weaker constraints than coding transcripts, and in some cases, structure rather than sequence might be under selection (Derrien *et al,* 2012; Smith *et al,* 2013); some are circular (Hansen *et al,* 2013; Memczak *et al,* 2013; Petkovic and Muller 2015; Zhang *et al,* 2013).

Databases established to describe lncRNAs include lncRNADisease (Chen *et al,* 2013), LncRBase (Chakraborty *et al,* 2014), NONCODE (Xie *et al,* 2014), LNCipedia (Volders *et al,* 2015b), lncRNAdb (Quek *et al,* 2015), lncRNAWiki (Ma *et al,* 2015) and RNAcentral (RNAcentral-Consortium 2015). As with many such databases, these are likely to include mis-annotated mRNAs. For example, some RNAs classified as lncRNAs associate with ribosomes (Chew *et al,* 2013; Ingolia *et al,* 2009; Ruiz-Orera *et al,* 2014). Additionally, mass spectrometry analyses of peptides in two cell lines revealed 69 of 9,640 so-called lncRNAs encode detectable peptides (Banfai *et al,* 2012; Derrien *et al,* 2012). Proteomic analysis of male germ cells in rat also identified peptide sequences derived from previously annotated lncRNAs (Chocu *et al,* 2014). It is not clear if these peptides are functional or represent

translational noise; regardless, the existence of a functional reading frame within an RNA does not exclude a non-coding function as well. As described below, SRY, *SRA* and oskar RNAs have both coding and non-coding functions.

It is likely that lncRNA classifications will be refined and subtypes of lncRNAs identified (Tuck and Tollervey 2013). Distinctions may be made according to lncRNA interacting partners, their functioning in cis vs. trans, whether they influence chromatin modification or organizational states, if activities are cytoplasmic or nuclear, lncRNA structural properties, or the kinds of sequences from which they originate. One example of the latter classification is represented by eRNAs, which are ncRNAs derived from enhancers, are typically less than 2kbp in length, and operate in cis (De Santa *et al,* 2010; Kim *et al,* 2010). They have been shown to regulate transcript elongation by interacting with mediator complex (Lai *et al,* 2013), and recruiting NELF from RNA pol II pause sites (Schaukowitch *et al,* 2014); and they also affect chromatin looping associated with enhancer function (Pefanis *et al,* 2015). eRNAs also regulate nucleosome remodeling processes (Mousavi *et al,* 2013). Impairing eRNA accumulation can attenuate enhancer activity (Lam *et al,* 2013). lncRNAs ncRNA-a3,4,5 and 7 themselves have enhancer like functions (Orom *et al,* 2010). Other eRNAs, such as *LED* (Leveille *et al,* 2015) and *LUNAR1* (Trimarchi *et al,* 2014), augment enhancer activity in *trans* by mechanisms that include recruiting mediator and RNA Pol II to enhancers. Additional examples of this classification include the extra

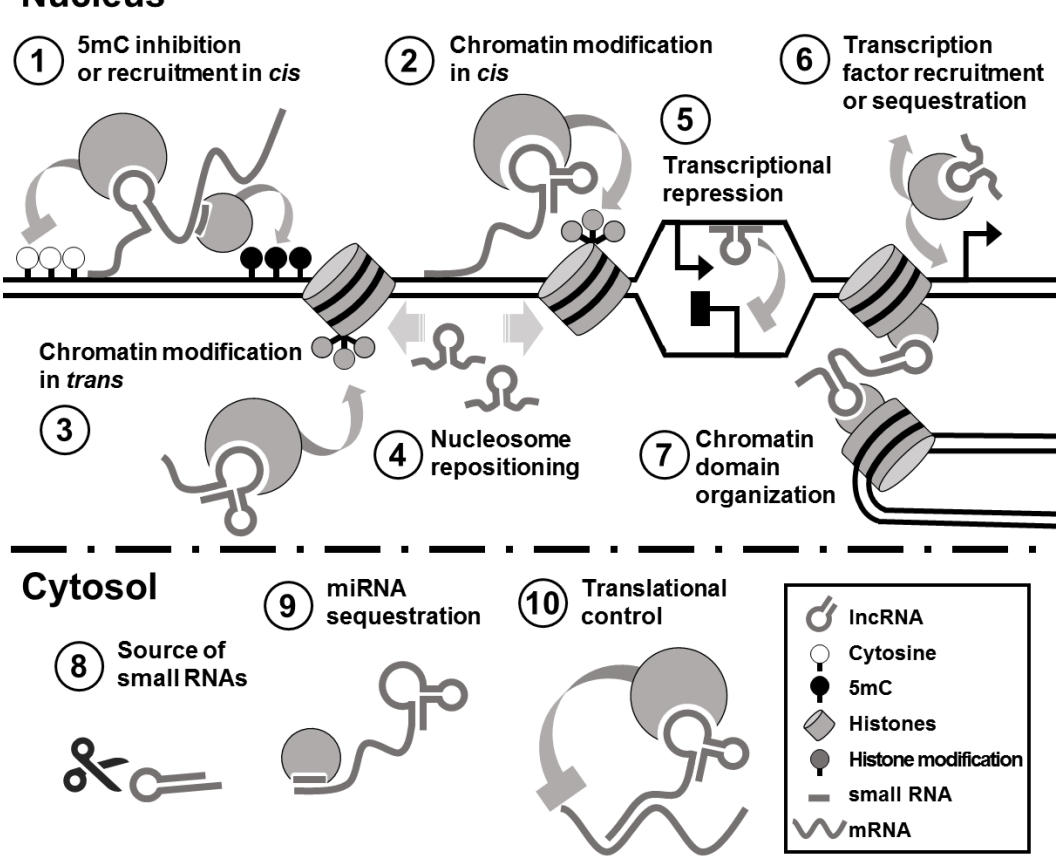
coding RNAs (ecRNA), which are Poly(A)– RNAs that extend beyond the gene bodies of coding genes (Di Ruscio *et al,* 2013), and pancRNAs that are promoter-associated (Hamazaki *et al,* 2015). The following four sections provide details of various molecular processes controlled by lncRNAs, which are summarized in **Table IV.1** and **Fig IV.1**.

**Table IV.1 Processes controlled by lncRNAs.**

| **Nuclear Processes** | **Examples** | **References** |
|---|---|---|
| DNA methylation inhibition in *cis* | ecRNA, CpG island R-loop RNAs | (Di Ruscio *et al,* 2013; Ginno *et al,* 2012) |
| DNA methylation recruitment in *cis* | rDNA pancRNA, pitRNA | (Hamazaki *et al,* 2015; Schmitz *et al,* 2010; Watanabe *et al,* 2011) |
| Histone modification recruitment in *cis* | *Xist, Airn*, *HOTTIP* | (Nagano *et al,* 2008; Sleutels *et al,* 2002; Wang *et al,* 2011; Zhao *et al,* 2008) |
| Histone modification recruitment in *trans* | *HOTAIR* | (Rinn *et al,* 2007) |
| Histone modification state maintenance | *Firre* | (Yang *et al,* 2015) |
| Nucleosome positioning | *SCHLAP1*, *Myheart, eRNAs* | (Han *et al,* 2014; Mousavi *et al,* 2013; Prensner *et al,* 2013) |
| Transcription interference | *Airn* | (Latos *et al,* 2012) |
| Transcription factor sequestration | *Gas5, PANDA* | (Hung *et al,* 2011; Kino *et al,* 2010) |
| Transcription factor recruitment | *BCAR4* | (Xing *et al,* 2014) |
| Organize chromatin domains and Nuclear bodies | *HOTTIP, Firre, MALAT1* | (Tripathi *et al,* 2010; Wang *et al,* 2011; Yang *et al,* 2015) |
| Enhancer control | *ncRNA-a3,4,5,7,* eRNAs, *LED, LUNAR* | (Lam *et al,* 2013; Leveille *et al,* 2015; Orom *et al,* 2010; Trimarchi *et al,* 2014) |
| Histone variant recruitment | centromeric lncRNA | (Quenet and Dalal 2014) |
| Splicing control | *MALAT1*, *FGF2R* antisense lncRNA | (Gonzalez *et al,* 2015; Tripathi *et al,* 2010) |
| **Cytoplasmic Processes** | **Examples** | **References** |

| | | |
|---|---|---|
| Source of miRNAs | *H19* | (Gao *et al,* 2012; Keniry *et al,* 2012) |
| miRNA sequestration | ceRNAs: *HULC, linc-MD1, lincRNA-RoR*, *H19;* circRNAs | (Cesana *et al,* 2011; Franco-Zorrilla *et al,* 2007; Hansen *et al,* 2013; Kallen *et al,* 2013; Wang *et al,* 2010; Wang *et al,* 2013) |
| Translation control | *Uchl1* antisense lncRNA, *lincRNA-p21* | (Carrieri *et al,* 2012; Yoon *et al,* 2012) |
| RNA stability control | *TINCR* | (Kretz *et al,* 2013) |
| Source of siRNAs and piRNAs | *pitRNA*, centromeric transcripts | (Hall *et al,* 2002; Volpe *et al,* 2002; Watanabe *et al,* 2011) |

**Fig IV.I. Mechanisms for lncRNA-mediate control of epigenetic state.**




## IV.C lncRNA control of histone states

A common theme with lncRNAs is their regulation of chromatin states, including histone and DNA modifications, nucleosome positioning and

placement of histone variants. The lncRNA *HOTAIR* that silences HOXD in trans binds PRC2, the major H3K27 histone methyltransferase containing complex, and is needed for deposition of H3K27me3 at HOXD (Rinn *et al,* 2007). The RepA lncRNA encoded within *Xist*, as well as *Xist* itself, bind PRC2, which is necessary for deposition of H3K27me3 on the inactive X-chromosome (Zhao *et al,* 2008). Because PRC2 binds RNA promiscuously (Davidovich *et al,* 2013; Kaneko *et al,* 2013), it was not clear if these interactions were specific. However, in RNA immunoprecipitation experiments only 20% of ~3,300 lncRNAs queried were observed to bind PRC2 (Khalil *et al,* 2009), indicating some specificity exists. Subsequent studies identified lncRNAs with high affinity for PRC2 (Davidovich *et al,* 2015; Herzog *et al,* 2014). There are additional determinants of specificity for PRC2 binding; for example the *Xist*-PRC2 interaction is partly regulated by the nucleosome remodeler ATRX (Sarma *et al,* 2014). Interestingly, maintenance of H3K27me3 on the inactive X-chromosome enabled by *Xist* requires an additional lncRNA. PRC2 is a heterogeneous complex (Margueron and Reinberg 2011) and various components were found to recruit it to lncRNAs. PRC2 binding to lncRNAs can occur through its component proteins JARID2 (Kaneko *et al,* 2014) or EZH2 (Kaneko *et al,* 2014; Zhao *et al,* 2008).

Besides binding PRC2, lncRNAs bind a variety of writers, erasers and readers of histone modifications, as well as other chromatin regulatory factors. In many cases, a given lncRNA can bind multiple chromatin regulatory factors (Guttman *et al,* 2011), though it is not yet known which binding events occur

simultaneously. Early findings reported Airn (Nagano *et al,* 2008) and Kcnq1ot1 (Pandey *et al,* 2008) bind the H3K9 methyltransferase EHMT2/G9a, with Kcnq1ot1 also binding PRC2 (Pandey *et al,* 2008). By binding EHMT2/G9a, Airn is able to direct it to the linked Slc22a3 promoter, and silence it (Nagano *et al,* 2008; Sleutels *et al,* 2002). *HOTTIP*, expressed from the HOXA locus, is brought into proximity to other sites in the HOXA cluster by looping, and through its recruitment of WRD5-containing MLL complexes, it promotes H3K4me3 deposition and gene transcription within the HOXA cluster (Wang *et al,* 2011). *HOTAIR* binds PRC2, and the lysine-specific demethylase KDM1A/LSD1 through distinct domains (Tsai *et al,* 2010). It is possible that these two factors are coordinated functionally, with KDM1A/LSD1 removing activating marks on H3K4 and PRC2 placing silencing marks on H3K27.

In addition to promoting chromatin modifications, or their removal, lncRNAs can restrict them to specific domains. A lncRNA from a pericentromeric region in *S. pombe* limits the spreading of H3K9me3 and binding of HP1, a reader of H3K9me3, beyond the centromeric region (Keller *et al,* 2013), and centromeric transcripts are the sources of siRNAs required for local placement of H3K9me3 (Hall *et al,* 2002; Volpe *et al,* 2002). Furthermore, in addition to binding PRC2, *Xist* binds SHARP/SPEN, HNRNPU/SAF-A and LBR, three factors necessary for X-inactivation. SHARP/SPEN interacts with the SMRT/NCOR2 co-repressor that activates the histone deacetylase HDAC3, which is likely to participate in restricting silencing to the inactive X-chromosome (McHugh *et al,* 2015).

Beyond their influence on covalent modifications to histone proteins, lncRNAs can control nucleosome position and placement of histone variants. The lncRNA *SCHLAP1* controls the localization and activity of SNF5/SMARCB1, a component of the SWI/SNF complex that repositions nucleosomes in an ATP-dependent manner (Prensner *et al,* 2013). A heart specific lncRNA, *MHRT*, interacts with the nucleosome remodeler SMARC4A/BRG1, restricting its activity at target sites (Han *et al,* 2014). Additionally, a centromeric lncRNA interacts with and is necessary for recruitment of the centromeric H3 variant, CENPA and its chaperone HJURP to human centromeres (Quenet and Dalal 2014).

IV.D lncRNA control of DNA methylation states

In addition to their effects on histone states, lncRNAs have been described that provide signals for deposition of DNA methylation in cis. At the rDNA locus, a promoter-spanning antisense lncRNA forms an R-loop, a triplex structure between double stranded DNA and a hybridized RNA, which recruits DNMT3B to the locus, leading to local methylation and rDNA silencing (Schmitz *et al,* 2010). At the imprinted *Rasgrf1* locus, pitRNA, a lncRNA spanning the domain carrying the methylation imprint, is required for local DNA methylation (Watanabe *et al,* 2011). This transcript normally functions in cis (Park *et al,* 2012); however, when expression patterns were perturbed, DNA methylation occurred in trans at the homologous locus (Herman *et al,* 2003). lncRNAs can also behave in the opposite manner, preventing deposition of DNA methylation. In contrast to their role in placing DNA

methylation at the rDNA locus, R-loops that form at CpG islands of other promoters have been implicated in preventing CpG island methylation (Ginno *et al,* 2012). The Poly(A)– ecRNA that extends across the CEBPA locus binds to the DNA methytransferase, DNMT1. This binding sequesters DNMT1, limiting DNA methylation of the transcribed locus, and enabling expression of the coding form of the Poly(A)+ mRNA (Di Ruscio *et al,* 2013). In RNA immunoprecipitation studies using antibody against DNMT1, many other Poly(A)– transcripts were isolated, suggesting ecRNA control of DNMT1 might be commonplace. The domains from which ecRNAs are transcribed tend to harbor less methylation and have more transcription relative to domains producing Poly(A)– transcripts that are unbound to DNMT1 (Di Ruscio *et al,* 2013), observations consistent with the notion that DNMT1 sequestration by ecRNAs frequently limits DNA methylation.

IV.E lncRNA control of transcriptional states

By influencing local chromatin states, lncRNAs can modify gene expression states. However, lncRNAs can affect transcription through means that are distinct from their effects on chromatin. For example, by binding transcription factors, lncRNAs such as *GAS5* (Kino *et al,* 2010) and *PANDA* (Hung *et al,* 2011) can limit factor access to their DNA targets; though the lncRNA BCAR4, can enable transcription factor recruitment to DNA (Xing *et al,* 2014). In some cases, such transcription factor interactions are sensitive to extracellular signaling molecules (Trimarchi *et al,* 2014; Xing *et al,* 2014) or also involve histone modifiers that affect local chromatin states (Wang *et al,*

2008; Xing *et al,* 2014). In addition to binding and recruiting EHMT2/G9a to some target sites, *Airn*, the lncRNA that regulates Igf2r imprinted expression (Sleutels *et al,* 2002), exerts this effect by transcriptional interference at the silenced paternal *Igf2r* allele (Latos *et al,* 2012). At other loci, transcriptional interference by lncRNAs appears sufficient to control local gene expression (Latos *et al,* 2012; Martianov *et al,* 2007; Santoro *et al,* 2013).

In addition to establishing chromatin and gene expression states, lncRNAs may respond to these states. For example, the lncRNAs *NEAT1* and *MALAT1* bind chromatin, but exhibited novel patterns of chromatin localization after treatment of cells with the transcription elongation inhibitor flavopiridol. This demonstrated that lncRNA localization can respond to RNA polymerase II activity, in addition to influencing that activity (West *et al,* 2014). Such results might reflect splicing choices influenced by lncRNAs (described further below); alternatively lncRNAs might maintain chromatin or expression states, once they are established. This is the case for the lncRNA *Firre*, which is required to maintain previously established H3K27me3 on the inactive X-chromosome (Yang *et al,* 2015).

IV.F lncRNA control of other functions

lncRNAs influence a variety of other cellular functions beyond control of chromatin and transcriptional states, including nuclear architecture, splicing, and mRNA translation by multiple mechanisms. In addition to its role in maintaining H3K27me3 on the inactive X-chromosome, *Firre* is also required for nucleolar localization of the inactive X-chromosome in mammals. This

occurs through a mechanism involving its own interaction with CTCF (Yang *et al,* 2015). *Firre* also binds the nuclear matrix factor hnRNPU through a sequence repeated within the lncRNA, and localizes to distinct regions in the genome in a manner dependent on hnRNPU expression (Hacisuleyman *et al,* 2014).

The lncRNA *MALAT1* binds SR splicing factors and influences their localization in nuclear speckles (Tripathi *et al,* 2010). *MALAT1* also interacts with pre-mRNAs (Engreitz *et al,* 2014) and its depletion causes changes in alternative splicing (Tripathi *et al,* 2010). An additional example of splicing control by lncRNAs involves intron-encoded lncRNAs that are processed by snoRNA-dependent mechanisms. These so-called sno-lncRNAs influence splicing by their association with Fox family splicing factors (Yin *et al,* 2012). An antisense lncRNA from the human *FGFR2* locus controls local alternative splicing choices by affecting local histone methylation state (Gonzalez *et al,* 2015).

A lncRNA corresponding to an antisense transcript from the coding gene *Uchl1* regulates Uchl1 translation (Carrieri *et al,* 2012). Interestingly, this lncRNA exerts its translational control through sequences with similarity to the SINEB2 repetitive element. lincRNA-p21 also regulates translation of specific transcripts, likely by a mechanism that involves physical interaction with its targets (Yoon *et al,* 2012). Though not controlling translation, the lncRNA TINCR can also affect protein levels after transcription and splicing by regulating mRNA stability (Kretz *et al,* 2013).

lncRNA effects can involve functional interactions with other regulatory ncRNAs. miRNAs can be sequestered by lncRNAs that were referred to as competing endogenous RNAs (ceRNAs), some of which are circular (circRNA) (Hansen *et al,* 2013; Memczak *et al,* 2013; Tay *et al,* 2014), and these limit the capacity of miRNAs to regulate translation of their mRNA targets. This was originally reported in *Arabidopsis* (Franco-Zorrilla *et al,* 2007), and also was found in mammalian liver cancer cells (Wang *et al,* 2010), myoblasts (Cesana *et al,* 2011), and embryonic stem cells (Wang *et al,* 2013). Interestingly, *H19* was shown to act in this capacity as well (Kallen *et al,* 2013), while also being a precursor for distinct miRNAs (Gao *et al,* 2012; Keniry *et al,* 2012). This is in addition to *H19*'s ability to bind the methylated DNA binding factor MBD1 and regulate other imprinted genes, both in *cis* and *trans* (Monnier *et al,* 2013). Also, miR-9 can target the lncRNA *MALAT1* for degradation (Leucci *et al,* 2013). Beyond miRNAs, the piRNA pathway is necessary for lncRNA-mediated control of DNA methylation at the imprinted locus *Rasgrf1* (Watanabe *et al,* 2011).

The examples of genomic regulation provided here are simply illustrative of known lncRNA activities; additional studies are likely to reveal further activities, and the necessity of individual lncRNAs for physiological processes in vivo. In one study of mice deficient for 18 lncRNAs, three were found to be essential for viability, and two affected growth (Sauvageau *et al,* 2013). Important questions, whose answers are beginning to emerge, include the following: By what mechanisms are the lncRNAs themselves regulated

(Amin *et al,* 2015), what are the details of the mechanisms by which lncRNAs exert their effects, and what health and disease relevant phenotypes are controlled by lncRNAs? Understanding mechanisms of lncRNA action will require further knowledge of lncRNA structure and its impact on function (Brown *et al,* 2014a; Somarowthu *et al,* 2015); interacting factors including proteins, other RNAs, and possibly metabolites; factors controlling lncRNA subcellular localization; and for chromatin based phenomena, understanding how lncRNAs localize to, and/or restrict their activities at specific genomic domains. For *cis* acting lncRNAs, many are likely to function co-transcriptionally while still tethered to their DNA template. How trans acting lncRNAs become targeted to and act at specific loci is less clear. For *Xist*, YY1 is important for recruiting *Xist* to the inactive X-chromosome (Jeon and Lee 2011), but it is not clear how *Xist* is excluded from other genomic locations. This might involve licensing enabled by X-chromosome pairing prior to X-inactivation (Xu *et al,* 2006), a mechanism that might be limited to X-inactivation. *HOTAIR* has many binding sites in the genome that are enriched for a GA-rich DNA motif, indicating DNA sequence specific binding factors might recruit the RNA (Chu *et al,* 2011). In human cells, transcriptional targets of the lncRNA *ANRIL/CDKN2B-AS1* both contain and require Alu elements for their *ANRIL/CDKN2B-AS1* responses. Not all Alu elements respond to *ANRIL/CDKN2B-AS1*, and it is not clear what provides specificity for Alu elements at *ANRIL/CDKN2B-AS1* target genes (Holdt *et al,* 2013). Additional issues requiring more study include the functional importance, if any, of post

transcriptional modifications to lncRNAs (Batista *et al,* 2014; Fu *et al,* 2014; Kiani *et al,* 2013; Liu *et al,* 2015a; Schwartz *et al,* 2014; Wang and He 2014; Zheng *et al,* 2013b), and the roles transposable elements have played in lncRNA diversity (Kapusta *et al,* 2013; Kelley and Rinn 2012; Liang *et al,* 2012). Approaches that systematically characterize proteins bound to specific lncRNAs will continue to be informative (McHugh *et al,* 2015).

Many excellent and recent reviews describe the discovery, cataloging, and activities controlled by lncRNAs, as well as approaches towards functional analysis (Batista and Chang 2013; Cech and Steitz 2014; Chu *et al,* 2015; Engreitz *et al,* 2015; Fatica and Bozzoni 2014; Flynn and Chang 2014; Geisler and Coller 2013; Ghosal *et al,* 2013; Holoch and Moazed 2015; Hu *et al,* 2012; Iyer *et al,* 2015; Morris and Mattick 2014; Quinodoz and Guttman 2014; Sun and Kraus 2013; Troy and Sharpless 2012; Ulitsky and Bartel 2013; Wang and Chang 2011). Additional reviews focus on mechanisms of lncRNA control of sex chromosome dosage compensation (Autuoro *et al*, 2014; Briggs and Reijo Pera 2014; Deng *et al,* 2014; Galupa and Heard 2015; Lee and Bartolomei 2013), and control of development including stem cell maintenance and differentiation (Batista and Chang 2013; Fatica and Bozzoni 2014; Flynn and Chang 2014; Ghosal *et al*, 2013; Yao and Jin 2014). Given the diverse roles lncRNAs play in essential biological processes common to many cell types, it should come as no surprise that lncRNAs play a vital role in reproduction, which is the focus of the remainder of this review. Though many ongoing studies are descriptive, functional and mechanistic studies exist and will be

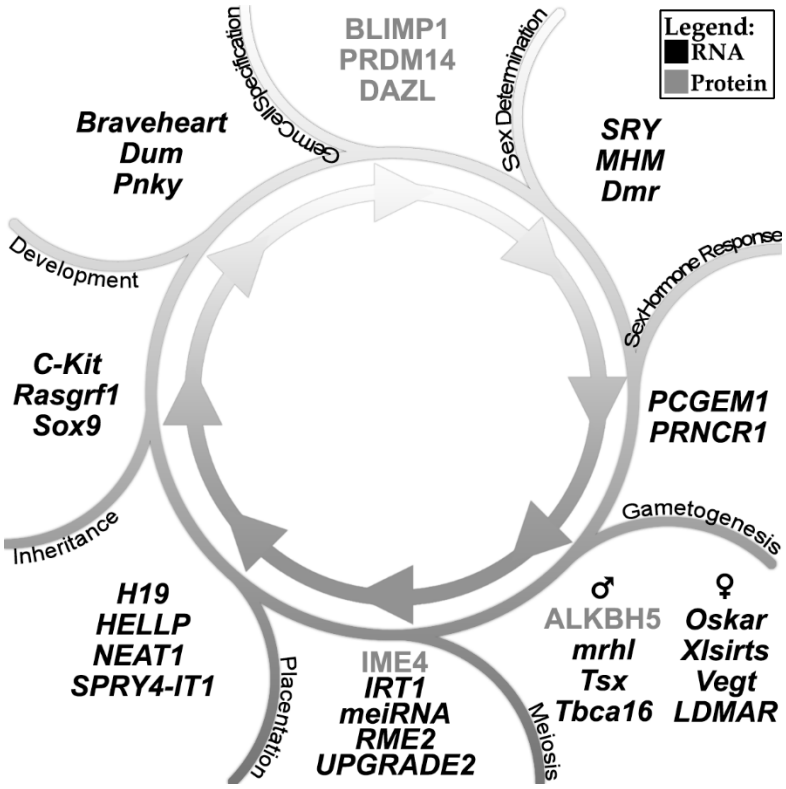
highlighted; these are summarized in **Table IV.2** and **Fig IV.2**. Studies from a diversity of species will be presented as they are informative for evolutionarily-conserved processes.

**Table IV.2. lncRNAs with identified functions in reproduction**

| Reproductive process | lncRNAs involved | Mechanisms | References |
|---|---|---|---|
| Gonadogenesis | *Sxl* promoter proximal RNAs | Activates *Sxl* and recruits chromatin modifiers like Polycomb and Trithorax. | (Mulvey *et al,* 2014) |
| Sex determination | *SRY* | Acts as a miRNA sponge that competitively binds miR-138 to potentially influence sex determination. | (Hansen *et al,* 2013) |
| | *Dmr* | Regulates splicing of *DMRT1.* | (Zhang *et al,* 2010) |
| Sex hormone responses | eRNAs from ER-bound enhancers | Controls ER regulated enhancer activity. | (Li *et al,* 2013) |
| | *PCGEM1* | Binds AR, and lncRNA *PRNCR1*; alters chromatin topology affecting androgen responses. | (Yang *et al,* 2013) |
| | *PRNCR1* | Binds AR, and methyltransferase DOT1L; alters AR modification state and chromatin topology affecting androgen responses. | (Yang *et al,* 2013) |
| | eRNAs from *KLK3* enhancer cluster | Recruits mediator and AR to *KLK3* and *KLK2.* | (Hsieh *et al,* 2014) |
| | *SRA* | Binds to and modulates activity of AR, ER, and PR; binds additional transcriptional regulators. | (Hatchell *et al,* 2006; Lanz *et al,* 1999; Lanz *et al,* 2002; Shi *et al,* 2001; Watanabe *et al,* 2001) |
| Meiosis | *UPGRADE2* | Expressed specifically in *Boechera* species capable of asexual reproduction. | (Mau *et al,* 2013) |
| | *IRT1* | Increases nucleosome occupancy and repressive modifications to silence the meiotic regulator *IME1.* | (van Werven *et al,* 2012) |

| | | | |
|---|---|---|---|
| | *RME2* | Silences the oppositely transcribed meiotic regulator *IME4*, possibly by transcriptional interference. | (Gelfand *et al,* 2011; Hongay *et al,* 2006; van Werven *et al,* 2012). |
| | *meiRNA* | Regulates nuclear import of the meiosis regulator Mei2, and formation of nuclear dot where Mmi1, which is capable of degrading meiosis regulating transcripts, is sequestered. | (Harigaya *et al,* 2006; Shichino *et al,* 2014; Yamashita *et al,* 1998) |
| Spermatogenesis | *mhrl* | Regulates Wnt signaling in spermatogonial cells. | (Arun *et al,* 2012) |
| | *Tsx* | Suppresses apoptosis of pachytene spermatocytes. | (Anguera *et al,* 2011) |
| | *LDMAR* | Required for anther development in rice. | (Ding *et al,* 2012) |
| Oogenesis | *oskar* | Non-coding functions of *oskar* mRNA required for early establishment of oocyte polarity. | (Jenny *et al,* 2006) |
| | *XIsirt* and *VegT* | Required for cytoskeletal organization and oocyte polarity in *Xenopus.* | (Kloc *et al,* 2005) |
| | *XLOC_057324* | Regulates flowering time and seed setting in rice. | (Zhang *et al,* 2014) |
| Placentation | *H19* | Controls placental growth regulated by NOMO1 by encoding miR-675 that controls *NOMO1* translation. | (Gao *et al,* 2012; Keniry *et al,* 2012) |
| | *HELLP* | Controls cell survival and migration of trophoblast cells. | (van Dijk *et al,* 2012) |
| | *SPRY4-IT1* | Controls cell survival and migration of trophoblast cells. | (Zou *et al,* 2013) |
| | *Airn* | Regulates imprinted expression of *Igf2r*, *Slc22a2* and *Slc22a3* which influence placental growth. | (Nagano *et al,* 2008; Zwart *et al,* 2001) |
| | *Kcnq1ot1* | Regulates imprinted expression of eight linked genes which influence placental growth. | (Pandey *et al,* 2008) |
| Reproductive disease | *NEAT1* | Required for corpus luteum formation and pregnancy maintenance | (Nakagawa *et al,* 2014) |

**Fig IV.2. Compilation of lncRNAs and proteins involved in reproductive processes.**




## IV.G Germ cell specification

Formation of the animal germ line begins with specification of primordial germ cells (PGCs), the pluripotent cells that derive from a cluster of cells posterior to the definitive primitive streak in the extraembryonic mesoderm of mid-primitive-streak stage embryos, 7-7.5 days post coitum (dpc) in mouse. They later migrate along the genital ridge where they contribute to the developing gonad (Ginsburg *et al,* 1990). Three proteins required for PGC specification include BLIMP/PRDM1, TFAP2C/AP2γ, and PRDM14 (Magnusdottir *et al,* 2013). In PGCs, BLIMP/PRDM1 represses genes involved in somatic development, in agreement with its prior characterization as a transcriptional repressor (Gyory *et al,* 2003; Keller and Maniatis 1991). However, BLIMP/PRDM1 is also bound near other genes that are activated

and important for PGC specification such as Tcfap2c/Ap2γ) and Cbx7. Among the 5,046 BLIMP/PRDM1 binding sites in mouse PGCs are 313 associated with non-coding genes (Magnusdottir *et al,* 2013), whose functions in PGC specification are unknown. Given that lncRNAs are known to directly repress transcription, it is possible that BLIMP/PRDM1 activates targets indirectly by negatively regulating uncharacterized repressive lncRNAs. Further profiling of germ line lncRNAs, and interacting factors, could help resolve this.

Another set of proteins important for PGC specification and meiotic progression are the RNA binding proteins DAZL, DAZ1, and BOLL/BOULE (Kee *et al,* 2009). These bind RNA in the cytoplasm, regulating translational initiation (Collier *et al,* 2005); however, they have also been shown to translocate into and out of the nucleus during germ line development (Reijo *et al,* 2000). During their nuclear phase, these proteins might functionally regulate coding and/or non-coding RNAs important for PGC formation or differentiation. This could be addressed by identification and functional characterization of RNAs associated with DAZ family members at different time points during germ line differentiation.

lncRNAs may be involved in PGC specification by the variety of mechanisms cited above. By way of example, Braveheart/Bvht is a lncRNA that controls the expression of MesP1, a master regulator controlling differentiation of mesodermal precursors into cardiomyocytes (Klattenhoff *et al,* 2013). Using this system as a model, it is possible to envision that master regulators of PGC fate like BLIMP/PRDM1 or DAZL may also be under the

control of an unknown lncRNA. lncRNA regulation might be active at every level of PGC development, both as an initiator and as a downstream response element.

IV.H Sex determination and gonadogenesis

Several lines of evidence suggest the importance of lncRNAs for sex determination and sex-specific patterns of development. The number or identity of sex chromosomes is typically the genetic determinant of sex; however, there are cases of lncRNA involvement in this process as well. *Drosophila* become female upon the early expression of the X-encoded Sex-lethal (*Sxl*) gene. The expression of this gene is dependent upon the ratio of X chromosomes to autosomes (A) by a complex chromosome counting mechanism that requires the competition of gene products from each chromosome. If X≥A, then the X transcription factors Sisterless-a and Sisterless-b activate *Sxl* expression; if X<A, then proteins including Deadpan and Extramacrochaetae directly or indirectly block binding to *Sxl* promoter, resulting in male determination (Schutt and Nothiger 2000). Interestingly, expression of a panoply of lncRNAs located ~1kb upstream of the promoter add to the complexity of this counting mechanism, with RNAs from the R1 region causing repression, and RNAs from the R2 region causing activation of *Sxl* (Mulvey *et al,* 2014). These RNA species were also shown to recruit chromatin modifiers like Polycomb and Trithorax, indicating that *Sxl* expression is regulated by a complex interaction network involving many lncRNAs.

In mice, expression of the Y-chromosome encoded *Sry* gene is sufficient to drive male sex determination (Koopman *et al,* 1991). SRY protein activates transcriptional cascades specific for male development (Kashimada and Koopman 2010), but its RNA independently functions as a circRNA miRNA sponge that competitively binds miR-138 *in vitro,* potentially contributing to its ability to positively regulate male specification (Hansen *et al,* 2013).

The DMRT1 protein has been implicated in sex determination in a variety of vertebrate and invertebrate species, including human, and acts both as a transcriptional repressor and activator (reviewed in Matson and Zarkower 2012)). In mice, the *Dmrt1* transcript participates in *trans* splicing with the lncRNA *Dmr*, producing a transcript that encodes a protein with an altered C-terminus. Overexpressing *Dmr* in primary Sertoli cell cultures increased the abundance of the altered form of DMRT1 protein, reduced the abundance of the cannonical DMRT1 isoform, and led to impaired expression of DMRT1 target genes, mimicking the DMRT1 loss of function phenotype. It is clear that *trans* splicing negatively regulated DMRT1; what is unclear is whether the noncannonical isoform has its own regulatory activity. Interestingly, reporters carrying a 3' UTR from *Dmr* exhibit enhanced expression (Zhang *et al,* 2010). It is not known if these contrasting results are due to idiosyncrasies of the specific assay system, or if they reflect the range of regulation controlled by *Dmr*.

## IV.I Sex hormone responses

The study of lncRNA activity in the context of sex hormone response has largely been restricted to the steroid sex hormones, which utilize nuclear receptors (NRs), including the estrogen receptor (ESR1), androgen receptor (AR), and progesterone receptor (PGR). Roles for lncRNAs have not been demonstrated for signaling by follicle stimulating hormone and luteinizing hormone, which utilize G-protein coupled receptors and cytosolic signal transduction cascades. However, given the cytosolic localization and activity of many lncRNAs, such roles are plausible.

Functional studies of ncRNA involvement in AR and ER responses include roles for eRNAs and other lncRNAs. For example, eRNAs transcribed from ER-bound enhancers are known to recruit transcriptional activators to drive expression of nearby ER-responsive genes. Loss of the eRNAs by siRNA knockdown led to reduced transcription of target genes without affecting ER recruitment; moreover, tethering specific eRNAs to a reporter gene enabled reporter activation (Li *et al,* 2013). In a similar manner, the lncRNAs *PCGEM1* and *PRNCR1* associate with the AR. Knockdown of *PCGEM1* or *PRNCR1* reduced transcription of a number of canonical AR-targeted genes. This was followed up with chromatin conformation capture (3C), where it was demonstrated that enhancer-promoter interactions were reduced in the absence of *PCGEM1* and *PRNCR1* (Yang *et al,* 2013). At a single AR-driven locus, Kallikrein-related peptidase 3 (*KLK3*), an enhancer cluster 4 kb upstream of the *KLK3* promoter binds AR in a hormone-

dependent manner (Hsieh *et al,* 2014). In this case the eRNA acts as part of a scaffolding apparatus that includes Mediator and AR to enhance transcriptional activity at the endogenous *KLK3* locus and at the downstream *KLK2* locus. Additional AR and ER target genes might also be influenced by lncRNAs.

The lncRNA *SRA* was originally identified in a yeast two hybrid screen for human PR interacting factors, indicating it has a functional reading frame. However, its steroid hormone receptor activator activity does not require its translation, or an open reading frame (Lanz *et al,* 1999; Lanz *et al,* 2002). *SRA* has since been shown to modulate the activity of AR, ESR1, and PGR through direct association with the hormone receptors (Lanz *et al,* 1999) and through recruitment of a variety of transcriptional activators and repressors (Hatchell *et al,* 2006; Shi *et al,* 2001; Watanabe *et al,* 2001). This lncRNA has another unique property in that it can be spliced and translated into *SRA* Protein (SRAP/SRA1) which also enhances steroid hormone-mediated gene expression (Kawashima *et al,* 2003).

In addition to acting as a transcriptional coactivator, *SRA* has recently been shown to associate with a repressive histone modifying complex containing unliganded PGR (uPGR), and chromatin binding and modifying factors including CBX5/HP1, KDM1A/LSD1, HDAC1 and 2, and RCOR1/CoREST in breast cancer cells. uPGR localizes this complex to approximately 20% of steroid-responsive genomic loci. Depletion of *SRA* led to destabilization of the complex and aberrant gene expression

patterns. Upon progesterone treatment, the repressive complex is evicted and replaced by ligand-bound PR and basal transcription factors (Vicent *et al,* 2013).

While *SRA* is known to enhance the activity of various steroid receptors including, but not limited to the steroid sex hormones, the regulation of *SRA* transcription itself is unknown. Similarly, the factors that determine its differential splicing to become a transcript coding for SRAP/SRA1 are also unknown.

### IV.J Meiosis

lncRNAs have been implicated in control of meiosis from studies in both plants and yeasts. In several plant species, seeds can form asexually through a variety of processes collectively referred to as apomixis (Koltunow and Grossniklaus 2003). Shared features include female gamete formation in the absence of recombination or reductive division seen normally in meiosis, followed by parthenogenic embryo development in the absence of fertilization. The resulting plants harbor their maternal genotype. In several apomictic species of the genus *Boechera*, microarray analyses identified a conserved lncRNA, UPGRADE2, that is present and highly upregulated in pollen mother cells. No homolog was found in sexually reproducing species of the same genus (Mau *et al,* 2013). It remains to be determined if this lncRNA is simply associated with or required for apomixis in *Boechera*.

In the budding yeast, *Saccharomyces cerevisiae*, *IME1* (inducer of meiosis) is kept transcriptionally silent by the repressor, RME1, in vegetative cells growing in a nutrient rich environment and in haploid cells encountering no partners of opposite mating type. RME1 induces expression of a lncRNA, *IRT1*, which spans the *IME1* promoter, and works in *cis* to increase local nucleosome occupancy, and to recruit the SET3 complex that deposits repressive histone modifications at the promoter (van Werven *et al,* 2012). Interestingly, many SET3 repressed genes have overlapping lncRNA transcripts (Kim *et al,* 2012).

Another inducer of meiosis, IME4 (Shah and Clancy 1992), is also regulated in *cis* by another lncRNA, *RME2*, which is transcribed antisense relative to *IME4* and might block its expression by transcriptional interference rather than by recruiting chromatin modifying factors (Gelfand *et al,* 2011; Hongay *et al,* 2006; van Werven *et al,* 2012). Antisense transcripts to these lncRNAs activated sporulation (van Werven *et al,* 2012). Interestingly, IME4 is a methyltransferase capable of placing the RNA modification, N6-methyladenine (m6A) (Agarwala *et al,* 2012); RNAs harboring this modification are less stable than those lacking it (Batista *et al,* 2014). The *Drosophila* ortholog of IME4, METTL3, is essential for gametogenesis and embryo viability (Hongay and Orr-Weaver 2011). These functions in *Drosophila* are mediated by the influence of IME4 on Notch signaling, though it is not clear how m6A modifications affect Notch signaling. Many RNAs with IME4-dependent m6A modifications have been described (Schwartz *et al,*

2013). Other than *RMA2*, other meiosis specific lncRNAs have been described in *S. cerevisiae*, raising the likelihood that additional lncRNA dependent mechanisms exist that regulate meiosis (Lardenois *et al,* 2011).

lncRNAs have also been implicated in meiosis control in the fission yeast *Schizosaccharomyces pombe.* Mei2 protein is the master regulator of meiosis in *S. pombe* (Watanabe *et al,* 1997; Watanabe and Yamamoto 1994). It is recruited to the nucleus by the lncRNA *meiRNA* (Shichino *et al,* 2014; Yamashita *et al,* 1998), forming a nuclear dot (Shimada *et al,* 2003) that includes the lncRNA, *Mei2* and *Mmi1*. This occurs at the *sme2* locus from which meiRNA is transcribed, identifying a *cis* acting function for meiRNA. The dot promotes meiosis by sequestering Mmi1, an RNA binding protein that degrades meiosis promoting transcripts (Harigaya *et al,* 2006). Degradation requires polyadenylation and involves nuclear exosomes. Interestingly, Mmi1, whose function is antagonized by meiRNA, is required for meiRNA recruitment to *sme2* (Shichino *et al,* 2014), suggesting that though meiRNA functions in *cis*, its localization involves *trans* acting factors.

Additional lines of evidence suggest, indirectly, other mechanisms that potentially involve lncRNA influences on meiosis. The example of m6A controlled by IME4 was cited above, however, other methyltransferases have been implicated in processes critical to RNA function in vertebrates, invertebrates, and plants (Li and Mason 2014; Schwartz *et al,* 2013; Zhong *et al,* 2008). More thorough characterization of both m6A modified RNAs and

lncRNAs influencing meiosis may identify the importance of m6A modifications to their action. Another potential mechanism influencing lncRNA action involves RNA binding proteins known to be important in mammalian meiosis, including DAZL and DDX4/VASA (Medrano *et al,* 2012). As mentioned previously, the DAZ family RNA binding proteins, which are required for PGC specification, are found in the nucleus and cytoplasm of fetal germ cells, in the cytoplasm of developing oocytes, and in the nucleus of spermatogonia. Their translocation between nucleus and cytoplasm during meiosis (reviewed in (Brook *et al,* 2009; Smorag *et al,* 2014)) implicates additional functions than just translational control (Collier *et al,* 2005). Two family members, DAZ and BOLL/BOULE, are required for later stages of meiosis (Kee *et al,* 2009). Immunoprecipitates of DAZL from rat testis homogenate contained many mRNAs, but these were also detected by microarray, which might have missed recently characterized lncRNAs. Associations between DAZ family proteins and lncRNAs might reveal important participants in meiosis control. DDX4/VASA is an RNA helicase that regulates mRNA translation and piRNA production (reviewed in Kotov *et al,* 2014). DDX4/VASA immunoprecipitates from mouse testicular cells contained 858 mRNAs (Nagamori *et al,* 2011). These were identified by microarrays designed to detect mRNAs. RNA-seq analysis would reveal more reliably if DDX4/VASA also binds and functionally regulates lncRNAs associated with meiosis. MOV10L1 is another RNA helicase that is expressed at increasing levels in germ cells between the gonocyte and pachytene spermatocyte

stages. It binds the PIWI proteins PIWIL1/MIWI and PIWIL2/MILI (Frost *et al,* 2010; Zheng *et al,* 2010) and piRNA precursor transcripts (Vourekas *et al,* 2015), which, formally, may be considered a class of lncRNAs. MOV10L1 is required for primary piRNA biogenesis (Vourekas *et al,* 2015; Zheng and Wang 2012; Zheng *et al,* 2010) and silencing retrotransposons in the male germ line (Frost *et al,* 2010). Male mice lacking MOV10L1, or carrying a point mutation in the ATP binding domain of the helicase exhibit meiotic arrest in prophase I (Frost *et al,* 2010; Vourekas *et al,* 2015; Zheng and Wang 2012; Zheng *et al,* 2010); females deficient for the protein are fertile. Special requirements for helicases during piRNA biogenesis may relate to G-quadruplex structures potentially present in precursor transcripts (Vourekas *et al,* 2015).

## IV.K Gametogenesis

### IV.K.1 Spermatogenesis

lncRNAs are dynamically expressed and appear to be highly regulated in spermatogenesis. Several studies have profiled the transcriptomes of the developing male germ line and have defined a clear pattern. First, transcript levels dramatically increase as spermatogonia enter meiosis. These increase further as spermatocytes give rise to spermatids. This is followed by a rapid depletion of RNA in spermatozoa (Bao *et al,* 2013; Chalmel *et al,* 2014; Laiho *et al,* 2013; Liang *et al,* 2014; Margolin *et al,* 2014; Soumillon *et al,* 2013). RNAs identified in these total RNA profiling studies include lncRNAs, however, few of these have been characterized functionally. Recent RNA-seq

profiling at different stages of spermatogenesis highlight examples of potential regulation of spermatogenesis by lncRNAs.

In array-based profiling of lncRNAs and mRNAs, stage-specific, differentially-expressed lncRNAs have been found within 30kb of coding gene clusters during spermatogenesis. Depending on the coding gene cluster, there were positive or negative correlations between lncRNA expression and local mRNA expression. The most pronounced changes in expression occurred after the onset of meiosis, with changes in lncRNA expression correlating with expression of nearby mRNA clusters. A subset of these lncRNAs were characterized via CLIP-QPCR (Bao *et al,* 2013); many were found to interact with EZH2 and KDM1A/LSD1 to potentially regulate nearby expression and methylation states. The coordinated change in expression of lncRNAs and corresponding gene clusters was also observed in an array-based profile (Liang *et al,* 2014). This is unsurprising in the context of the promiscuous binding of PRC2 (Davidovich *et al,* 2015); however, the physiological relevance of these data remains unknown and further characterization is required.

At birth, spermatogonia in mice express a comparably low fraction of the total testis lncRNA profile (Soumillon *et al,* 2013). An interesting transcript detected in this phase, *mhrl*, is a nuclear localized lncRNA that has been shown suppress the Wnt signaling pathway in a spermatogonial cell line by regulating beta-catenin nuclear translocation (Arun *et al,* 2012). Wnt signaling is a regulator of “stem cell-ness” and implicated in maintaining a self-renewing

population of spermatogonial stem cells (Golestaneh *et al,* 2009; Yeh *et al,* 2011). Though *mhrl*-mediated repression of Wnt signaling suggests it influences spermatocyte differentiation, its specific function needs to be explored by *in vivo* manipulations of *mhrl* expression.

Upon induction of meiosis there is a considerable increase in lncRNA transcription in mouse spermatocytes (Soumillon *et al,* 2013). In pachytene spermatocytes, *Tsx,* a predominantly nuclear testis specific X-linked RNA, becomes highly expressed and escapes X-inactivation (Anguera *et al,* 2011). A *Tsx* knockout produces viable and fertile offspring, however, males have decreased testis size and increased levels of apoptosis of pachytene spermatocytes. Interestingly, *Tsx* knockout mice showed deficiencies in learning and increased *Xist* expression. The nuclear localization and X-linked expression of *Tsx* suggests a role in X-inactivation in pachytene spermatocytes; however, this regulatory role of *Tsx* is not yet understood.

As previously mentioned, RNA modifications are relevant in gametogenesis. The importance of RNA methyltransferases such as IME4 and METTL3 were previously discussed in reference to gamete development. However, RNA demethylases such as ALKBH5 are also vital during this time, specifically at the pachytene stage of spermatogenesis (Zheng *et al,* 2013a). *Alkbh5* knockout mice exhibited decreased testis size, sterility, an increase of m6A-modified mRNAs, altered RNA localization, and significant changes in gene expression. The increased half-life of demethylated RNAs at this stage (Batista *et al,* 2014) may contribute to the

increased expression and overall abundance of ncRNAs in spermatocytes, potentially affecting the recruitment of other chromatin readers/writers such as PRC2 to loci. Alternatively, m6A may be regulating RNA-Protein interactions or affinities *via* altered RNA base pairing (Liu *et al,* 2015a). m6A-Seq has not yet been performed in developing testis.

Tubulin cofactor A chromosome 13 isoform (TBCA13), a protein involved in tubulin assembly, increases in expression from 14 days post partum (dpp) to a highly expressed state at 25 dpp in mouse testis. In a spermatocyte cell line, the transcription of *Tbca13* is regulated via a pseudogene, *Tbca16*, which is a duplication of *Tbca13* with both sense and antisense transcription on chromosome 16. The antisense product of *Tbca16* appears to negatively regulate *Tbca13*. When *Tbca16* mRNA was depleted with shRNA, *Tbca13* escaped silencing (Nolasco *et al,* 2012). The mechanism of the silencing of *Tbca16* and escape of *Tbca13* during spermatogenesis has not been elucidated; however, a related mechanism might be found at the 3' actin pause site where antisense transcription and R-loop formation recruits AGO2, EHMT2/G9a, and the repressive H3K9me2 mark to enhance mRNA termination (Skourti-Stathaki *et al,* 2014). The *in vivo* importance of this regulation and its involvement in spermatogenesis has not been further explored.

A majority of the transcriptome is depleted upon spermatozoa maturation. Most recently, the transcriptomes of the nucleus and periphery of mature spermatozoa were profiled (Johnson *et al,* 2015). These data indicate

that the majority of the RNA in a spermatozoon is localized to the cytoplasm, while a minority (roughly 34%) localizes to the nucleus. *Malat1* highlights the potential for lncRNA-mediated chromatin organization in the male germ line – despite the expulsion of most RNAs, it is enriched in the sperm nucleus. However, *Malat1* knockouts do not exhibit defects in fertility, underlining the fact that its function at this stage is unclear (Zhang *et al,* 2012). Several RNAs are present in mature spermatozoa that are not present in unfertilized oocytes and are delivered to the zygote upon fertilization (Johnson *et al,* 2015). It is currently unclear if spermatozoon localized lncRNAs are vital for gamete formation or zygotic function after fertilization.

IV.K.2 Oogenesis

The developing mammalian oocyte exists in a cumulus-oocyte complex (COC), with cumulus cells (CCs) forming a network of cells surrounding the oocyte in close contact through gap junctions. Early in oogenesis, CCs form a compact layer around immature oocytes, which are arrested at Prophase I. Surges of FSH/LH at ovulation cause the COC to expand and detach from the follicle wall, coincident with the oocyte resuming meiosis (Yokoo and Sato 2004). Though the transcriptional change in CCs before and after the COC expansion is considerable, a small number of lncRNAs were detected as differentially expressed by RNA-Seq (Yerushalmi *et al,* 2014). Ninety-six non-coding RNAs, 45 antisense and 44 lincRNAs, were identified in this screen as having differential expression between compact and expanded CCs. Though not evaluated functionally, the presence of antisense transcripts during this

interval suggests a regulatory role. Another study investigated lncRNAs in "high-quality" CCs (H-CCs) versus "poor-quality" CCs (P-CCs) by microarray in human (Xu *et al,* 2014). The samples were derived from *in vitro* fertilization and quality was defined by their morphology. Of the 20,000 lncRNAs examined, 633 were identified as being differentially expressed between H-CCs and P-CCs.

These CC lncRNA profile data are especially important because of the evidence that cytoplasm and its contents are shared in a limited way between CCs and the oocyte. In mammals, CCs that have been independently transfected with a GFP reporter allowed GFP mRNA to move into the oocytes, resulting in GFP expressing oocytes containing no plasmid (Macaulay *et al,* 2014). In *Drosophila*, nurse cells transfer RNA and other cytoplasmic components to oocytes (Cha *et al,* 2001; Nicolas *et al,* 2009); similar phenomena are seen in hydra (Alexandrova *et al,* 2005), and in mouse (Cossetti *et al,* 2014), which might be exosome mediated (Gezer *et al,* 2014; Pefanis *et al,* 2015). This communication by cytoplasmic sharing is a perfect medium by which regulatory lncRNAs may be moved from a somatic cell type into the developing germ line.

In a single-cell RNA-seq profile of MII oocytes and preimplantation embryos, 8,700 maternal lncRNAs were identified in the preimplantation embryo (Yan *et al,* 2013). Here, 660 lncRNAs were identified as being differentially expressed from MII oocytes to zygotes. These lncRNAs may affect gene activation during the maternal-zygotic transition. In this context,

many lncRNAs in the transition from MII oocytes to 2-cell embryos have been identified as having possible functional relevance (Hamazaki *et al,* 2015). In an impressive screen using strand-specific RNA-seq, more than 1,000 potentially functional lncRNA/mRNA pairs have been identified; a subset acting as pancRNAs in zygotes. A further screen on identified pairs would elucidate their functions. Similarly, a great deal of antisense transcription was found near promoters in drosophila oocytes (Brown *et al,* 2014b).

Several lncRNAs have been characterized in oogenesis in non-mammalian systems. Much like the previously mentioned *SRA* and SRY gene products, the *Drosophila oskar* gene is a good example of an RNA with coding and noncoding functions. Loss of the oskar protein caused defects in, oocyte polarity, and embryonic germ line specification, and abdominal development; the loss of the oskar RNA, however, caused the additional phenotype of early arrest in oogenesis (Jenny *et al,* 2006). This lncRNA is only translated upon localization to the posterior pole, where its 3'UTR is necessary to rescue the oogenesis defect, possibly by recruitment of other factors involved in the establishment of cell polarity (Kugler and Lasko 2009). The independence of this 3'UTR was underlined when it was identified years later in a genome wide profile of 3'UTR associated RNAs (uaRNAs) (Mercer *et al,* 2011). A similar RNA scaffolding function is found in *Xenopus* where two RNAs are required for proper cytoskeletal organization and oocyte polarity (Kloc *et al,* 2005). *XIsirt* is a lncRNA composed of short tandem repeats which are suspected to form stem loop structures for correct localization (Allen *et al,*

2003). *VegT* mRNA, another dual purpose transcript, is necessary for cytokeratin network assembly while the VegT protein is a required transcription factor for mesoderm and endoderm differentiation (Kofron *et al,* 1999; Xanthos *et al,* 2001). These structural functions are unlikely to transfer directly to mammals since mammals do not have the same asymmetric distribution of molecules associated with oocyte development, though this fact does not preclude their function altogether.

Beyond these studies of spermatogenesis and oogenesis done in animals, studies in plants are informing lncRNA-based mechanisms essential for gamete formation. An RNA-seq screen for lncRNAs expressed in reproductive tissues of rice identified a number of transcripts. One, *XLOC_057324*, was expressed exclusively in young panicles and pistils. A strains with a T-DNA insertion in *XLOC_057324* flowered prematurely and set fewer seeds (Zhang *et al,* 2014b). In another study done in rice, hybrids exhibiting long-day-specific male sterility carried a mutation in the lncRNA *LDMAR*. A point mutation in *LDMAR* was associated with increased DNA methylation of the locus, reduced *LDMAR* expression under long daylight conditions, and premature apoptosis of developing anthers (Ding *et al,* 2012). The mechanisms underlying these effects are not known.

IV.L Placentation

Initial data, though limited, are consistent with a role for lncRNAs in placenta formation and function, with some of the strongest results coming from studies of *H19* in placental function. *H19* is a source for miR-675

biogenesis, a miRNA that has been shown to directly downregulate Nodal Modulator 1 (NOMO1) and inhibit its ability to stimulate proliferation of a human trophoblast cell line (Gao *et al,* 2012; Keniry *et al,* 2012). In normal placenta, *H19* and its miR-675 repress NOMO1 mediated proliferation, but in preeclamptic placentas, *H19* and the miRNA are repressed, allowing NOMO1 misregulation to cause placental overgrowth.

In studies of HELLP syndrome – a maternal condition of hemolysis, elevated liver enzymes, and low platelets, which has its origins in placental insufficiency – a locus associated with this syndrome was mapped to an intergenic region harboring a lncRNA that is expressed in several trophoblast subtypes in human placenta. When knocked down in extravillous trophoblast cells, gene expression changes were associated with increased G1/S and cell death functions, as well as decreased G2/M, cell survival, and migration functions. Increasing accumulation of the *HELLP* lncRNA had the opposite effects, and decreased cell invasion (van Dijk *et al,* 2012). The mechanisms by which the *HELLP* lncRNA exerts these effects are unknown.

In other studies of term pregnancies with intrauterine growth restriction (IUGR), the *NEAT1* lncRNA was found to be enriched four-fold in IUGR *vs.* control placentae (Gremlich *et al,* 2014). This species is present in nuclear paraspeckles and is essential for their assembly (Clemson *et al,* 2009). It is not clear if increased *NEAT1* contributes to, or is a consequence of IUGR. In other studies of preeclampsia, the lncRNA *SPRY4-IT1*, which is expressed in placenta, was reported to be overexpressed in preeclamptic placentae. siRNA

knockdown of *SPRY4-IT1* in a transformed human trophoblast line increased cell migration and reduced apoptosis, whereas overexpressing *SPRY4-IT1* had the opposite effects (Zou *et al,* 2013). Additional lncRNAs were reported to exhibit differential expression in preeclamptic *vs*, control placentae, though the functional relevance has not be tested (He *et al,* 2013). *In vivo* manipulations are necessary to assess directly the importance of these lncRNAs in placenta function.

The placenta is associated with unique lncRNA-mediated control of some previously mentioned imprinted loci as well. *Airn* controls imprinted expression of *Igf2r* globally, and controls imprinted expression of two additional adjacent genes specifically in placental tissue, *Slc22a2* and *Slc22a3* (Nagano *et al,* 2008; Zwart *et al,* 2001). Similarly, the lncRNA *Kcnq1ot1* regulates the imprinting of four nearby genes in all tissues, but controls four additional, more distantly placed genes in placental tissue (Pandey *et al,* 2008). Both of these lncRNAs have been shown to directly interact with chromatin modifying machinery in a lineage-specific way, suggesting that other lncRNAs might work similarly. These imprinting mechanisms are probably more tightly regulated in the placenta due to the tissue's more direct role in embryonic growth.

IV.M Inheritance

Though DNA is responsible for genetic inheritance, non-genetic transmission of traits through meiosis has been observed, a phenomenon

referred to as transgenerational epigenetic inheritance (TEI) (Rakyan and Whitelaw 2003). Mechanisms underlying TEI are mediated by histone modifications, DNA methylation, prions, and most importantly for this review, RNA species. The first evidence for the involvement of RNA in TEI came from studies of paramutation, a form of TEI, involving the *b1* locus in maize. Two alleles of *b1* exist, *B-I* and *B'*, which are genetically identical, but *B'* is silent and lacks DNA methylation in a repeat region necessary for paramutation, whereas *B-I* is active and methylated in the region (Stam *et al,* 2002). When present in the same plant, *B'* converts *B-I* to its own state, and the conversion is stable through meiosis for several generations. The role of RNA in paramutation was demonstrated when it was shown that a 6kbp tandem repeat 100kbp upstream of the locus that has enhancer activity is transcribed, processed into siRNAs, and that this production and the paramutation phenotype requires the RNA-dependent RNA polymerase MOP1 (Alleman *et al,* 2006; Arteaga-Vazquez *et al,* 2010; Dorweiler *et al,* 2000; Stam *et al,* 2002).

Though not as robust as their plant counterparts, animal systems display similar RNA-mediated phenotype inheritance. Studies of *Kit+* (Rassoulzadegan *et al,* 2006), *Sox9* (Grandjean *et al,* 2009), *Cdk9* (Wagner *et al,* 2008), *Rasgrf1* (Herman *et al,* 2003), and work in stressed mice (Gapp *et al,* 2014) revealed paramutation-like effects, with evidence consistent with RNA-mediated mechanisms. Some of these early studies endeavored to prove the sufficiency of small RNAs to induce the phenotype of interest by

injecting miRNA species into wild type zygotes to recapitulate the phenotype. While making a strong argument for sufficiency, these studies do not answer all questions. For instance, elimination of the miRNA pathway by *DROSHA* knockout, or the piRNA pathway by *Mov10l1* knockout increased the penetrance of the *Kit*+ phenotype, suggesting that miRNAs and piRNAs were acting as suppressors rather than activators of paramutation (Yuan *et al,* 2015). The mechanisms controlling TEI is unknown in mammals, but it is impossible to eliminate either the importance of indirect lncRNA control *via* small RNA regulation, or of direct lncRNA transmission upon fertilization. RNA modifications appear to play a role, as the *Kit*+ and *Sox9* phenotypes are dependent on DNMT2, an RNA methyltransferase, though the universality of the effect must be studied further (Kiani *et al,* 2013).

IV.N Development

lncRNAs associated with preimplantation development have been characterized by RNA-seq of zygotes and other preimplantation stages of development (Caballero *et al,* 2014; Hamazaki *et al,* 2015; Paranjpe *et al,* 2013; Yan *et al,* 2013; Zhang *et al,* 2014a). By comparing these lncRNA profiles with profiles of parental gametes, it is possible to identify those lncRNA arising immediately after zygote activation. Evaluating the importance of these lncRNAs for early embryonic events will require their experimental manipulation (Sauvageau *et al,* 2013).

Some of the most extensive functional studies relating to preimplantation embryos come from studies of embryonic stem cells. Cultured embryonic stem cells express at least 226 lncRNAs, 137 of which have been shown to affect gene expression, and 26 of which are necessary to repress differentiation and maintain pluripotency (Guttman *et al,* 2011). These lncRNAs are involved in many functions. For instance, *Meg3* interacts with JARID2 to specifically recruit PRC2 and its repressive activity to embryonic development genes in *trans* (Kaneko *et al,* 2014). lncRNAs can play activating roles too. Six lncRNAs were shown to interact with WDR5, a protein shown to recruit MLL and its H3K4me3 activity (Wang *et al,* 2011; Yang *et al,* 2014). The lncRNA RoR does not control chromatin remodelers, but instead maintains expression of the core pluripotency factors by acting as a sponge to titrate out repressive miRNAs which would downregulate their translation (Wang *et al,* 2013). Through a variety of mechanisms, the central role for lncRNAs at this stage is to maintain self-renewal characteristics (reviewed in (Flynn and Chang 2014)).

Many critical steps in postimplantation somatic development have been shown to be regulated by lncRNAs. Table IV.3 summarizes several key examples. The reader is referred to recent reviews addressing this issue, including stem cell maintenance and differentiation (Batista and Chang 2013; Fatica and Bozzoni 2014; Flynn and Chang 2014; Ghosal *et al,* 2013; Yao and Jin 2014).

**Table IV.3. Roles for lncRNAs in post-implantation developmental processes.**

| Tissue | lncRNA | Function | References |
|---|---|---|---|
| Cardiovascular | *Fendrr* | Regulates cardiac development | (Grote and Herrmann 2013; Grote *et al,* 2013) |
| | *Braveheart* | Regulates cardiovascular development | (Klattenhoff *et al,* 2013) |
| | *tie1AS* | Regulates tie1 and vascular development | (Li *et al,* 2010) |
| Hematopoietic | *H19* | Maintains hematopoietic stem cell quiescence | (Venkatraman *et al,* 2013) |
| | *lncRNA-αGT* | Necessary for embryonic to adult α-globin switching | (Arriaga-Canon *et al,* 2014) |
| | *7 species* | Controls terminal erythroid differentiation | (Paralkar *et al,* 2014) |
| Musculoskeletal | *SRA* | Enhances myogenic differentiation and myogenic conversion of non-muscle cells | (Hube *et al,* 2011) |
| | *lincMD1* | Enhances myoblast differentiation | (Cesana *et al,* 2011) |
| | *Dum* | Regulates myoblast differentiation | (Wang *et al,* 2015) |
| | *MUNC* | Induces myoblast differentiation | (Mueller *et al,* 2015) |
| Neural | *Six3OS* | Regulates Six3 and retinal development | (Rapicavoli *et al,* 2011) |
| | *TUNAR* | Regulates pluripotency and neural differentiation | (Lin *et al,* 2014) |
| | *Evf2* | Represses Dlx5 and controlling GABA circuitry | (Berghoff *et al,* 2013) |
| | *Pnky* | Regulates neurogenesis from neural stem cells | (Ramos *et al,* 2015) |
| Mammary | *Pinc1* | Regulates alveolar development | (Shore *et al,* 2012) |
| | *Zfas1* | Regulates alveolar development | (Askarian-Amiri *et al,* 2011) |
| Endoderm | *DEANR1/LINC00261* | Regulates endoderm differentiation | (Jiang *et al,* 2015) |

| | | | |
|---|---|---|---|
| Adipose | *HOTAIR* | Regulates preadipocyte differentiation | (Divoux *et al,* 2014) |
| | 10 species | Regulates preadipocyte differentiation | (Sun *et al,* 2013) |

## IV.O Reproductive Disease

Beyond placental insufficiencies discussed above that are associated with perturbations in lncRNA regulatory mechanisms, several lines of evidence document additional roles for lncRNAs in various reproductive pathologies. For example, in a cohort study of nineteen men with idiopathic infertility and histologically confirmed meiotic arrest, copy number variants of three genes including the lncRNA *LOC100507205* were found to be unique to the meiotic arrest patients as compared to 95 fertile controls (Eggers *et al,* 2015). In a second example, a screen in women for lncRNAs associated with premature rupture of the placental membranes (PPROM) identified thirteen lncRNAs that were differentially expressed in PPROM *vs*. full-term placentae. These lncRNAs appear to play roles in the inflammatory response, smooth muscle contraction, and ligand-receptor interactions (Luo *et al,* 2015; Luo *et al,* 2013). In a third study, women suffering from polycystic ovary syndrome (PCOS), characterized by high serum androgens, absence or irregular menstruation, and infertility, the lncRNAs *SRA* and *CTBP1-AS1* were demonstrated to be overexpressed as compared to healthy controls (Liu *et al,* 2015b; Liu *et al,* 2014). Many additional studies have focused on the role of lncRNAs in various reproductive cancers. For example, *SRA*, already discussed as a regulator of nuclear hormone responses, is elevated in

estrogen-responsive ovarian and breast cancer (Hussein-Fikret and Fuller 2005; Leygue *et al,* 1999; Murphy *et al,* 2000). *PGEM1* and *PCNR1* lncRNAs were first identified as being overexpressed in aggressive prostate adenocarcinomas (Yang *et al,* 2013), and *NEAT1* once again was implicated in the progression of androgen-insensitive prostate tumors (Chakravarty *et al,* 2014). Finally, two related genital malformation syndromes are associated with epigenetic alterations at *H19*, which is methylated on the paternal allele and thereby silenced. In addition to silencing the paternal copy of *H19*, methylation is required for expression of the paternal copy of *Igf2*, to which *H19* is linked. Silver-Russell syndrome is clinically and genetically heterogeneous, with some patients exhibiting hypomethylation of *H19*. The most severely hypomethylated females show congenital aplasia of the uterus and upper vagina, and severely hypomethylated males exhibit cryptorchidism and testicular agenesis (Bliek *et al,* 2006; Bruce *et al,* 2009). *H19* hypomethylation is also associated with some Müllerian aplasia patients, whose congenital abnormalities of the female genital tract produce vaginal and uterine malformations that limit reproduction to methods involving surrogacy (Sandbacka *et al,* 2011). Because paternal silencing of *H19* and paternal expression of *Igf2r* are coupled, it is not clear whether aberrant expression of either or both loci is responsible for the reproductive phenotypes of these patients. For each of these examples from human clinical studies, lines of evidence beyond associations are required to demonstrate the importance of candidate lncRNAs in reproductive processes, and further investigation is

necessary to reveal their mechanisms of action. Animal studies will be important in this regard, such as those demonstrating that mice deficient for the lncRNA *NEAT1* have impaired corpus luteum formation and failure to maintain pregnancy (Nakagawa *et al,* 2014).

### IV.P Conclusions

While lncRNAs have been definitively recognized as important mediators of cellular fate and function, their role in the reproductive processes that initiate in the early embryo and continue through an organism's reproductive lifespan are only now being elucidated. Descriptive, hypothesis-generating studies that characterize lncRNAs associated with reproduction represent the low hanging fruit in the field. These studies are heavily concentrated in specific reproductive events but remain sparse in others, and have been applied to a limited number of organisms. The more challenging studies entail identifying which lncRNAs discovered in such studies in fact influence reproductive processes, and by what mechanisms. Detailed mechanistic studies will require manipulating lncRNA expression and evaluating reproductive phenotypes; characterizing lncRNA structures, ideally those assumed *in vivo*; identifying proteins and other factors interacting with lncRNAs; cataloging the chemical modifications present on lncRNAs and their interacting partners, and assessing the importance of those modifications for structure and function. Expanding such analyses to a diversity of species and a diversity of individuals within human populations will reveal the evolutionary conservation of lncRNA-mediated mechanisms affecting reproduction, and

genetic variants important to reproductive health. Given the vast array of lncRNAs transcribed from complex genomes and their range of activities, such studies will rival the complexity and importance of functional genomic analyses of the coding genome.

# V METHYLATION AS A POTENTIAL DRIVER OF PHENOTYPIC DIVERSITY IN *CANIS LUPUS FAMILIARIS*

## V.A Introduction: DNA methylation and the dog

The domestic dog, *Canis lupus familiaris*, has served as man's guard, partner, and companion for centuries. In recent years, the dog has proven to be an excellent model species. With its unique population structure composed of hundreds of highly inbred, morphologically and behaviorally distinct extant breeds, the dog has massively expanded our understanding of morphology, disease, and behavioral genetics (reviewed in Sutter and Ostrander, 2001).

As fruitful as the canine genome has proven for the discovery of genetic variants, the investigation of the canine methylome is in its infancy. Recent studies have implicated DNA methylation as a possible driver of phenotypic variation in different dog breeds. For example, a potential epigenetic contribution to behavior differences in the Beagle, German Shepherd Dog, and Sapsaree dog breeds, dogs bred and known for highly diverse behavior, became evident when investigators found that brain tissue from these dogs displayed distinct differences in methylation levels at the promoter of *Monoamine Oxidase A*, with accompanying differences in MAOA mRNA levels (Eo *et al,* 2016). Monoamine Oxidase A is a known modulator of behavior: Mouse knockouts display aggression and anxiety-like behavior (Cases *et al*, 1995; Chen *et al*, 2004); in human, *MAOA* has gained the nickname "warrior gene" due to genotypic predictiveness of aggressive behavior (McDermott *et al*, 2009) and is the target of the MAOI class of antidepressants (as reviewed

by Shulman, Herrmann, and Walker, 2013).

Another study highlights the interaction of genetics and epigenetics: Mutations in the *cyclooxygenase-2* (*Cox-2*) promoter are thought to increase risk for renal dysplasia, but have variable penetrance (Whiteley *et al*, 2011). However, bisulfite sequencing of a CpG island within the *Cox-2* promoter revealed that methylation state was more strongly correlated to clinical renal dysplasia than genotype (Whiteley, 2014).

Whole methylome analysis in the dog is comparably nascent. From previous investigators, we know that the canine methylome has 1.5 and 3 times respectively more CpG islands than human and rodent; however, the dog has comparably fewer promoter-associated CpG islands (Han and Zhao, 2009). Comparing the dog to its parent species, *Canis lupus*, reveals that dog-wolf DMRs tend to overlap transposable element sequence and are more commonly hypermethylated in the dog relative to the wolf (Koch *et al*, 2016). In a more targeted comparison of cloned dogs with or without gonadal dysgenesis, whole genome bisulfite sequencing revealed aberrant methylation of the *SRY* gene as a potential cause of incomplete sexual development (Jeong *et al,* 2016).

Here, we examine whether DNA methylation could be a contributing factor to canine body size. We first took a candidate approach, assessing a CpG island in the promoter of *Igf1*. A deletion in intron 2 of *Igf1* has a dramatic impact on body size, explaining over 20% of body size variation across dog breeds (Sutter *et al,* 2007). We test the hypothesis that this allelic variant could

affect *Igf1* activity through variable methylation of the *Igf1* CpG island.

Second, we took a genome-wide approach to identify DMRs that vary between two phenotypically distinct dog breeds, the Great Dane and the Yorkshire Terrier. We perform reduced-representation bisulfite sequencing (RRBS) of genomic DNA of two Great Danes and two Yorkshire Terriers and identify 37 DMRs associated with body size, at genes implicated in body size and cancer phenotypes. This work presents one of the first forays into what is likely to be a rich source of discovery for other phenotypic diversity such as behavior and disease risk.

## V.B Materials and Methods

**Genomic DNA collection:** Peripheral blood mononuclear cell (PBMC)-derived genomic DNA (gDNA) from twelve dogs, four *Igf1* ancestral allele homozygotes, four heterozygotes, and four derived allele homozygotes were a gift from the Boyko Lab; three additional gDNA samples were acquired from the Cornell Biobank and used for mRRBS (**Table V.1**).

**Table V.1. Name, breed, age, sex, weight, genotype, and source information for fifteen dogs used for *Igf1* promoter methylation analysis and mRRBS.** Abbreviations: “F,” female; “S,” spayed; “M,” male; “C,” castrated; “I,” intact; “a,” ancestral (no SINE insertion); “d,” derived (SINE insertion); *denotes unknown. Assays performed on each gDNA sample, either RRBS or *IgF1* targeted sequencing and COBRA (IgF1), is indicated at far right column.

| ID | Breed | Age | Sex | Weight (kg) | *Igf1* Genotype | Source | Assay |
|---|---|---|---|---|---|---|---|
| PFZ15E01 | English Bulldog | 5y 4m | FS | 42.4 | ad | Boyko | IgF1 |
| FJ67 | Village Dog - Fiji | Adult* | F* | 11.4 | dd | Boyko | IgF1 |
| Stn_399 | Village Dog - Belize | 2y* | MC | 17.7 | dd | Boyko | IgF1 |
| 18396 | Great Dane | 3y 5m | MC | 56 | aa | Boyko | IgF1, RRBS |
| Bix | American Water Spaniel | 5y 4m | MC | 18.6 | aa | Boyko | IgF1 |
| Scooby | Mix | 1y | MC | 7.4 | dd | Boyko | IgF1 |
| 16232 | Rough Collie | 2y 5m | MC | 27 | aa | Boyko | IgF1 |
| Ruby | American Water Spaniel | 3y 8m | FI | 13.2 | ad | Boyko | IgF1 |
| Maggie | Mix | 9y 9m | FS | 7.6 | dd | Boyko | IgF1 |
| FJ87 | Village Dog - Fiji | Adult* | M* | 18.6 | ad | Boyko | IgF1 |
| FJ61 | Village Dog - Fiji | Adult* | M* | 15.2 | ad | Boyko | IgF1 |
| 14930 | German Shorthaired Pointer | 8y 4m | FS | 26.3 | aa | Boyko | IgF1 |
| 18421 | Yorkshire Terrier | Adult* | MC | * | * | Biobank | RRBS |
| 4089 | Yorkshire Terrier | Adult* | MC | * | * | Biobank | RRBS |
| 13856 | Great Dane | Adult* | MC | * | * | Biobank | RRBS |

**Bisulfite conversion, PCR, and COBRA for *Igf1*:** Bisulfite conversion was performed on 250 ng of canine PBMC-derived gDNA (provided by the Boyko lab) using the EZ DNA Lightning Kit (Zymo D5001). BS-PCR was performed with the following primers for the Igf1 CpG island: IGF1BS1F 5'-AYGGTTAGAAGTGTAATTGTTGTGGTTTGG-3'; IGF1BS1R 5'-CTACACCCRAAACTCCTCCAAAAAC-3' and EpiMark Hot Start *Taq* DNA Polymerase (NEB M0490) following the manufacturer's protocol. COBRA was performed by digesting BS-PCR products with 5 U *BstUI* at 60C for 1 hour. Digestion products were separated on a 4% agarose gel.

**Targeted bisulfite sequencing:** The method for targeted bisulfite sequencing is described in Chapter II. The grep and reference sequence used for canine Igf1 are included in **Table V.2**. Methylation analysis was performed with QUMA (Kumaki *et al*, 2008). Total reads per sample are listed in **Table V.3**.

**Table V.2. Sequences used for QUMA analysis of the *Igf1* CpG island.** In preparation for analysis, raw files were probed for reads containing amplicon-specific sequences using the grep function in Linux. QUMA was performed on these reads.

| | |
|---|---|
| **grep** | GGAGTAGAAGAGTTTTGGGTTA |
| **Reference for QUMA** | ggagcagaagagctctgggtcacggcaaagaacgggtcagtgcagggccg agggcgcttcggggtgggggagcctggggctgaacgacagtgaggaggga ggctccatcccaagcgcccaaatacgcactttctttcggaatcgaaaact gtgcgcgccgcctagtggccc |

**Table V.3. Total bisulfite reads per sample for IGF1BS1F and R BS-PCR amplicon.**

| Dog ID | Genotype | Total Reads |
|---|---|---|
| 18396 | aa | 2024 |
| Bix | aa | 1246 |
| 16232 | aa | 1100 |
| 14930 | aa | 755 |
| PFZ15E01 | ad | 951 |
| FJ87 | ad | 1988 |

| | | |
|---|---|---|
| FJ61 | ad | 811 |
| FJ67 | dd | 1195 |
| Stn_399 | dd | 1046 |
| Scooby | dd | 742 |
| Maggie | dd | 739 |

**Multiplex RRBS and DMR discovery:**

Multiplexed reduced representation bisulfite sequencing (mRRBS) libraries were prepared by Roman Spektor, following the protocol described in Boyle *et al,* 2012. Library quality control and DMR discovery were performed with MethylKit package in R (Akalin *et al,* 2012). GO Analysis was performed with PANTHER (Thomas *et al,* 2003). Total reads, median CpG coverage, and average reads per CpG for each RRBS library were reported with Bismark (https://www.bioinformatics.babraham.ac.uk/projects/bismark/) and are listed in **Table V.4**.

**Table V.4. Total reads, total cytosines analyzed, and average coverage per per cytosine for mRRBS libraries of two Yorkshire Terriers (YT) and two Great Danes (GD).**

| Dog ID (Breed) | Total sequences analyzed | Total cytosines analyzed | Avg coverage per cytosine |
|---|---|---|---|
| 19421 (YT) | 14055094 | 127150911 | 8.11 |
| 4089 (YT) | 13534273 | 125153838 | 8.23 |
| 13856 (GD) | 13331698 | 119917584 | 7.96 |
| 18396 (GD) | 14733531 | 135799226 | 9.23 |

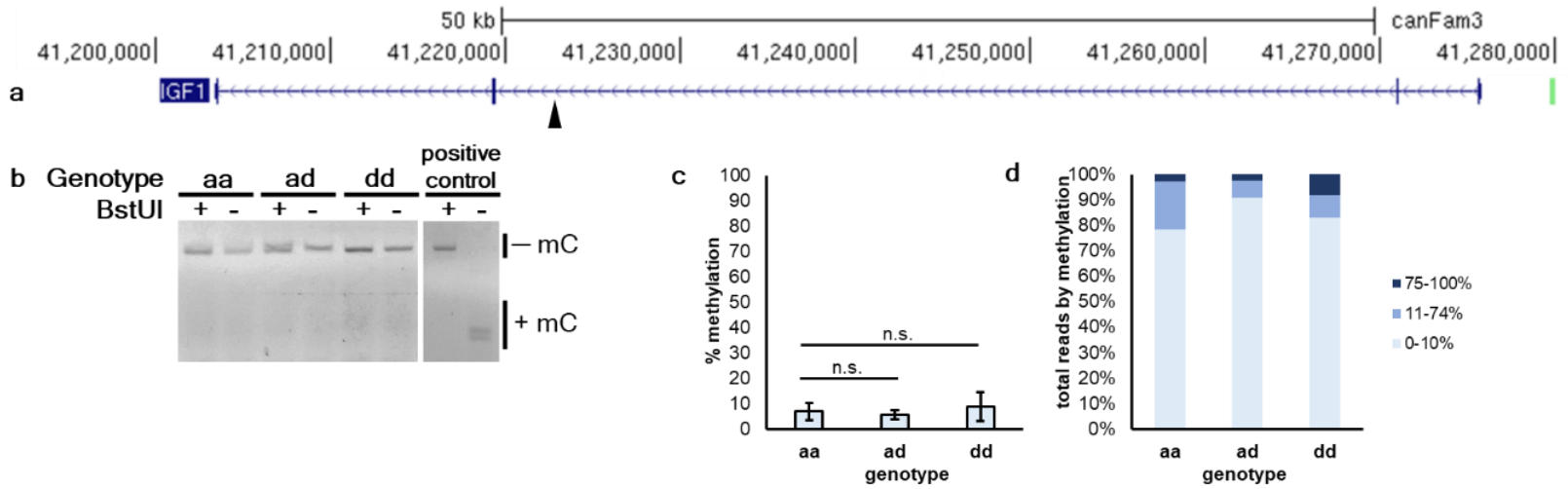
**HapFLK analysis:** Adam Boyko intersected coordinates of DMR-associated genes with regions of strong positive selection, using hapFLK data from Schlamp *et al*, 2017.

## V.C Results

### V.C.1 Bisulfite analysis of the *Igf1* CpG island shows invariable hypomethylation regardless of *Igf1* SINE genotype.

Canine *Igf1* harbors a 284bp CpG island roughly 4.6kb upstream of the Igf1 promoter (**Fig V.1a**) that is not conserved in either mouse or human. To test the hypothesis that variable methylation fo the CpG could correspond to the *Igf1* SINE genotype correlated to body size, we designed bisulfite PCR primers that encompass >80% of CpG island sequence, spanning 17 CpG dinucleotides. Our PCR product includes two BstUI sites by which we evaluated methylation qualitatively via COBRA. Compared to a positive control (the murine *Rasgrf1* DMR in wild-type sperm, which is 100% methylated), our amplicon is hypomethylated in all genotypes (**Fig V.1b**). Targeted bisulfite sequencing confirmed COBRA results, with average methylation below 10% in all genotypes (**Fig V.1c**). Stratification of bisulfite PCR reads by methylation status indicates that in all cases, the vast majority of sequencing reads are under 10% methylated (**Fig V.1d**).

**Fig V.1. Methylation of a CpG island upstream of the canine *Igf1* promoter does not vary with presence or absence of the *Igf1* SINE insertion associated with body size. a)** A 284bp long CpG island (green bar) lies 4.6kb upstream of *Igf1* coding sequence on CFA15. Black arrowhead denotes the location of the SINE_Ccf insertion most positively correlated with body size variation (Sutter *et al*, 2007). **b)** Representative COBRA for homozygous ancestral (aa), heterozygous (ad), and homozygous derived (dd) genotypes at *Igf1*. *BstuI* digestion queries two BstUI sites within the *Igf1* CpG island. Sperm COBRA for the murine *Rasgrf1* DMR is shown as a positive control for BstUI digestion. **c)** Average *Igf1* DMR methylation of aa, ad, and dd dogs. Error bars represent standard deviation of biological triplicate at minimum. **d)** Total reads by methylation of three dogs with respectively the aa, ad, and dd genotypes. Key for methylation at right. n.s., not significant.

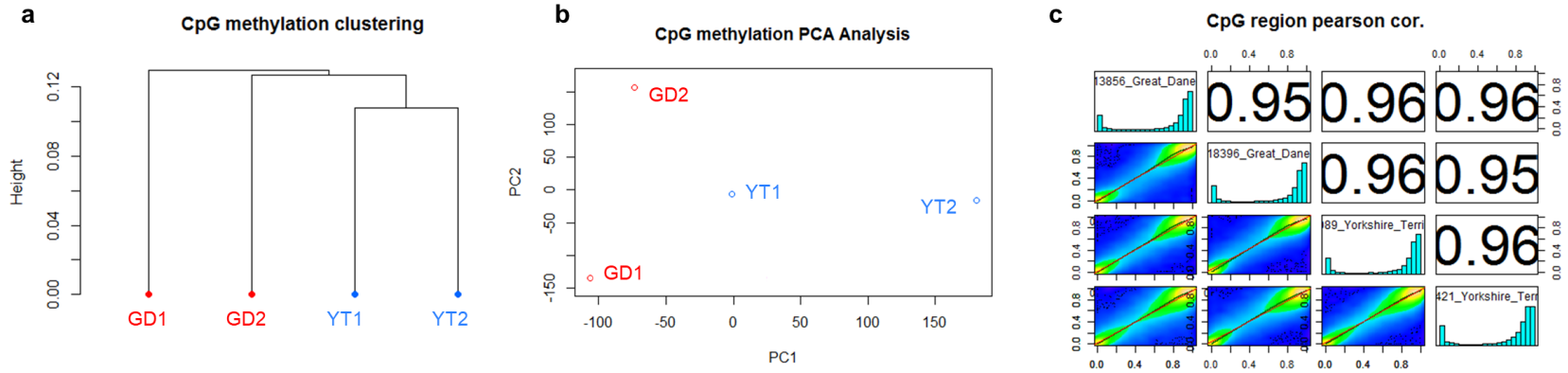


### V.C.2 RRBS of Yorkshire Terriers and Great Danes

We next took an unbiased approach to DMR discovery between pairs of adult male castrated Yorkshire Terriers (YTs) and Great Danes (GDs) via multiplexed reduced representation bisulfite sequencing (mRRBS).

Using the methylKit package in R, we evaluated the similarity between and across paired samples (**Fig V.2**). We found that YT and GD samples clustered together hierarchically (**Fig V.2a**) and in Principal Components space (**Fig V.2b**); all four samples tended to be very highly methylated or unmethylated at cytosines; ie, did not display levels of intermediate methylation (Scatterplots and histograms in **Fig V.2c**) and are similar in their methylation compositions by Pearson correlation (Numbers at right in **Fig V.2c**).

**Fig V.2. Analysis of mRRBS libraries of two Yorkshire Terriers and two Great Danes. a)** Hierarchical clustering of Yorkshire Terrier (YT1 and 2) and Great Dane (GD1 and 2) using Pearson correlation distance. YT samples cluster together, as do GD samples. **b)** Principal Components Analysis of paired GD and YT samples demonstrate that GD1 and GD2 are closer together in principal component space, as are YT1 and YT2. **c)** CpG region correlation across all four samples. Scatterplots of percent methylation values at left; showing that most cytosines are either hypomethylated or hypermethylated. Histograms along diagonal denote percent methylation histograms per cytosine of each sample and support scatterplot results. Numbers at right indicate pairwise Pearson correlation scores; all sample pairs have high Pearson correlation scores, indicating similar percent methylation profiles across all samples.



We queried DMRs in 1kb sliding windows that differed in methylation by greater than 50%. We discovered 57 DMRs between YTs and GDs; 26 were hypermethylated in YT relative to GD; 31 were hypomethylated (**Table V.3**). We then assayed proximity to genes annotated in Ensembl.

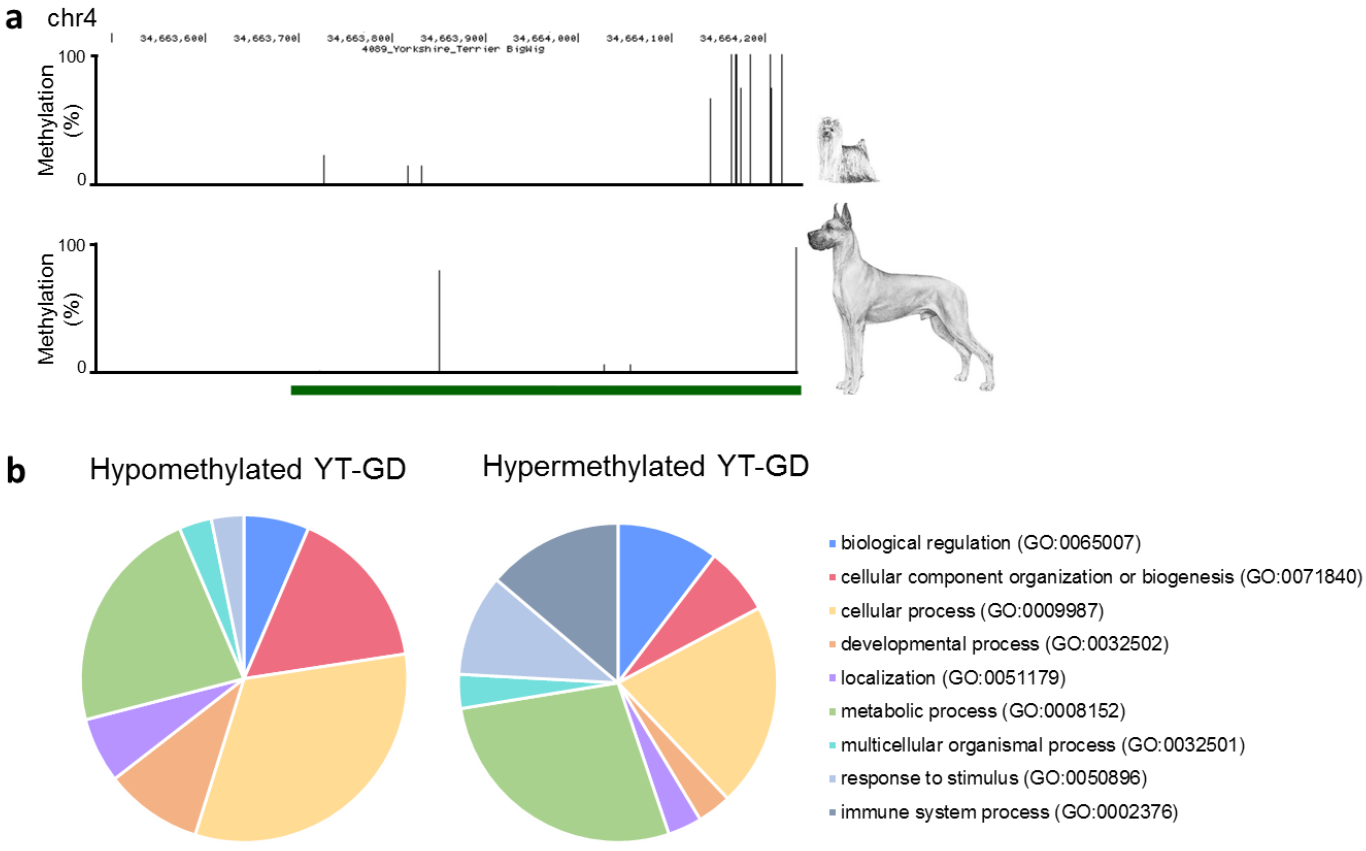
**Table V.5. List of genes associated with DMRs that are a) hypermethylated and b) hypomethylated in YTs relative to GDs.** Genes are listed by EnsemblID by order of significance; all genes have a significance value of $p < 0.01$. Distance from the gene transcriptional start site is included.

| Gene (Ensembl ID) | DMR Distance from Gene (bp) |
|---|---|
| **a) Hypermethylated in YT** | |
| ENSCAFT00000049616.1 | 0 |
| ENSCAFT00000042926.1 | 10344 |
| ENSCAFT00000047842.2 | 9802 |
| ENSCAFT00000037072.2 | 0 |
| ENSCAFT00000053566.1 | -107324 |
| ENSCAFT00000016313.3 | -25316 |
| ENSCAFT00000016894.3 | -325 |
| ENSCAFT00000022329.3 | 0 |
| ENSCAFT00000050995.1 | 4956 |
| ENSCAFT00000027130.3 | -11138 |
| ENSCAFT00000028741.3 | 6607 |
| ENSCAFT00000021177.3 | 2213 |
| ENSCAFT00000023460.4 | 0 |
| ENSCAFT00000042829.1 | 589 |
| ENSCAFT00000047988.1 | 6756 |
| ENSCAFT00000043153.1 | -8085 |
| ENSCAFT00000005482.3 | 192 |
| ENSCAFT00000006283.3 | 0 |
| ENSCAFT00000051932.1 | 30879 |
| ENSCAFT00000052641.1 | 14322 |
| ENSCAFT00000058662.1 | -24888 |
| ENSCAFT00000030552.4 | 29563 |
| ENSCAFT00000031243.3 | 921 |
| ENSCAFT00000039732.2 | 0 |
| ENSCAFT00000012950.3 | 0 |
| ENSCAFT00000017342.4 | -3456 |
| ENSCAFT00000044991.1 | 16327 |
| ENSCAFT00000050441.1 | 17197 |
| ENSCAFT00000041364.1 | 0 |
| ENSCAFT00000020794.3 | -476 |
| ENSCAFT00000046108.1 | -334 |
| ENSCAFT00000030537.3 | -3279 |
| **b) Hypomethylated in YT** | |

| | |
|---|---|
| ENSCAFT00000000203.4 | 17583 |
| ENSCAFT00000052171.1 | -55291 |
| ENSCAFT00000046618.1 | 0 |
| ENSCAFT00000022760.3 | 21525 |
| ENSCAFT00000023926.2 | 19286 |
| ENSCAFT00000031783.3 | -377 |
| ENSCAFT00000037005.1 | 857 |
| ENSCAFT00000019096.3 | 2124 |
| ENSCAFT00000008978.4 | -57028 |
| ENSCAFT00000025903.3 | 12976 |
| ENSCAFT00000047889.1 | 225 |
| ENSCAFT00000052151.1 | 23122 |
| ENSCAFT00000052842.1 | -25661 |
| ENSCAFT00000046319.1 | -29958 |
| ENSCAFT00000049388.2 | -19456 |
| ENSCAFT00000030476.4 | 3467 |
| ENSCAFT00000008665.3 | -16556 |
| ENSCAFT00000020170.3 | -16507 |
| ENSCAFT00000016266.3 | -3458 |
| ENSCAFT00000025797.3 | -10933 |
| ENSCAFT00000052265.1 | -54414 |
| ENSCAFT00000046493.2 | -21 |
| ENSCAFT00000015169.2 | -7884 |
| ENSCAFT00000051011.1 | 22947 |
| ENSCAFT00000021351.3 | -19428 |
| ENSCAFT00000018069.3 | -17797 |

The most significant DMR between YTs and GDs ($p = 3.59 \times 10^{-54}$) is located upstream of the *GLUD1* gene on CFA 4; this and others evaluated visually in the UCSC Genome Browser to confirm differential methylation across the annotated CpG island (**Fig V.2a**). GO analysis of significantly associated DMRs ($p < 0.01$) did not reveal significantly enriched biological processes (**Fig V.2b**).

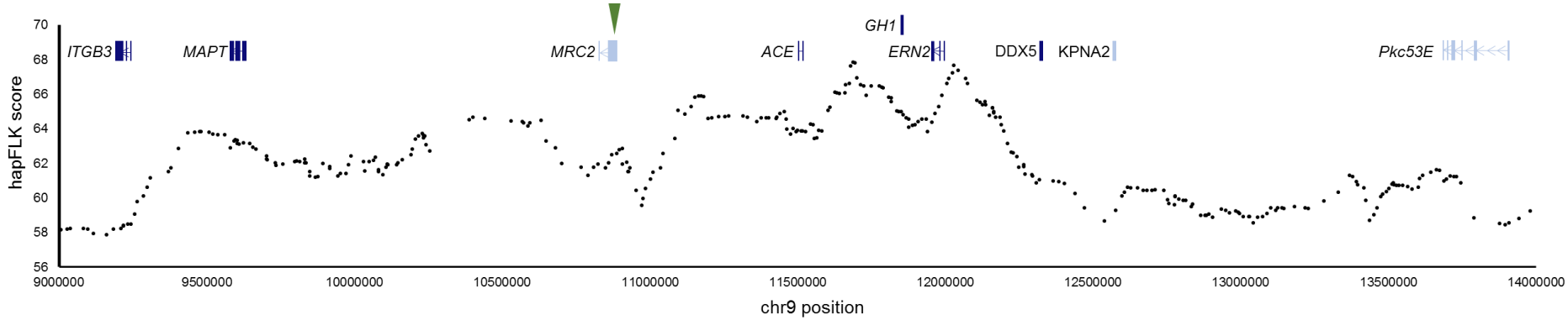
**Fig V.3. Manual inspection of the *GLUD1* DMR and GO analysis of all identified DMRs. a)** Visualization of differential methylation at the CpG island (green bar) upstream of the *GLUD1* gene in the UCSC Genome Browser. YT ID #4089 and GD ID #13856 are shown. The YT is hypermethylated along the distal end of the CpG island, whereas the GD is hypomethylated. **b)** GO Analysis of DMRs that are hypomethylated or hypermethylated by 50% in YTs relative to the GD. GO term key at right. None of these terms are significantly enriched in either hypomethylated or hypermethylated DMRs.



### V.C.2.i Positive selection signatures at YT-GD DMRs.

Chromosomal coordinates of statistically significant DMRs were intersected with coordinates of the canine genome under strong positive selection, identified using the statistic hapFLK. hapFLK utilizes haplotype frequencies within samples, and is therefore quite utile for evaluating breed-specific regions of selection (Fariellio *et al*, 2013). HapFLK scan of 25 purebred dog breeds, with 25 dogs per breed, reveals an average genome hapFLK of approximately 55; 68 genomic regions had statistically significant ($p < 0.003$) elevated hapFLK scores (Schlamp *et al*, 2015; Adam Boyko personal communications). One DMR, which lies intragenic to *MRC2* was within a 1Mb block of positive selection on chr9; however, the highest hapFLK scores in this region centered around the known size gene *GH1* (**Fig V.4**). No other DMRs were located within regions of positive selection.

**Fig V.4. The *GH1* positive selection sweep on chr9 contains a DMR at the *MRC* gene.** Note that the highest hapFLK scores center around *GH1*. The YT-GD DMR is located within *MRC2*, demarcated by the green arrowhead. Annotated genes from *C. familiaris* are indicated in dark blue; annotated genes from other species are indicated in light blue.



## V.D Discussion

### V.D.1 Methylation of the *Igf1* CpG island does not correlate with *Igf1* genotype.

We used COBRA and targeted bisulfite sequencing to assay correlation of the SINE insertion in intron 2 of *Igf1* with methylation at a CpG island upstream of the *Igf1* promoter. In a panel of twelve dogs of known genotype, the CpG island is invariably hypomethylated. These findings do not rule out changes in DNA methylation in other tissues; tissue-specific patterns of DNA methylation have been well documented (Doi *et al,* 2009; Chaithanya Ponnaluri *et al,* 2017). In addition, while all dogs tested were adults at the time of DNA collection, they varied in sex and intact status, which could affect DNA methylation (Christensen *et al,* 2009). However, with these caveats in mind, our data support that the *Igf1* SINE insertion that associates with body size variation does not act by imparting variable methylation to the DMR in upstream of the *Igf1* promoter.

DNA methylation is, of course, one of many means by which a noncoding variant can affect gene activity. Another strong candidate for the mechanism of the *Igf1* SINE is enhancer disruption or decommissioning. While *in silico* analysis of the orthologous regions in mouse and human did not strongly support disruption of an intragenic enhancer, enhancers are poorly conserved across mammalian species (Villar *et al,* 2015). While the same considerations must be taken regarding the tissue-specificity of enhancers (Visel *et al,* 2009 and reviewed in Ong and Corces 2011), further investigation could involve ChIP-qPCR for histone markers of active and inactive enhancers within *Igf1* or, in a genome-wide iteration, ChIP-Seq in canine tissues.

### V.D.2 Genome wide YT-GD DMR discovery

RRBS analysis revealed 57 DMRs between YTs and GDs. One DMR fell within a region of positive selection on chr9; however, the highest hapFLK peaks center over the *Growth Hormone 1* gene, which is known to contribute to body size. We conclude that, at least in our limited data set and with the caveat that we utilize mRRBS rather than whole genome bisulfite sequencing, that the body size variation that so distinguishes the YT and GD breed is more likely to be driven by genetic changes rather than differential methylation.

An additional consideration, besides our restricted data set and the reduced, but not comprehensive, nature of mRRBS, regards the intersection of DMRs with potential enhancer or other regulatory regions, as discussed above. None of the DMRs discovered appear to impact putative regulatory sequences as annotated by ORegANNO (Griffith *et al*, 2007). A canine ENCODE project does not yet exist as it does for mouse, human, worm, and fly (The ENCODE Project Consortium, 2012; Sloan *et al*, 2016); comparing the location of canine DMRs to putative regulatory regions in other species may again prove uninformative given the poor conservation of enhancers.

Our findings do bring to light the possibility of differential methylation as a potential source of disease risk. While this has been explored extensively in humans (discussed in Chapter I), purebred dog breeds have known predispositions to numerous diseases, many of which have a human homologue. This study reveals two DMRs closely associated with two genes that echo some of these disease risks

in the Great Dane and the Yorkshire Terrier; first, *MRC2* and osteosarcoma risk in the Great Dane, and second, *GLUD1* and hypoglycemia in the Yorkshire Terrier.

**MRC2 and Osteosarcoma:** Different cancers in the domestic dog have heritability values of up to 0.69 (Dobson *et al,* 2013); however, only a handful of causative mutations are identified. These include a single point mutation is associated with squamous cell carcinoma of the digit in the Poodle (Karyadi *et al*, 2013) point mutations in *BRCA1*, *BRCA2*, and *CDK5RAP2* for mammary cancer risk in English Springer Spaniels (Rivera *et al*, 2009; Melin *et al*, 2016). A syndrome that includes renal cystadenocarcinoma has been mapped to a mutation in the *BHD* gene in German Shepherds (Lingaas *et al*, 2003), and risk haplotype associations have been made in mast cell tumor and lymphoma risk in the Golden Retriever (Tonomura *et al*, 2015). However, genetic bases for other prevalent canine cancers including hemangiosarcoma, prostate adenocarcinoma, and osteosarcoma have yet to be identified. And while myriad reasons exist for why genetic markers for these highly prevalent cancers have not been identified yet in dogs, a strong hypothesis could be that environmental and epigenetic factors influence cancer risk. In human, many cancers have demonstrated interactions between genetic and environmental etiologies, among them breast cancer (Nickels *et al*, 2013) and non-small cell lung cancer (Davidson *et al*, 2016). Methylation analysis of cancers as a means of surveillance has also shown promising. In humans, methylation analysis of single genes in tumor cells have been shown to predict metastatic risk and/or disease-free interval in colon cancer (Chen *et al*, 2005) and breast cancer (Chimonidou *et al*, 2013; Ulirsch *et al*, 2012). In the past year, a handful of studies have explored the

utility of whole blood-derived DNA methylome analysis as a surveillance method for cancer risk (Roos *et al*, 2016; Ambatipudi *et al,* 2017; reviewed in Yokoi *et al*, 2017) and as a substitute DNA methylation profile to predict allergy risk (Langie *et al*, 2017).

The single DMR within a positive selective signal on chr9 falls directly within the gene body of C-Type Mannose Receptor 2 (MRC2). A key factor for collagen turnover, dysregulation of MRC2 is implicated in cancer metastasis, namely bone metastasis, as well as primary bone cancers such as osteosarcoma (St Croix *et al,* 2000, reviewed in Melander *et al,* 2015 and Sturge 2016). In dogs, osteosarcoma (OSA) is diagnosed almost exclusively in giant breed dogs including the GD (Brodey 1979; Misdrop 1980; Ru *et al,* 1998; Rosenberger *et al,* 2007; McNeill *et al,* 2007; Selvarajah and Kirpensteijn 2010). Some researchers have already targeted MRC2 as a means of OSA therapy *in vitro* and in a mouse model (Engelholm *et al,* 2015) with promising results. In dog, MRC2 has already been demonstrated to be overexpressed on canine OSA cell surfaces relative to control lines (Milovancev *et al,* 2013). Appropriate follow up for this finding would involve additional bisulfite analysis of other toy and giant breed individuals to determine methylation status of the MRC2 DMR, as well as MRC2 methylation and expression analysis in canine OSA lines.

**Hypoglycemia:** Hypoglycemia, or low blood glucose, is a malady of puppies, especially toy and hunting dog breeds, thought to occur due to inadequate hepatic stores of glycogen (Fieberger 1986; Bistner *et al*, 2000). If left untreated, hypoglycemia can cause listlessness, low energy, and can progress quickly to seizures and shock (Mila *et al*, 2017). Fortunately, hypoglycemia in juvenile dogs is fairly manageable with frequent, regular feeding and close monitoring. To our

knowledge, differences in resting insulin ranges in these dog breeds have not been investigated, most likely due to the costs and the risks associated with fasting hypoglycemia-prone puppies for resting insulin levels. Our highest significance DMR, shown in **Fig V.2a**, lies 1kb upstream of the *GLUD1* gene. Gain of function mutations in *GLUD1* are associated with several cases of hyperinsulinemic hypoglycemia in humans (Miki *et al*, 2000; Yasuda *et al,* 2001; Balasubramaniam *et al*, 2011; Odom *et al*, 2016; reviewed in Stanley 2011). While this could represent an interesting connection between differential methylation and disease risk, a number of preliminary assays are warranted, beginning with measuring resting insulin levels in toy, hunting, medium, and large breed puppies. Should resting insulin levels vary between the groups, bisulfite analysis of the *GLUD1* DMR in more animals of these groups could be performed.

# VI EXPANDED DISCUSSION

## VI.A Alternative mechanisms for the *Rasgrf1* repeats in directing methylation in the male embryonic germline

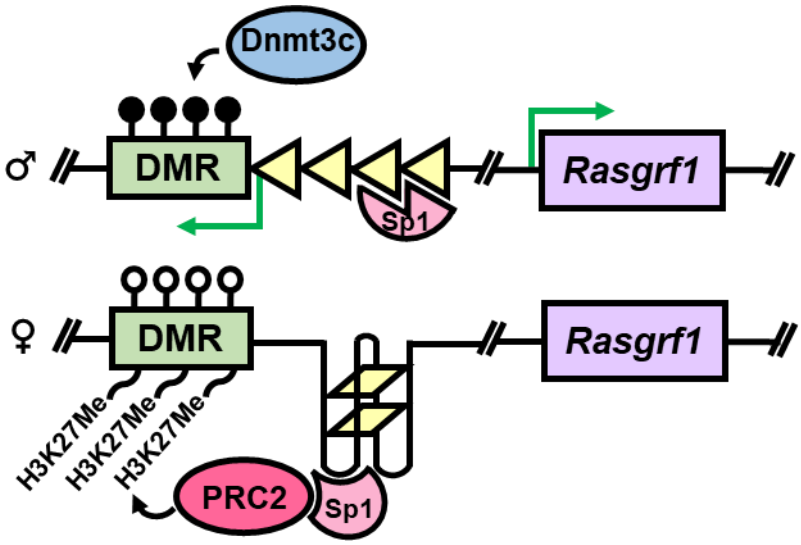
In Chapters II and IV, additional components of the mechanism underlying DNA methylation at *Rasgrf1* were exposed: First, that the repeats themselves can act in a pitRNA-independent manner to impart methylation to the DMR; and second, that Sp1 binds the repeats and drives pitRNA expression. In light of these data, a predominant question arises: By what mechanism is this *cis*-element dictating differential methylation? Potential mechanisms are discussed in Chapters II and III. Here, I will expand more on one hypothesis—that a differential G4 structure could exist between the maternal and paternal alleles. As discussed in Chapter III, the *Rasgrf1* repeats are predicted to form a stable G-Quadruplex (G4). G4 structures require at least two guanine-rich tracts (five to seven consecutive guanines) separated by three to five nucleotides. The guanine tracts form a secondary planar arrangement via Hoogsteen hydrogen bonds, (Dipankar and Gilbert, 1988). Genome-wide, G4s appear to mark nucleosome depleted, regulatory chromatin; enhanced G4 formation is associated with increased local transcription (Hansel-Hertsch *et al,* 2016), and is negatively associated with DNA methylation (Halder *et al,* 2010).

At specific loci, G4s have assumed both repressive and activating roles (Siddiqui-Jain *et al*, 2002; Cogoi and Xodo, 2006; Rankin *et al*, 2005). G4s have been demonstrated to be biologically relevant: As targets for telomerase (Moye *et al*, 2015), they have implications for aging and senescence, Further, they have drawn attention as a target for cancer therapy (Riou *et al*, 2002; Zhou *et al*, 2006). G4s have

also been implicated as a class of *cis*-elements that could help direct the nuanced temporospatial events that occur in embryonic development (David *et al*, 2016). Even more recently, a G4 has been characterized at the imprinted *H19* locus and requires Sp1 binding for suppression of *H19* expression (Fukuhara *et al*, 2017).

The *Rasgrf1* repeats harbor 112 runs of GGGG tetrads, and is therefore predicted to form a stable G4. Notably, Sp1 binds both its canonical 5'-GGGCGG-3' sequence as well as G4 structures. As such, a potential mechanism could involve differential G4 formation and Sp1 binding effects (**Fig VI.1**).

**Fig VI.1. Potential model for the mechanism of methylation at *Rasgrf1*.** Male germline (upper diagram). Sp1 binds the repeats, which do not assume secondary G4 structure in the embryonic male germline. This drives expression of the pitRNA but does not recruit PRC2, allowing for Dntm3c-mediated deposition of DNA methylation at the DMR. In neonatal brain, paternal methylation leads to paternal expression of *Rasgrf1*. Female germline (lower diagram). G4 formation at the repeats are bound by Sp1 which could in this configuration recruit PRC2, which deposits H3K27Me at the DMR and precludes DNA methylation. In neonatal brain, the unmethylated maternal allele is not expressed.




Several aspects of the model suggested above represent just one of many

possible permutations. First, G4 formation could occur in the male rather than the female germline, as G4 have also been shown to activate gene expression. Second, Sp1 may not be present at the repeats in the female germline. However, the experiments to test this model in all of its variants remain the same. First, differential G4 formation at the repeats in the female and male germlines could be assayed. Classic methods involve circular dichroism (Giraldo *et al*, 1994), though an alternative approach that may be more amenable to comparing the male and female states would involve G4 ChIP-qPCR for the repeats. Anti-G4 antibodies have been generated (Biffi *et al*, 2013) and are now commercially available (Millipore MABE1126). Second, Sp1 binding should be assayed in the female germline. Should Sp1 bind the repeats in the female germline, physical interactions between Sp1 and PRC2 should be assayed as a potential link between the *Rasgrf1* repeats, PRC2, and the deposition of H3K27Me on the maternal allele.

## VI.B Summary: Imprinted loci as model systems for *cis* regulation of DNA methylation

As genes that are defined by a parent-of-origin-specific pattern of DNA methylation, imprinted loci have proven to be extremely valuable in the identification of *cis* elements that direct DNA methylation. However, the mechanisms by which DNA methylation is directed remain relatively obscure. Certainly, as reviewed in Chapter IV, lncRNAs have been implicated as molecular bridges and scaffolds upon DNA methylatranferases and other effectors of epigenetic state might build, including the pitRNA at *Rasgrf1*. However, our work work in combination with others as

described in Chapter II, support a requirement for the *cis* element, rather than the lncRNA transcript that it drives. How the *cis* element at *Rasgrf1* might function separately of the pitRNA was the focus of discussion in Chapters II, III, and Section VI.A. In general, how *cis* elements might function to recruit effectors of epigenetic state require a transcription factor that recognize *cis* element sequence or potentially secondary structure, as demonstrated at *H19* (Fukuhara *et al*, 2017) and as we speculate upon at *Rasgrf1*. Returning to the former, a transcription factor might influence DNA methylation by binding unmethylated DNA and precluding or promoting DNA methyltransferase activity. Much *in vitro* data exist in support of both of these mechanisms (reviewed in Blattler and Farnam, 2013), but it largely remains to be seen whether any or all of these mechanisms act *in vivo* during crucial times of embryonic reprogramming.

An additional, upstream consideration includes preexisting chromatin structure. As discussed in Chapter I and demonstrated in Chapter II, chromatin architecture can and does impact the extent of a strong transcriptional activator's effects. Reducing the local effects of a transcriptional activator by maintaining distinct regions of chromatin interactions could also allow for the encroachment of repressive complexes—this has already been demonstrated in some cases, as loss of the histone demethylase LSD1 leads to loss of methylation at transposable element sequences, likely due to an increase in methylated H3K4 residues, which are typically marks of transcriptional activation, but which reduce DNMT binding (Wang *et al*, 2009). Alternatively, a compact versus a more permissive chromatin domain might preclude pioneer factors from binding target sequences and recruiting DNA

methylation. Certainly, alternate chromatin environments differ between the sexes, and could contribute to differential methylation at imprinted loci. Mechanisms that regulate chromatin structure are already well known in mature gametes (De La Fuente *et al*, 2001; Zhao *et al*, 2004; Burns *et al*, 2003) and could be at play during sex determination: For example, the male sex-specific factors SRY and SF1 activate an enhancer of Sox9, which is known to interact with p300 to regulate transcription and chromatin structure (Sekido *et al*, 2008; Furumatsu *et al*, 2005).

### VI.C Differential DNA methylation in the dog: Potential predictors of ancestry, age, and disease

The data presented in Chapter V explore differential methylation between two phenotypically diverse dog breeds, the Yorkshire Terrier and the Great Dane. Having examined just two individuals of two different dog breeds, this dataset may have limited power to identify biologically relevant DMRs. However, within this dataset, we have identified two potentially informative DMRs, one within the *GLUD1* gene, known to be involved in glucose homeostasis, and one within the *MRC2* gene, which is known to be upregulated in primary bone cancers such as osteosarcoma. As such, our data in combination with the work of others indicate that the canine methylome has great potential as a rich source of variation for other traits.

Recently, a large cohort of domestic dogs and wolves were sequenced by mRRBS (Koch *et al*, 2016). Multiple wolf-dog single methyl polymorphisms (SMPs) were identified, though a functional basis of any of these DMRs remain to be characterized. From this same data set, a “molecular clock” was proposed using 67 and 41 SMPs in dogs and wolves respectively (Thompson *et al*, 2017). While much

remains to be discovered in the functional activity of these annotated SMPs, in human, genetic-epigenetic interactions have been implicated in conditions including oral cancer (Chan *et al*, 2008), Friedrich's ataxia (Al-Madahwi *et al*, 2008), opiate addiction susceptibility (Oertal *et al*, 2012) and Graves' disease (Stefan *et al*, 2014). In the context of haplotype-dependent allele-specific methylation, these SMPs may prove as informative in dogs as they have in humans (reviewed in Do *et al*, 2017).

In our own work, we chose to examine differentially methylated regions composed of dense stretches of CpG dinucleotides, or CpG islands which RRBS captures efficiently (Gu *et al*, 2011) and which have known potential for transcriptional regulation (discussed in Chapter I). Our limited data set, composed of four dogs, have indicated the potential utility of DMRs in explaining phenotypic variation in dogs. To date, the dataset used by Thompson and Koch has yet to be mined for DMRs that differ between the fifty dogs representing nineteen different breeds, among them breeds of unique ancestry and appearance. Ancestry and relatedness amongst dog breeds and wild canids is extensively characterized (Lindblad-Toh *et al*, 2005; vonHoldt *et al*, 2010; Cronin *et al*, 2015; Parker *et al*, 2017); however, whether DMRs can predict ancestral relationships, and to what resolution of relationships, is unknown, and could likely involve hierarchical clustering of mRRBS samples of different breeds. Such an experiment could also indicate the function of identified DMRs in driving dog breed diversification, which has fixed predispositions for behavior and disease in the domestic dog.

The methylome of mixed breed dogs also carries immense potential, but has not been explored by us or others. Purebred dogs are typically emphasized in canine

genomics due to high degrees of inbreeding and trait fixation, but some studies do demonstrate the utility of mixed breed dogs, either in confirming findings from purebreds, such as mutations contributing to body size (Boyko *et al*, 2010), or aiding in the discovery of trait-associated mutations themselves (Huson *et al*, 2011). In many prevalent conditions, most with a known genetic contribution, mixed breed dogs are as susceptible as purebreds (Bellumori *et al*, 2013). Given that over half of the owned dogs in the United States are of mixed breed ancestry (AVMA, 2012), the predictiveness of DNA methylation for disease or behavior predisposition of mixed breed dogs could be incredibly valuable.

## REFERENCES

### CHAPTER I


Alexander KA, Wang X, Shibata M, Clark AG, García-García MJ. TRIM28 controls genomic imprinting through distinct mechanisms during and after early genome-wide reprogramming. Cell reports. 2015;13(6):1194-1205. doi:10.1016/j.celrep.2015.09.078.

Algar, Elizabeth, Samantha Brickell, Gillian Deeble, David Amor, and Peter Smith. Analysis of CDKN1C in Beckwith Wiedemann syndrome. Human mutation. 2000; 15(6): 497.

Alleman M, Sidorenko L, McGinnis K, Seshadri V, Dorweiler JE, White J, *et al*. An RNA-dependent RNA polymerase is required for paramutation in maize. Nature 2006, 442: 295-8.

Allen E, Horvath S, Tong F, Kraft P, Spiteri E, Riggs AD, Marahrens Y. High concentrations of long interspersed nuclear element sequence distinguish monoallelically expressed genes. Proceedings of the National Academy of Sciences. 2003 Aug 19;100(17):9940-5.

Amir RE, Van den Veyver IB, Wan M, Tran CQ, Francke U, Zoghbi HY. Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. Nat Genet. 1999 Oct;23(2):185-8.

Andrews SC, Wood MD, Tunster SJ, Barton SC, Surani MA, John RM. Cdkn1c (p57 Kip2) is the major regulator of embryonic growth within its imprinted domain on mouse distal chromosome 7. BMC Developmental Biology. 2007 May 21;7(1):53.

Andrey G, Montavon T, Mascrez B, Gonzalez F, Noordermeer D, Leleu M, Trono D, Spitz F, Duboule D. A switch between topological domains underlies HoxD genes collinearity in mouse limbs. Science. 2013 Jun 7;340(6137):1234167.

Anne J, Ollo R, Ephrussi A, Mechler BM. Arginine methyltransferase Capsuleen is essential for methylation of spliceosomal Sm proteins and germ cell formation in Drosophila. Development. 2007 Jan 1;134(1):137-46.

Aravin A, Gaidatzis D, Pfeffer S, Lagos-Quintana M, Landgraf P, Iovino N, Morris P, Brownstein MJ, Kuramochi-Miyagawa S, Nakano T, Chien M. A novel class of small RNAs bind to MILI protein in mouse testes. Nature. 2006 Jul 13;442(7099):203.

Aravin AA, Hannon GJ, Brennecke J. The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. science. 2007 Nov 2;318(5851):761-4.

Aravin AA, Sachidanandam R, Bourc'his D, *et al*. A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. Molecular cell. 2008;31(6):785-799. doi:10.1016/j.molcel.2008.09.003.

Aravin AA, Van Der Heijden GW, Castañeda J, Vagin VV, Hannon GJ, Bortvin A. Cytoplasmic compartmentalization of the fetal piRNA pathway in mice. PLoS genetics. 2009 Dec 11;5(12):e1000764.

Arrighi FE and Hsu TC. Localization of heterochromatin in human chromosomes. Cytogenetics 1971;10:81–86. DOI:10.1159/000130130.

Arteaga-Vazquez MA, Chandler VL. Paramutation in maize: RNA mediated trans-generational gene silencing. Curr Opin Genet Dev. 2010;20: 156–163. pmid:20153628

Ashe A, Morgan DK, Whitelaw NC, *et al*. A genome-wide screen for modifiers of transgene variegation identifies genes with critical roles in development. Genome Biology. 2008;9(12):R182. doi:10.1186/gb-2008-9-12-r182.

Azuara V, Perry P, Sauer S, Spivakov M, Jorgensen HF, John RM, Gouti M, Casanova M, Warnes G, Merkenschlager M *et al* (2006) Chromatin signatures of pluripotent cell lines. Nat Cell Biol 8: 532–538

Baccarelli A, Wright R, Bollati V, *et al*. Ischemic Heart Disease and Stroke in Relation to Blood DNA Methylation. Epidemiology (Cambridge, Mass). 2010;21(6):819-828. doi:10.1097/EDE.0b013e3181f20457.

Baguisi A, Behboodi E, Melican DT, Pollock JS, Destrempes MM, Cammuso C, Williams JL, Nims SD, Porter CA, Midura P, Palacios MJ. Production of goats by somatic cell nuclear transfer. Nature biotechnology. 1999 May 1;17(5).

Balhorn R. The protamine family of sperm nuclear proteins. Genome Biology. 2007;8(9):227. doi:10.1186/gb-2007-8-9-227.

Baniahmad A., Steiner, C., Kohne, A. C. & Renkawitz, R. Modular structure of a chicken lysozyme silencer: involvement of an unusual thyroid hormone receptor binding site. Cell 1990 61, 505–514

Banine F, Bartlett C, Gunawardena R, Muchardt C, Yaniv M, Knudsen ES, Weissman BE, Sherman LS. SWI/SNF chromatin-remodeling factors induce changes in DNA methylation to promote transcriptional activation. Cancer research. 2005 May 1;65(9):3542-7.

Bannister AJ, Zegerman P, Partridge JF, Miska EA, Thomas JO, Allshire RC, Kouzarides T. Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. Nature. 2001 Mar 1;410(6824):120-4.

Barau J, Teissandier A, Zamudio N, Roy S, Nalesso V, Herault Y, *et al* (2016). The DNA methyltransferase DNMT3C protects male germ cells from transposon activity. Science 354(6314): 909-912.

Bartolomei MS, Ferguson-Smith AC. Mammalian genomic imprinting. Cold Spring Harbor perspectives in biology. 2011 Jul 1;3(7):a002592.

Bashtrykov P, Jankevicius G, Smarandache A, Jurkowska RZ, Ragozin S, Jeltsch A. Specificity of Dnmt1 for methylation of hemimethylated CpG sites resides in its catalytic domain. Chem Biol. 2012 May 25;19(5):572-8. doi: 10.1016/j.chembiol.2012.03.010.

Bastepe M. The GNAS locus and pseudohypoparathyroidism. Genomic imprinting. 2008:27-40.

Bateson W, Pellew C. On the genetics of “rogues” among culinary peas (Pisum sativum). Journal of Genetics. 1915 Jul 1;5(1):13-36.

Beaujean N, Taylor J, Gardner J, Wilmut I, Meehan R, Young L. Effect of limited DNA methylation reprogramming in the normal sheep embryo on somatic cell nuclear transfer. Biology of reproduction. 2004 Jul 1;71(1):185-93.

Belele CL, Sidorenko L, Stam M, Bader R, Arteaga-Vazquez MA, Chandler VL (2013) Specific Tandem Repeats Are Sufficient for Paramutation-Induced Trans-Generational Silencing. PLoS Genet 9(10): e1003773. https://doi.org/10.1371/journal.pgen.1003773

Bell AC, Felsenfeld G. Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. Nature. 2000 May 25;405(6785):482.

Bell AC, West AG, Felsenfeld G. The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. Cell. 1999 Aug 6;98(3):387-96.

Bennett ST, Wilson AJ, Esposito L, Bouzekri N, Undlien DE, Cucca F, Nisticò L, Buzzetti R, Bosi E, Pociot F, Nerup J, Cambon-Thomsen A, Pugliese A, Shield JP, McKinney PA, Bain SC, Polychronakos C, Todd JA. Insulin VNTR allele-specific effect in type 1 diabetes depends on identity of untransmitted paternal allele. The IMDIAB Group. Nat Genet. 1997 Nov;17(3):350-2.

Benton MC, Johnstone A, Eccles D, *et al*. An analysis of DNA methylation in human adipose tissue reveals differential modification of obesity genes before and after gastric bypass and weight loss. Genome Biology. 2015;16(1):8. doi:10.1186/s13059-014-0569-x.

Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, Fry B, Meissner A, Wernig M, Plath K, Jaenisch R, Wagschal A, Feil R, Schreiber SL, Lander ES. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell. 2006 Apr 21;125(2):315-26.

Bestor TH, Edwards JR, Boulard M. Notes on the role of dynamic DNA methylation in mammalian development. Proceedings of the National Academy of Sciences of the United States of America. 2015;112(22):6796-6799. doi:10.1073/pnas.1415301111.

Bestor TH. Activation of mammalian DNA methyltransferase by cleavage of a Zn binding regulatory domain. The EMBO Journal. 1992;11(7):2611-2617.

Bhutani N, Brady JJ, Damian M, Sacco A, Corbel SY, and Blau HM. Reprogramming towards pluripotency requires AID-dependent DNA demethylation. Nature 2010,463(7284): 1042-1047.

Binda O, LeRoy G, Bua DJ, Garcia BA, Gozani O, Richard S. Trimethylation of histone H3 lysine 4 impairs methylation of histone H3 lysine 9: Regulation of lysine methyltransferases by physical interaction with their substrates. Epigenetics. 2010;5(8):767-775. doi:10.4161/epi.5.8.13278.

Blelloch R, Wang Z, Meissner A, Pollard S, Smith A, Jaenisch R. Reprogramming efficiency following somatic cell nuclear transfer is influenced by the differentiation and methylation state of the donor nucleus. Stem Cells. 2006 Sep 1;24(9):2007-13.

Bolzer A, Kreth G, Solovei I, Koehler D, Saracoglu K, Fauth C, Müller S, Eils R, Cremer C, Speicher MR, Cremer T. Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes. PLoS biology. 2005 Apr 26;3(5):e157.

Borgel J, Guibert S, Li Y, Chiba H, Schübeler D, Sasaki H, Forné T, Weber M. Targets and dynamics of promoter DNA methylation during early mouse development. Nat Genet. 2010 Dec;42(12):1093-100. doi: 10.1038/ng.708.

Boros J, Arnoult N, Stroobant V, Collet J-F, Decottignies A. Polycomb Repressive Complex 2 and H3K27me3 Cooperate with H3K9 Methylation To Maintain Heterochromatin Protein 1α at Chromatin. Molecular and Cellular Biology. 2014;34(19):3662-3674. doi:10.1128/MCB.00205-14.az

Bostick M, Kim JK, Estève PO, Clark A, Pradhan S, Jacobsen SE. UHRF1 plays a role in maintaining DNA methylation in mammalian cells. Science. 2007 Sep 21;317(5845):1760-4.

Bourc'his D, Xu GL, Lin CS, Bollman B, and Bestor TH (2001). Dnmt3L and the establishment of maternal genomic imprints. Science 294(5551): 2536-9.

Bourc'his D, Bestor TH. Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. Nature. 2004 Sep 2;431(7004):96-9.

Brattås PL, Jönsson ME, Fasching L, Nelander Wahlestedt J, Shahsavani M, Falk R, Falk A, Jern P, Parmar M, Jakobsson J. TRIM28 Controls a Gene Regulatory

Network Based on Endogenous Retroviruses in Human Neural Progenitor Cells. Cell Rep. 2017 Jan 3;18(1):1-11. doi: 10.1016/j.celrep.2016.12.010.f

Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, Sachidanandam R, Hannon GJ. Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. Cell. 2007 Mar 23;128(6):1089-103.

Brink RA. A Genetic Change Associated with the R Locus in Maize Which Is Directed and Potentially Reversible. Genetics. 1956;41(6):872-889.

Brower-Toland B, Findley SD, Jiang L, *et al*. Drosophila PIWI associates with chromatin and interacts directly with HP1a. Genes & Development. 2007;21(18):2300-2311. doi:10.1101/gad.1564307.

Brzeska K, Brzeski J, Smith J, Chandler VL. Transgenic expression of CBBP, a CXC domain protein, establishes paramutation in maize. Proceedings of the National Academy of Sciences of the United States of America. 2010;107(12):5516-5521. doi:10.1073/pnas.1001576107.

Buiting K, Groß S, Lich C, Gillessen-Kaesbach G, El-Maarri O, Horsthemke B. Epimutations in Prader-Willi and Angelman syndromes: a molecular study of 136 patients with an imprinting defect. The American Journal of Human Genetics. 2003 Mar 31;72(3):571-7.

Busslinger GA, Stocsits RR, van der Lelij P, Axelsson E, Tedeschi A, Galjart N, Peters JM. Cohesin is positioned in mammalian genomes by transcription, CTCF and Wapl. Nature. 2017 Apr 27;544(7651):503-7.

Campbell KH, McWhir J, Ritchie WA, Wilmut I. Sheep cloned by nuclear transfer from a cultured cell line. Nature. 1996 Mar 7;380(6569):64.

Canapa A, Barucca M, Biscotti MA, Forconi M, Olmo E. Transposons, genome size, and evolutionary insights in animals. Cytogenetic and genome research. 2015;147(4):217-39.

Capovilla M, Robichon A, Rassoulzadegan M. A new paramutation-like example at the Delta gene of Drosophila. Skoulakis EMC, ed. PLoS ONE. 2017;12(3):e0172780. doi:10.1371/journal.pone.0172780.

Carmell MA, Girard A, van de Kant HJ, Bourc'his D, Bestor TH, de Rooij DG, Hannon GJ. MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline. Dev Cell. 2007 Apr;12(4):503-14. Epub 2007 Mar 29.

Catcheside DG. TheP-locus position effect inOenothera. Journal of genetics. 1947 Apr 1;48(1):31-42.

Cattanach BM. Parental origin effects in mice. Development. 1986 Oct 1;97(Supplement):137-50.

Chang J, Zhang B, Heath H, Galjart N, Wang X, Milbrandt J. Nicotinamide adenine dinucleotide (NAD)–regulated DNA methylation alters CCCTC-binding factor (CTCF)/cohesin binding and transcription at the BDNF locus. Proceedings of the National Academy of Sciences. 2010 Dec 14;107(50):21836-41.

Chang Y, Sun L, Kokura K, *et al*. MPP8 mediates the interactions between DNA methyltransferase Dnmt3a and H3K9 methyltransferase GLP/G9a. Nature Communications. 2011;2:533. doi:10.1038/ncomms1549.

Chen BH, Marioni RE, Colicino E, *et al*. DNA methylation-based measures of biological age: meta-analysis predicting time to death. Aging (Albany NY). 2016;8(9):1844-1859. doi:10.18632/aging.101020.

Chen Z, Hagen DE, Elsik CG, Ji T, Morris CJ, Moon LE, Rivera RM. Characterization of global loss of imprinting in fetal overgrowth syndrome induced by assisted reproduction. Proceedings of the National Academy of Sciences. 2015 Apr 14;112(15):4618-23.

Chénais B, Caruso A, Hiard S, Casse N. The impact of transposable elements on eukaryotic genomes: from genome size increase to genetic adaptation to stressful environments. Gene. 2012 Nov 1;509(1):7-15.

Cheng JC, Yoo CB, Weisenberger DJ, Chuang J, Wozniak C, Liang G, Marquez VE, Greer S, Orntoft TF, Thykjaer T, Jones PA. Preferential response of cancer cells to zebularine. Cancer Cell. 2004 Aug;6(2):151-8.

Chotalia M, Smallwod SA, Ruf N, Dawson C, Lucifero DFrontera M, *et al*. Transcription is required for establishment of germline methylation marks at imprinted genes. Genes and Development 2008, 23: 105-117.

Chuma S, Hosokawa M, Kitamura K, *et al*. Tdrd1/Mtr-1, a tudor-related gene, is essential for male germ-cell differentiation and nuage/germinal granule formation in mice. Proceedings of the National Academy of Sciences of the United States of America. 2006;103(43):15894-15899. doi:10.1073/pnas.0601878103.

Cirio MC, Martel J, Mann M, *et al*. DNA methyltransferase 1o functions during preimplantation development to preclude a profound level of epigenetic variation. Developmental biology. 2008;324(1):139-150. doi:10.1016/j.ydbio.2008.09.015.

Clapcott SJ, Peters J, Orban PC, Brambilla R, Graham CF (2003) Two ENU-induced mutations in *Rasgrf1* and early mouse growth retardation. Mamm Genome 14:495–505:125–126

Clapier CR, Iwasa J, Cairns BR, Peterson CL. Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. Nat Rev Mol Cell Biol. 2017 Jul;18(7):407-422. doi: 10.1038/nrm.2017.26.

Clapier CR, Längst G, Corona DF, Becker PB, Nightingale KP. Critical role for the histone H4 N terminus in nucleosome remodeling by ISWI. Molecular and Cellular Biology. 2001 Feb 1;21(3):875-83.

Clark LA, Wahl JM, Rees CA, Murphy KE. Retrotransposon insertion in SILV is responsible for merle patterning of the domestic dog. Proceedings of the National Academy of Sciences of the United States of America. 2006;103(5):1376-1381. doi:10.1073/pnas.0506940103.

Coppieters N, Dieriks BV, Lill C, Faull RL, Curtis MA, Dragunow M. Global changes in DNA methylation and hydroxymethylation in Alzheimer's disease human brain. Neurobiol Aging. 2014 Jun;35(6):1334-44. doi: 10.1016/j.neurobiolaging.2013.11.031.

Cortellino S, Xu J, Sannai M, Moore R, Carette E, Cigliano A, *et al*.Thymine DNA Glycosylase Is Essential for Active DNA Demethylation by Linked Deamination-Base Excision Repair. Cell 2011, 146(1): 67-79.

Cox DN, Chao A, Baker J, Chang L, Qiao D, Lin H. A novel class of evolutionarily conserved genes defined by piwi are essential for stem cell self-renewal. Genes & development. 1998 Dec 1;12(23):3715-27.

Creyghton MP, Cheng AW, Welstead GG, *et al*. Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proceedings of the National Academy of Sciences of the United States of America. 2010;107(50):21931-21936. doi:10.1073/pnas.1016071107.

Croft JA, Bridger JM, Boyle S, Perry P, Teague P, Bickmore WA. Differences in the localization and morphology of chromosomes in the human nucleus. The Journal of cell biology. 1999 Jun 14;145(6):1119-31.

Crouse HV. The controlling element in sex chromosome behavior in Sciara. Genetics. 1960 Oct;45(10):1429.

Cui K, Zang C, Roh TY, Schones DE, Childs RW, Peng W, Zhao K. Chromatin Signatures in Multipotent Human Hematopoietic Stem Cells Indicate the Fate of Bivalent Genes during Differentiation. Cell Stem Cell 9 January 2009, 4(1): 80-93. https://doi.org/10.1016/j.stem.2008.11.011

Cutter A, Hayes JJ. A Brief Review of Nucleosome Structure. FEBS letters. 2015;589(20 0 0):2914-2922. doi:10.1016/j.febslet.2015.05.016.

Darricarrère N, Liu N, Watanabe T, Lin H. Function of Piwi, a nuclear Piwi/Argonaute protein, is independent of its slicer activity. Proceedings of the National Academy of Sciences. 2013 Jan 22;110(4):1297-302.

Davis TL, Trasler JM, Moss SB, Yang GJ, Bartolomei MS. Acquisition of theH19Methylation imprint occurs differentially on the parental alleles during spermatogenesis. Genomics. 1999 May 15;58(1):18-28.

Dayeh T, Volkov P, Salö S, Hall E, Nilsson E, Olsson AH, *et al*. (2014) Genome-Wide DNA Methylation Analysis of Human Pancreatic Islets from Type 2 Diabetic and Non-Diabetic Donors Identifies Candidate Genes That Influence Insulin Secretion. PLoS Genet 10(3): e1004160. https://doi.org/10.1371/journal.pgen.1004160

De Fazio S, Bartonicek N, Di Giacomo M, Abreu-Goodger C, Sankar A, Funaya C, Antony C, Moreira PN, Enright AJ, O'Carroll D. The endonuclease activity of Mili fuels piRNA amplification that silences LINE1 elements. Nature. 2011 Oct 23;480(7376):259-63. doi: 10.1038/nature10547.

De Jager P, Srivastava G, Lunnon K, *et al*. Alzheimery's disease pathology is associated with early alterations in brain DNA methylation at ANK1, BIN1, RHBDF2 and other loci. Nature neuroscience. 2014;17(9):1156-1163. doi:10.1038/nn.3786.

de Vanssay A, Boug AL, Boivin A, Hermant C, Teysset L, Delmarre V, *et al*. Paramutation in Drosophila linked to emergence of a piRNA-producing locus. Nature 2012, 490: 112-5.

de Vanssay A, Bougé A-L, Boivin A, Hermant C, Teysset L, Delmarre V, *et al*. Paramutation in Drosophila linked to emergence of a piRNA-producing locus. Nature. 2012;490: 112–115. pmid:22922650

de Wit E, Vos ES, Holwerda SJ, Valdes-Quezada C, Verstegen MJ, Teunissen H, Splinter E, Wijchers PJ, Krijger PH, de Laat W. CTCF binding polarity determines chromatin looping. Molecular cell. 2015 Nov 19;60(4):676-84.

DeBarry JD, Ganko EW, McCarthy EM, McDonald JF. The contribution of LTR retrotransposon sequences to gene evolution in Mus musculus. Molecular biology and evolution. 2005 Dec 20;23(3):479-81.

Deng W, Lin H. Miwi, a murine homolog of piwi, encodes a cytoplasmic protein essential for spermatogenesis. Developmental cell. 2002 Jun 30;2(6):819-30.

Dewannieux M, Dupressoir A, Harper F, Pierron G, Heidmann T. Identification of autonomous IAP LTR retrotransposons mobile in mammalian cells. Nat Genet. 2004 May;36(5):534-9.

Dindot SV, Person R, Strivens M, Garcia R, Beaud*et al*. Epigenetic profiling at mouse imprinted gene clusters reveals novel epigenetic and genetic features at differentially methylated regions. Genome research. 2009 Aug 1;19(8):1374-83.

Dixon JR, Jung I, Selvaraj S, Shen Y, Antosiewicz-Bourget JE, Lee AY, Ye Z, Kim A, Rajagopal N, Xie W, Diao Y. Chromatin architecture reorganization during stem cell differentiation. Nature. 2015 Feb 19;518(7539):331.

Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature. 2012 May 17;485(7398):376.

Dorweiler JE, Carey CC, Kubo KM, Hollick JB, Kermicle JL, Chandler VL. mediator of paramutation1 Is Required for Establishment and Maintenance of Paramutation at Multiple Maize Loci. The Plant Cell. 2000;12(11):2101-2118.

Drake NM, DeVito LM, Cleland TA, Soloway PD. Imprinted *Rasgrf1* expression in neonatal mice affects olfactory learning and memory. Genes, brain, and behavior. 2011;10(4):392-403. doi:10.1111/j.1601-183X.2011.00678.x.

Ecco G, Cassano M, Kauzlaric A, *et al*. Transposable elements and their KRAB-ZFP controllers regulate gene expression in adult tissues. Developmental cell. 2016;36(6):611-623. doi:10.1016/j.devcel.2016.02.024.

Ecco G, Cassano M, Kauzlaric A, *et al*. Transposable elements and their KRAB-ZFP controllers regulate gene expression in adult tissues. Developmental cell. 2016;36(6):611-623. doi:10.1016/j.devcel.2016.02.024.

Edmunds JW, Mahadevan LC, Clayton AL. Dynamic histone H3 methylation during gene induction: HYPB/Setd2 mediates all H3K36 trimethylation. The EMBO Journal. 2008;27(2):406-420. doi:10.1038/sj.emboj.7601967.

Erhard KF, Stonaker JL, Parkinson SE, Lim JP, Hale CJ, Hollick JB. RNA polymerase IV functions in paramutation in Zea mays. Science. 2009 Feb 27;323(5918):1201-5.

Fang J, Cheng J, Wang J, *et al*. Hemi-methylated DNA opens a closed conformation of UHRF1 to facilitate its histone recognition. Nature Communications. 2016;7:11197. doi:10.1038/ncomms11197.

Fasching L, Kapopoulou A, Sachdeva R, *et al*. TRIM28 Represses Transcription of Endogenous Retroviruses in Neural Progenitor Cells. Cell reports. 2015;10(1):20-28. doi:10.1016/j.celrep.2014.12.004.

Fedoriw AM, Stein P, Svoboda P, Schultz RM, Bartolomei MS. Transgenic RNAi reveals essential function for CTCF in H19 gene imprinting. Science. 2004 Jan 9;303(5655):238-40.

Feinberg AP, Vogelstein B. Hypomethylation distinguishes genes of some human cancers from their normal counterparts. Nature 1983 Jan 6;301(5895):89-92.

Feinberg AP, Vogelstein B. Hypomethylation of ras oncogenes in primary human cancers. Biochemical and biophysical research communications. 1983 Feb 28;111(1):47-54.

Ferguson-Smith AC, Sasaki H, Cattanach BM, Surani MA. Parental-origin-specific epigenetic modification of the mouse H19 gene. Nature. 1993 Apr 22;362(6422):751.

Frohlich LF, Mrakovcic M, Steinborn R, Chung U, Bastepe M, and Juppner H. Targeted deletion of the Nesp55 DMR defines another Gnas imprinting control region and provides a mouse model of autosomal dominant PHP-Ib. PNAS 2010, 107(20): 9275-9280.

Fuks F, Burgers WA, Brehm A, Hughes-Davies L, Kouzarides T. DNA methyltransferase Dnmt1 associates with histone deacetylase activity. Nat Genet. 2000 Jan;24(1):88-91.

Galli C, Lagutina I, Crotti G, Colleoni S, Turini P, Ponderato N, Duchi R, Lazzari G. Pregnancy: a cloned horse born to its dam twin. Nature. 2003 Aug 7;424(6949):635-.

Gardner RJ, Mackay DJ, Mungall AJ, Polychronakos C, Siebert R, Shield JP, Temple IK, Robinson DO. An imprinted locus associated with transient neonatal diabetes mellitus. Human molecular genetics. 2000 Mar 1;9(4):589-96. Arima T, Drewell RA, Arney KL, Inoue J, Makita Y, Hata A, Oshimura M, Wake N,

Gicquel C, Rossignol S, Cabrol S, Houang M, Steunou V, Barbu V, Danton F, Thibaud N, Le Merrer M, Burglen L, Bertrand AM. Epimutation of the telomeric imprinting center region on chromosome 11p15 in Silver-Russell syndrome. Nature genetics. 2005 Sep 1;37(9):1003.

Giese KP, Friedman E, Telliez J, Fedorov NB, Wines M *et al* (2001) Hippocampus-dependent learning and memory is impaired in mice lacking the Ras-guanine-nucleotide releasing factor 1 (Ras-Grf1). Neuropharmacology 41:791–800

Gifford WD, Pfaff SL, Macfarlan TS. Transposable elements as genetic regulatory substrates in early development. Trends in cell biology. 2013;23(5):218-226. doi:10.1016/j.tcb.2013.01.001.

Girard A, Sachidanandam R, Hannon GJ, Carmell MA. A germline-specific class of small RNAs binds mammalian Piwi proteins. Nature. 2006 Jul 13;442(7099):199-202.

Goelz SE, Vogelstein B, Hamilton SR, Feinberg AP. Hypomethylation of DNA from benign and malignant human colon neoplasms. Science. 1985 Apr 12;228(4696):187-90.

Goh WSS, Falciatori I, Tam OH, *et al*. piRNA-directed cleavage of meiotic transcripts regulates spermatogenesis. Genes & Development. 2015;29(10):1032-1044. doi:10.1101/gad.260455.115.

Gonsalvez GB, Tian L, Ospina JK, Boisvert FM, Lamond AI, Matera AG. Two distinct arginine methyltransferases are required for biogenesis of Sm-class ribonucleoproteins. J Cell Biol. 2007 Aug 27;178(5):733-40.

Gonzalez-Zulueta M, Bender CM, Yang AS, Nguyen T, Beart RW, Van Tornout JM, Jones PA. Methylation of the 5' CpG island of the p16/CDKN2 tumor suppressor gene in normal and transformed human tissues correlates with gene silencing. Cancer Res. 1995 Oct 15;55(20):4531-5.

Gou L-T, Dai P, Yang J-H, *et al*. Pachytene piRNAs instruct massive mRNA elimination during late spermiogenesis. Cell Research. 2014;24(6):680-700. doi:10.1038/cr.2014.41.

Gowher H, Liebert K, Hermann A, Xu G, Jeltsch A. Mechanism of stimulation of catalytic activity of Dnmt3A and Dnmt3B DNA-(cytosine-C5)-methyltransferases by Dnmt3L. J Biol Chem. 2005 Apr 8;280(14):13341-8.

Grandjean V, Gounon P, Wagner N, Martin L, Wagner KD, Bernex F, *et al*. The miR-124-Sox9 paramutation: RNA-mediated epigenetic control of embryonic and adult growth. Development. 2009;136: 3647–3655. pmid:19820183

Greally JM. Short interspersed transposable elements (SINEs) are excluded from imprinted regions in the human genome. Proceedings of the National Academy of Sciences of the United States of America. 2002;99(1):327-332. doi:10.1073/pnas.012539199.

Grivna ST, Beyret E, Wang Z, Lin H. A novel class of small RNAs in mouse spermatogenic cells. Genes Dev. 2006 Jul 1;20(13):1709-14.

Gu TP, Guo F, Yang H, Wu HP, Xu GF, Liu W, Xie ZG, Shi L, He X, Jin SG, Iqbal K, Shi YG, Deng Z, Szabó PE, Pfeifer GP, Li J, Xu GL. The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. Nature. 2011 Sep 4;477(7366):606-10. doi: 10.1038/nature10443.

Guo H, Zhu P, Yan L, Li R, Hu B, Lian Y, *et al*. The DNA methylation landscape of human early embryos. Nature 511, 606–610 (31 July 2014) doi:10.1038/nature13544

Guo X, Wang L, Li J, Ding Z, Xiao J, Yin X, He S, Shi P, Dong L, Li G, Tian C. Structural insight into autoinhibition and histone H3-induced activation of DNMT3A. Nature. 2015 Jan 29;517(7536):640.

Guo Y, Xu Q, Canzio D, Shou J, Li J, Gorkin DU, *et al*. CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. Cell 2015 162(4): 900-10.

Haberland M, Montgomery RL, Olson EN. The many roles of histone deacetylases in development and physiology: implications for disease and therapy. Nature reviews Genetics. 2009;10(1):32-42. doi:10.1038/nrg2485.

Hackett JA, Sengupta R, Zylicz JJ, *et al*. Germline DNA Demethylation Dynamics and Imprint Erasure through 5-hydroxymethylcytosine. Science (New York, NY). 2013;339(6118):10.1126/science.1229277. doi:10.1126/science.1229277.

Hagarman JA, Motley MP, Kristjansdottir K, Soloway PD (2013) Coordinate Regulation of DNA Methylation and H3K27me3 in Mouse Embryonic Stem Cells. PLoS ONE 8(1): e53880. https://doi.org/10.1371/journal.pone.0053880

Hagemann R, Berg W. Paramutation at the sulfurea locus of Lycopersicon esculentum Mill. TAG Theoretical and Applied Genetics. 1978 May 1;53(3):113-23.

Hale CJ, Stonaker JL, Gross SM, Hollick JB. A Novel Snf2 Protein Maintains trans-Generational Regulatory States Established by Paramutation in Maize. Marteinssen R, ed. PLoS Biology. 2007;5(10):e275. doi:10.1371/journal.pbio.0050275.

Han BW, Wang W, Li C, Weng Z, Zamore PD. piRNA-guided transposon cleavage initiates Zucchini-dependent, phased piRNA production. Science. 2015 May 15;348(6236):817-21.

Handler D, Olivieri D, Novatchkova M, *et al*. A systematic analysis of Drosophila TUDOR domain-containing proteins identifies Vreteno and the Tdrd12 family as essential primary piRNA pathway factors. The EMBO Journal. 2011;30(19):3977-3993. doi:10.1038/emboj.2011.308.

Handler D, Olivieri D, Novatchkova M, Gruber FS, Meixner K, Mechtler K, Stark A, Sachidanandam R, Brennecke J. A systematic analysis of Drosophila TUDOR domain-containing proteins identifies Vreteno and the Tdrd12 family as essential primary piRNA pathway factors. The EMBO journal. 2011 Oct 5;30(19):3977-93.

Hannum G, Guinney J, Zhao L, *et al*. Genome-wide Methylation Profiles Reveal Quantitative Views of Human Aging Rates. Molecular cell. 2013;49(2):359-367. doi:10.1016/j.molcel.2012.10.016.

Hansen AS, Pustova I, Cattoglio C, Tjian R, Darzacq X. CTCF and cohesin regulate chromatin loop stability with distinct dynamics. Elife. 2017;6.

Hansen RS, Wijmenga C, Luo P, *et al*. The DNMT3B DNA methyltransferase gene is mutated in the ICF immunodeficiency syndrome. Proceedings of the National Academy of Sciences of the United States of America. 1999;96(25):14412-14417.

Haring M, Bader R, Louwers M, Schwabe A, van Driel R, Stam M. The role of DNA methylation, nucleosome occupancy and histone modifications in paramutation. The Plant Journal. 2010 Aug 1;63(3):366-78.

Harris CJ, Husmann D, Liu W, El Kasmi F, Wang H, Papikian A, Pastor WA, Moissiard G, Vashisht AA, Dangl JL, Wohlschlegel JA. Arabidopsis AtMORC4 and AtMORC7 form nuclear bodies and repress a large number of protein-coding genes. PLoS genetics. 2016 May 12;12(5):e1005998.

Hashimoto H, Liu Y, Upadhyay AK, *et al*. Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. Nucleic Acids Research. 2012;40(11):4841-4849. doi:10.1093/nar/gks155.

Hassan HM, Kolendowski B, Isovic M, Bose K, Dranse HJ, Sampaio AV, Underhill TM4, Torchia J. Regulation of Active DNA Demethylation through RAR-Mediated Recruitment of a TET/TDG Complex. Cell Rep. 2017 May 23;19(8):1685-1697. doi: 10.1016/j.celrep.2017.05.007.

Heintzman ND, Hon GC, Hawkins RD, *et al*. Histone Modifications at Human Enhancers Reflect Global Cell Type-Specific Gene Expression. Nature. 2009;459(7243):108-112. doi:10.1038/nature07829.

Herman H, Lu M, Anggraini M, Sikora A, Chang Y, Yoon BJ, and Soloway PD. Trans allele methylation and paramutation-like effects in mice. Nature Genetics 2003, 34(2): 199-292.

Herman JG, Graff JR, Myöhänen S, Nelkin BD, Baylin SB. Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. Proceedings of the National Academy of Sciences of the United States of America. 1996;93(18):9821-9826.

Hermann A, Goyal R, and Jeltsch A. The Dnmt1 DNA-(cytosine-C5)-methyltransferase Methylates DNA Processively with High Preference for Hemimethylated Target Sites. JBC 2004 doi: 10.1074/jbc.M403427200

Hermant C, Boivin A, Teysset L, *et al*. Paramutation in Drosophila Requires Both Nuclear and Cytoplasmic Actors of the piRNA Pathway and Induces Cis-spreading of piRNA Production. Genetics. 2015;201(4):1381-1396. doi:10.1534/genetics.115.180307.

Hiura H, Obata Y, Komiyama J, Shirai M, Kono T. Oocyte growth-dependent progression of maternal imprinting in mice. Genes to Cells. 2006 Apr 1;11(4):353-61.

Hiura H, Okae H, Miyauchi N, Sato F, Sato A, Van De Pette M, John RM, Kagami M, Nakai K, Soejima H, Ogata T. Characterization of DNA methylation errors in patients with imprinting disorders conceived by assisted reproduction technologies. Human Reproduction. 2012 Jun 6;27(8):2541-8.

Holdt LM, Hoffmann S, Sass K, Langenberger D, Scholz M, Krohn K, Finstermeier K, Stahringer A, Wilfert W, Beutner F, Gielen S. Alu elements in ANRIL non-coding RNA

at chromosome 9p21 modulate atherogenic cell functions through trans-regulation of gene networks. PLoS genetics. 2013 Jul 4;9(7):e1003588.

Hollick JB and Chandler VL. Genetic Factors Required to Maintain Repression of a Paramutagenic Maize pl1 Allele. GENETICS January 1, 2001 vol. 157 no. 1 369-378

Hollick JB, Kermicle JL, and Parkinson SE. Rmr6 Maintains Meiotic Inheritance of Paramutant States in Zea mays. GENETICS October 1, 2005 vol. 171 no. 2 725-740; doi: 10.1534/genetics.105.045260

Holmes R, Chang Y, and Soloway PD. Timing and sequence requirements defined for embryonic maintenance of imprinted DNA methylation at *Rasgrf1*. Molecular and Cell Biology 2006, 26(24):9564-70.

Horvath S. DNA methylation age of human tissues and cell types. Genome Biology. 2013;14(10):R115. doi:10.1186/gb-2013-14-10-r115.

Hou C, Li L, Qin ZS, Corces VG. Gene density, transcription, and insulators contribute to the partition of the Drosophila genome into physical domains. Molecular cell. 2012 Nov 9;48(3):471-84.

Howell, Carina Y. *et al*. Genomic Imprinting Disrupted by a Maternal Effect Mutation in the Dnmt1 Gene. Cell , Volume 104 , Issue 6 , 829 – 838

Ioannides Y, Lokulo-Sodipe K, Mackay DJ, Davies JH, Temple IK. Temple syndrome: improving the recognition of an underdiagnosed chromosome 14 imprinting disorder: an analysis of 51 published cases. Journal of medical genetics. 2014 Aug 1;51(8):495-501.

Ipsaro JJ, Haase AD, Knott SR, Joshua-Tor L, Hannon GJ. The structural biochemistry of Zucchini implicates it as a nuclease in piRNA biogenesis. Nature. 2012 Nov 8;491(7423):279.

Itier J-M, Tremp GL, Leonard J-F, Multon M-C, Ret G *et al* (1998) Imprinted gene in postnatal growth role. Nature 393

Ito S, D'Alessio AC, Taranova OV, Hong K, Sowers LC, Zhang Y. Role of Tet proteins in 5mC to 5hmC conversion, ES cell self-renewal, and ICM specification. Nature. 2010;466(7310):1129-1133. doi:10.1038/nature09303.

Iyengar S, Farnham PJ. KAP1 Protein: An Enigmatic Master Regulator of the Genome. The Journal of Biological Chemistry. 2011;286(30):26267-26276. doi:10.1074/jbc.R111.252569.

Izumi N, Shoji K, Sakaguchi Y, *et al*. Identification and functional analysis of the pre-piRNA 3′ Trimmer in silkworms. Cell. 2016;164(5):962-973. doi:10.1016/j.cell.2016.01.008.

Jacobs FM, Greenberg D, Nguyen N, *et al*. An evolutionary arms race between KRAB zinc finger genes 91/93 and SVA/L1 retrotransposons. Nature. 2014;516(7530):242-245. doi:10.1038/nature13760.

Jain D, Meydan C, Lange J, Bouuaert CC, Mason CE, Anderson K, *et al* (2017). rahu is a mutant allele of Dnmt3c, encoding a DNA methyltransferase homolog required for meiosis and transposon repression in the mouse male germline. bioRxiv, https://doi.org/10.1101/121822.

James TC, Elgin SC. Identification of a nonhistone chromosomal protein associated with heterochromatin in Drosophila melanogaster and its gene. Molecular and Cellular Biology. 1986;6(11):3862-3872.

Javierre BM, Burren OS, Wilder SP, Kreuzhuber R, Hill SM, Sewitz S, Cairns J, Wingett SW, Várnai C, Thiecke MJ, Burden F. Lineage-specific genome architecture links enhancers and non-coding disease variants to target gene promoters. Cell. 2016 Nov 17;167(5):1369-84.

Jenuwein T and Allis CD. Translating the Histone Code. Science 10 Aug 2001; 293(5532): 1074-1080. DOI: 10.1126/science.1063127

Jin B, Tao Q, Peng J, Soo HM, Wu W, Ying J, Fields CR, Delmas AL, Liu X, Qiu J, Robertson KD. DNA methyltransferase 3B (DNMT3B) mutations in ICF syndrome lead to altered epigenetic modifications and aberrant expression of genes regulating development, neurogenesis and immune function. Hum Mol Genet. 2008 Mar 1;17(5):690-709.

Jones B, Su H, Bhat A, Lei H, Bajko J, Hevi S, *et al*. The Histone H3K79 Methyltransferase Dot1L Is Essential for Mammalian Development and Heterochromatin Structure. PLoS Genet 2008 4(9): e1000190. https://doi.org/10.1371/journal.pgen.1000190

Jones P, Veenstra GJ, Wade PA, Vermaak D, Kass SU, Landsberger N, *et al*. Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. Nature 1998, 19: 187-191.

Jüppner, H., Schipani, E., Bastepe, M., Cole, D.E.C., Lawson, M.L., Mannstadt, M., Hendy, G.N., Plotkin, H., Koshiyama, H., Koh, T. and Crawford, J.D., 1998. The gene responsible for pseudohypoparathyroidism type Ib is paternally imprinted and maps in four unrelated kindreds to chromosome 20q13. 3. Proceedings of the National Academy of Sciences, 95(20), pp.11798-11803.

Kacem S, Feil R. Chromatin mechanisms in genomic imprinting. Mammalian genome. 2009 Oct 1;20(9-10):544-56.

Kagami M, Sekita Y, Nishimura G, Irie M, Kato F, Okada M, Yamamori S, Kishimoto H, Nakayama M, Tanaka Y, Matsuoka K. Deletions and epimutations affecting the

human 14q32. 2 imprinted region in individuals with paternal and maternal upd (14)-like phenotypes. Nature genetics. 2008 Feb 1;40(2):237-42.

Kagiwada S, Kurimoto K, Hirota T, Yamaji M, Saitou M. Replication-coupled passive DNA demethylation for the erasure of genome imprints in mice. The EMBO Journal. 2013;32(3):340-353. doi:10.1038/emboj.2012.331.

Kaneda M, Okano M, Hata K, Sado T, Tsujimoto N, Li E, Sasaki H.Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. Nature. 2004 Jun 24;429(6994):900-3.

Kantor B, Kaufman Y, Makedonski K, Razin A, and Shemer R.Establishing the epigenetic status of the Prader–Willi/Angelman imprinting center in the gametes and embryo. Human Molecular Genetics 2004, 12(22): 2767-2779.

Kapusta A, Feschotte C. Volatile evolution of long noncoding RNA repertoires: mechanisms and biological implications. Trends in Genetics. 2014 Oct 31;30(10):439-52.

Kato Y, Tani T, Sotomaru Y, Kurokawa K, Kato JY, Doguchi H, Yasue H, Tsunoda Y. Eight calves cloned from somatic cells of a single adult. Science. 1998 Dec 11;282(5396):2095-8.

Kato Y, Kaneda M, Hata K, Kumaki K, Hisano M, Kohara Y, *et al* (2007). Role of the Dnmt3 family in de novo methylation of imprinted and repetitive sequences during male germ cell development in the mouse. Human Molecular Genetics 16(19): 2272-80.

Kawahara M, Wu Q, Takahashi N, Morita S, Yamada K, Ito M, Ferguson-Smith AC, Kono T. High-frequency generation of viable mice from engineered bi-maternal embryos. Nature biotechnology. 2007 Sep 1;25(9):1045.

Kawaoka S, Izumi N, Katsuma S, Tomari Y. 3′ end formation of PIWI-interacting RNAs in vitro. Molecular cell. 2011 Sep 16;43(6):1015-22.

Kawasaki Y, Lee J, Matsuzasa A, Kohda T, *et al*. Active DNA demethylation is required for complete imprint erasure in primordial germ cells. Scientific Reports 2014 4, Article number: 3658 doi:10.1038/srep03658

Ke X, Thomas SN, Robinson DO, Collins A. The distinguishing sequence characteristics of mouse imprinted genes. Mammalian genome. 2002 Nov 21;13(11):639-45.

Kehle J, Beuchle D, Treuheit S, Christen B, Kennison JA, Bienz M, Müller J. dMi-2, a hunchback-interacting protein that functions in polycomb repression. Science. 1998 Dec 4;282(5395):1897-900.

Kermicle JL, Eggleston WB, Alleman M. Organization of paramutagenicity in R-stippled maize. Genetics. 1995 Sep 1;141(1):361-72.

Kim S, Günesdogan U, Zylicz JJ, Hackett JA, Cougot D, Bao S, Lee C, Dietmann S, Allen GE, Sengupta R, Surani MA. PRMT5 protects genomic integrity during global DNA demethylation in primordial germ cells and preimplantation embryos. Molecular cell. 2014 Nov 20;56(4):564-79.

Kirino Y, Kim N, de Planell-Saguer M, Khandros E, Chiorean S, Klein PS, Rigoutsos I, Jongens TA, Mourelatos Z. Arginine methylation of Piwi proteins, catalyzed by dPRMT5, is required for Ago3 and Aub stability. Nature cell biology. 2009 May;11(5):652.

Kirino Y, Mourelatos Z. Mouse Piwi-interacting RNAs are 2 [variant prime]-O-methylated at their 3 [variant prime] termini. Nature structural & molecular biology. 2007 Apr 1;14(4):347.

Kirino Y, Vourekas A, Kim N, de Lima Alves F, Rappsilber J, Klein PS, Jongens TA, Mourelatos Z. Arginine methylation of vasa protein is conserved across phyla. Journal of Biological Chemistry. 2010 Mar 12;285(11):8148-54.

Klenov MS, Sokolova OA, Yakushev EY, Stolyarenko AD, Mikhaleva EA, Lavrov SA, Gvozdev VA. Separation of stem cell maintenance and transposon silencing functions of Piwi protein. Proceedings of the National Academy of Sciences. 2011 Nov 15;108(46):18760-5.

Klose RJ, Zhang Y. Regulation of histone methylation by demethylimination and demethylation. Nat Rev Mol Cell Biol. 2007 Apr;8(4):307-18.

Ko YG, Nishino K, Hattori N, Arai Y, Tanaka S, Shiota K. Stage-by-stage change in DNA methylation status of Dnmt1 locus during mouse early development. J Biol Chem. 2005 Mar 11;280(10):9627-34.

Kobayashi H, Sakurai T, Imai M, *et al*. Contribution of Intragenic DNA Methylation in Mouse Gametic DNA Methylomes to Establish Oocyte-Specific Heritable Marks. Reik W, ed. PLoS Genetics. 2012;8(1):e1002440. doi:10.1371/journal.pgen.1002440.

Koh KP, Yabuuchi A, Rao S, *et al*. Tet1 and Tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells. Cell stem cell. 2011;8(2):200-213. doi:10.1016/j.stem.2011.01.008.

Kohli RM, Zhang Y. TET enzymes, TDG and the dynamics of DNA demethylation. Nature. 2013 Oct 24;502(7472):472.

Koito A, Ikeda T. Intrinsic immunity against retrotransposons by APOBEC cytidine deaminases. Frontiers in microbiology. 2013;4.

Kojima K, Kuramochi-Miyagawa S, Chuma S, Tanaka T, Nakatsuji N, Kimura T, Nakano T. Associations between PIWI proteins and TDRD1/MTR-1 are critical for integrated subcellular localization in murine male germ cells. Genes to Cells. 2009 Oct 1;14(10):1155-65.

Kojima-Kita K, Kuramochi-Miyagawa S, Nagamori I, Ogonuki N, Ogura A, Hasuwa H, Akazawa T, Inoue N5, Nakano T. MIWI2 as an Effector of DNA Methylation and Gene Silencing in Embryonic Male Germ Cells. Cell Rep. 2016 Sep 13;16(11):2819-28. doi: 10.1016/j.celrep.2016.08.027.

Kono T, Obata Y, Wu Q, Niwa K, Ono Y, Yamamoto Y, Park ES, Seo JS, Ogawa H. Birth of parthenogenetic mice that can develop to adulthood. Nature. 2004 Apr 22;428(6985):860-4.

Kourmouli N, Jeppesen P, Mahadevhaiah S, Burgoyne P, Wu R, Gilbert DM, Bongiorni S, Prantera G, Fanti L, Pimpinelli S, Shi W, Fundele R, Singh PB. Heterochromatin and tri-methylated lysine 20 of histone H4 in animals. J Cell Sci. 2004 May 15;117(Pt 12):2491-501.

Kouskouti A, Talianidis I. Histone modifications defining active genes persist after transcriptional and mitotic inactivation. The EMBO Journal. 2005;24(2):347-357. doi:10.1038/sj.emboj.7600516.

Ku M, Koche RP, Rheinbay E, Mendenhall EM, Endoh M, Mikkelsen TS, *et al*. (2008) Genomewide Analysis of PRC1 and PRC2 Occupancy Identifies Two Classes of Bivalent Domains. PLoS Genet 4(10): e1000242. https://doi.org/10.1371/journal.pgen.1000242

Kuramochi-Miyagawa S, Kimura T, Ijiri TW, Isobe T, Asada N, Fujita Y, Ikawa M, Iwai N, Okabe M, Deng W, Lin H. Mili, a mammalian member of piwi family gene, is essential for spermatogenesis. Development. 2004 Feb 15;131(4):839-49.

Kuramochi-Miyagawa S, Kimura T, Yomogida K, Kuroiwa A, Tadokoro Y, Fujita Y, Sato M, Matsuda Y, Nakano T. Two mouse piwi-related genes: miwi and mili. Mechanisms of development. 2001 Oct 31;108(1):121-33.

Kuramochi-Miyagawa S, Watanabe T, Gotoh K, Takamatsu K, Chuma S, Kojima-Kita K, Shiromoto Y, Asada N, Toyoda A, Fujiyama A, Totoki Y. MVH in piRNA processing and gene silencing of retrotransposons. Genes & development. 2010 May 1;24(9):887-92.

Kuramochi-Miyagawa S, Watanabe T, Gotoh K, Totoki Y, Toyoda A, Ikawa M, Asada N, Kojima K, Yamaguchi Y, Ijiri TW, Hata K. DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. Genes & development. 2008 Apr 1;22(7):908-17.

Kurosawa K, Sasaki H, Sato Y, Yamanaka M, Shimizu M, Ito Y, Okuyama T, Matsuo M, Imaizumi K, Kuroki Y, Nishimura G. Paternal UPD14 is responsible for a distinctive malformation complex. American journal of medical genetics. 2002 Jul 1;110(3):268-72.

Kurz A, Lampel S, Nickolenko JE, Bradl J, Benner A, Zirbel RM, Cremer T, Lichter P. Active and inactive genes localize preferentially in the periphery of chromosome territories. The Journal of cell biology. 1996 Dec 1;135(5):1195-205.

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 & Development. 2002;16(22):2893-2905. doi:10.1101/gad.1035902.

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. 2001 Mar 1;410(6824):116-20.

Lai AY, Fatemi M, Dhasarathy A, Malone C, Sobol SE, Geigerman C, Jaye DL, Mav D, Shah R, Li L, Wade PA. DNA methylation prevents CTCF-mediated silencing of the oncogene BCL6 in B cell lymphomas. Journal of Experimental Medicine. 2010 Aug 23:jem-20100204.

Lane N, Dean W, Erhardt S, Hajkova P, Surani A, Walter J, Reik W. Resistance of IAPs to methylation reprogramming may provide a mechanism for epigenetic inheritance in the mouse. genesis. 2003 Feb 1;35(2):88-93.

Längst G, Bonte EJ, Corona DF, Becker PB. Nucleosome movement by CHRAC and ISWI without disruption or trans-displacement of the histone octamer. Cell. 1999 Jun 25;97(7):843-52.

Lau NC, Seto AG, Kim J, Kuramochi-Miyagawa S, Nakano T, Bartel DP, Kingston RE. Characterization of the piRNA complex from rat testes. Science. 2006 Jul 21;313(5785):363-7.

Law JA, Du J, Hale CJ, *et al*. Polymerase-IV occupancy at RNA-directed DNA methylation sites requires SHH1. Nature. 2013;498(7454):385-389. doi:10.1038/nature12178.

Le Thomas A, Rogers AK, Webster A, Marinov GK, Liao SE, Perkins EM, Hur JK, Aravin AA, Tóth KF. Piwi induces piRNA-guided transcriptional silencing and establishment of a repressive chromatin state. Genes & development. 2013 Feb 15;27(4):390-9.

Le Thomas A, Stuwe E, Li S, *et al*. Transgenerationally inherited piRNAs trigger piRNA biogenesis by changing the chromatin of piRNA clusters and inducing

precursor processing. Genes & Development. 2014;28(15):1667-1680. doi:10.1101/gad.245514.114.

Lee BC, Kim MK, Jang G, Oh HJ, Yuda F, Kim HJ, Shamim MH, Kim JJ, Kang SK, Schatten G, Hwang WS. Dogs cloned from adult somatic cells. Nature. 2005 Aug 4;436(7051):641-.

Lemos MC, Thakker RV. GNAS mutations in Pseudohypoparathyroidism type 1a and related disorders. Human mutation. 2015 Jan 1;36(1):11-9.

Lewis JD, Meehan RR, Henzel WJ, Maurer-Fogy I, Jeppesen P, Klein F, Bird A. Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA. Cell. 1992 Jun 12;69(6):905-14.

Li BZ, Huang Z, Cui QY, Song XH, Du L, Jeltsch A, Chen P, Li G, Li E, Xu GL. Histone tails regulate DNA methylation by allosterically activating de novo methyltransferase. Cell research. 2011 Aug;21(8):1172.

Li E, Beard C, Jaenisch R. Role for DNA methylation in genomic imprinting. Nature. 1993 Dec 25;366(6453):362-5.

Li JY, Lees-Murdock DJ, Xu GL, Walsh CP. Timing of establishment of paternal methylation imprints in the mouse. Genomics. 2004 Dec 31;84(6):952-60.

Li X, Ito M, Zhou F, *et al*. A maternal-zygotic effect gene Zfp57 maintains both maternal and paternal imprints. Developmental cell. 2008;15(4):547-557. doi:10.1016/j.devcel.2008.08.014.

Lieberman-Aiden E, Van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, Sandstrom R. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. science. 2009 Oct 9;326(5950):289-93.

Lim SL, Qu ZP, Kortschak RD, Lawrence DM, Geoghegan J, Hempfling AL, Bergmann M, Goodnow CC, Ormandy CJ, Wong L, Mann J. HENMT1 and piRNA stability are required for adult male germ cell transposon repression and to define the spermatogenic program in the mouse. PLoS genetics. 2015 Oct 23;11(10):e1005620.

Lin H, Spradling AC. A novel group of pumilio mutations affects the asymmetric division of germline stem cells in the Drosophila ovary. Development. 1997 Jun 15;124(12):2463-76.

Lindroth AM, Park YJ, McLean CM, Dokshin GA, Parker JM, Pasini D, *et al*. Antagonism between DNA and H3K27 Methylation at the Imprinted *Rasgrf1* Locus. PLoS Genetics 2008, 4(8): e1000145. https://doi.org/10.1371/journal.pgen.1000145.

Liu X, Gao Q, Li P, Zhao Q, Zhang J, Li J, *et al*. UHRF1 targets DNMT1 for DNA methylation through cooperative binding of hemi-methylated DNA and methylated H3K9. Nat Comm 2013 4, Article number: 1563 doi:10.1038/ncomms2562

Liu X, Want C, Liu W, Li J, Li C, Kou X, *et al*. Distinct features of H3K4me3 and H3K27me3 chromatin domains in pre-implantation embryos. Nature 537, 558–562 (22 September 2016) doi:10.1038/nature19362

Liu Z, Xie Z, Jones W, Pavlovicz RE, Liu S, Yu J, *et al*. Curcumin is a potent DNA hypomethylation agent. Bioorganic & Medicinal Chemistry Letters. Volume 19, Issue 3, 1 February 2009, Pages 706–709

Lobanenkov V. V. *et al*. A novel sequence-specific DNA binding protein which interacts with three regularly spaced direct repeats of the CCCTC-motif in the 5′-flanking sequence of the chicken c-mycgene. Oncogene 5, 1743–1753 (1990).

Lucifero D, Chaillet JR, Trasler JM. Potential significance of genomic imprinting defects for reproduction and assisted reproductive technology. Human reproduction update. 2004 Jan 1;10(1):3-18.

Luedi PP, Dietrich FS, Weidman JR, Bosko JM, Jirtle RL, Hartemink AJ. Computational and experimental identification of novel human imprinted genes. Genome research. 2007 Dec 1;17(12):1723-30.

Lupiáñez DG, Kraft K, Heinrich V, *et al*. Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions. Cell. 2015;161(5):1012-1025. doi:10.1016/j.cell.2015.04.004.

Lyle R, Watanabe D, te Vruchte D, Lerchner W, Smrzka OW, Wutz A, Schageman J, Hahner L, Davies C, Barlow DP. The imprinted antisense RNA at the Igf2r locus overlaps but does not imprint Mas1. Nat Genet. 2000 May;25(1):19-21.

Ma L, Buchold GM, Greenbaum MP, Roy A, Burns KH, Zhu H, Han DY, Harris RA, Coarfa C, Gunaratne PH, Yan W. GASZ is essential for male meiosis and suppression of retrotransposon expression in the male germline. PLoS genetics. 2009 Sep 4;5(9):e1000635.

Mabb AM, Judson MC, Zylka MJ, Philpot BD. Angelman syndrome: insights into genomic imprinting and neurodevelopmental phenotypes. Trends in neurosciences. 2011 Jun 30;34(6):293-303.

Macfarlan TS, Gifford WD, Driscoll S, Lettieri K, Rowe HM, Bonanomi D, Firth A, Singer O, Trono D, Pfaff SL. Embryonic stem cell potency fluctuates with endogenous retrovirus activity. Nature 2012 487: 57–63.

MacGregor GR, Zambrowicz BP, Soriano P. Tissue non-specific alkaline phosphatase is expressed in both embryonic and extraembryonic lineages during

mouse embryogenesis but is not required for migration of primordial germ cells. Development. 1995 May 1;121(5):1487-96.

Mackay DJ, Callaway JL, Marks SM, White HE, Acerini CL, Boonen SE, Dayanikli P, Firth HV, Goodship JA, Haemers AP, Hahnemann JM. Hypomethylation of multiple imprinted loci in individuals with transient neonatal diabetes is associated with mutations in ZFP57. Nature genetics. 2008 Aug 1;40(8):949-51.

Manakov SA, Pezic D, Marinov GK, Pastor WA, Sachidanandam R, Aravin AA. MIWI2 and MILI have differential effects on piRNA biogenesis and DNA methylation. Cell reports. 2015 Aug 25;12(8):1234-43.

Marchant TW, Johnson EJ, McTeir L, *et al*. Canine Brachycephaly Is Associated with a Retrotransposon-Mediated Missplicing of SMOC2. Current Biology. 2017;27(11):1573-1584.e6. doi:10.1016/j.cub.2017.04.057.

Matzke MA and Mosher RA. RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. Nature Reviews Genetics 15, 394–408 (2014) doi:10.1038/nrg3683

McClintock B (1950) The origin and behavior of mutable loci in maize. Proc Natl Acad Sci USA 36(6):344–355.

McGrath J, Solter D. Completion of mouse embryogenesis requires both the maternal and paternal genomes. Cell. 1984 May 1;37(1):179-83.

McGraw S, Oakes CC, Martel J, Cirio MC, de Zeeuw P, Mak W, *et al*. (2013) Loss of DNMT1o Disrupts Imprinted X Chromosome Inactivation and Accentuates Placental Defects in Females. PLoS Genet 9(11): e1003873. https://doi.org/10.1371/journal.pgen.1003873

McKinnell Z, Wessel G. Ligers and tigons and..... what?.... oh my!. Molecular reproduction and development. 2012 Aug 1;79(8).

Meehan RR, Lewis JD, McKay S, Kleiner EL, Bird AP. Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs. Cell. 1989 Aug 11;58(3):499-507.

Meng L, Ward AJ, Chun S, Bennett CF, Beaud*et al*, Rigo F. Towards a therapy for Angelman syndrome by reduction of a long non-coding RNA. Nature. 2015 Feb 19;518(7539):409.

Merlo A, Herman JG, Mao L, Lee DJ, Gabrielson E, Burger PC, Baylin SB, Sidransky D. 5' CpG island methylation is associated with transcriptional silencing of the tumour suppressor p16/CDKN2/MTS1 in human cancers. Nat Med. 1995 Jul;1(7):686-92.

Messerschmidt DM, de Vries W, Ito M, Solter D, Ferguson-Smith A, Knowles BB. Trim28 is required for epigenetic stability during mouse oocyte to embryo transition. Science. 2012 Mar 23;335(6075):1499-502. doi: 10.1126/science.1216154.

Meyer P, Heidmann I, Niedenhof I. Differences in DNA-methylation are associated with a paramutation phenomenon in transgenic petunia. The Plant Journal. 1993 Jul 1;4(1):89-100.

Michaelis C, Ciosk R, Nasmyth K. Cohesins: chromosomal proteins that prevent premature separation of sister chromatids. Cell. 1997 Oct 3;91(1):35-45.

Milagro FI, Gómez-Abellán P, Campión J, Martínez JA, Ordovás JM, Garaulet M. CLOCK, PER2 and BMAL1 DNA methylation: association with obesity and metabolic syndrome characteristics and monounsaturated fat intake. Chronobiol Int. 2012 Nov;29(9):1180-94. doi: 10.3109/07420528.2012.719967

Milot E, Strouboulis J, Trimborn T, Wijgerde M, de Boer E, Langeveld A, Tan-Un K, Vergeer W, Yannoutsos N, Grosveld F, Fraoser P. Heterochromatin Effects on the Frequency and Duration of LCR-Mediated Gene Transcription. Cell. 1996 Oct 4;87(1):105-14.

Mittelsten Scheid O, Afsar K, Paszkowski J. Formation of stable epialleles and their paramutation-like interaction in tetraploid Arabidopsis thaliana. Nat Genet. 2003 Aug;34(4):450-4.

Mohn F, Handler D, Brennecke J. piRNA-guided slicing specifies transcripts for Zucchini-dependent, phased piRNA biogenesis. Science. 2015 May 15;348(6236):812-7.

Moissiard G, Cokus SJ, Cary J, Feng S, Billi AC, Stroud H, Husmann D, Zhan Y, Lajoie BR, McCord RP, Hale CJ. MORC family ATPases required for heterochromatin condensation and gene silencing. Science. 2012 Jun 15;336(6087):1448-51.

Moleres A, Campión J, Milagro FI, Marcos A, Campoy C, Garagorri JM, Gómez-Martínez S, Martínez JA, Azcona-Sanjulián MC, Martí A; EVASYON Study Group. Differential DNA methylation patterns between high and low responders to a weight loss intervention in overweight or obese adolescents: the EVASYON study. FASEB J. 2013 Jun;27(6):2504-12. doi: 10.1096/fj.12-215566.

Monk D, Bentley L, Hitchins M, Myler RA, Clayton-Smith J, Ismail S, Price SM, Preece MA, Stanier P, Moore GE. Chromosome 7p disruptions in Silver Russell syndrome: delineating an imprinted candidate gene region. Human genetics. 2002 Oct 1;111(4):376-87.

Morgan HD, Sutherland HG, Martin DI, Whitelaw E. Epigenetic inheritance at the agouti locus in the mouse. Nat Genet. 1999 Nov;23(3):314-8.

Muramatsu M, Kinoshita K, Fagarasan S, Yamada S, Shinkai Y, Honjo T. Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. Cell. 2000 Sep 1;102(5):553-63.

Muramatsu M, Sankaranand VS, Anant S, Sugai M, Kinoshita K, Davidson NO, Honjo T. Specific expression of activation-induced cytidine deaminase (AID), a novel member of the RNA-editing deaminase family in germinal center B cells. J Biol Chem. 1999 Jun 25;274(26):18470-6.

Nabel CS, Jia H, Ye Y, *et al*. AID/APOBEC deaminases disfavor modified cytosines implicated in DNA demethylation. Nature chemical biology. 2012;8(9):751-758. doi:10.1038/nchembio.1042.

Nakamura T, Arai Y, Umehara H, Masuhara M, Kimura T, Taniguchi H, Sekimoto T, Ikawa M, Yoneda Y, Okabe M, Tanaka S, Shiota K, Nakano T. PGC7/Stella protects against DNA demethylation in early embryogenesis. Nat Cell Biol. 2007 Jan;9(1):64-71.

Nakamura T, Liu YJ, Nakashima H, Umehara H, Inoue K, Matoba S, Tachibana M, Ogura A, Shinkai Y, Nakano T. PGC7 binds histone H3K9me2 to protect against conversion of 5mC to 5hmC in early embryos. Nature. 2012 Jun 3;486(7403):415-9. doi: 10.1038/nature11093.

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

Nguyen AT, Zhang Y. The diverse functions of Dot1 and H3K79 methylation. Genes & Development. 2011;25(13):1345-1358. doi:10.1101/gad.2057811.

Nishimasu H, Ishizu H, Saito K, Fukuhara S, Kamatani MK, Bonnefond L, Matsumoto N, Nishizawa T, Nakanaga K, Aoki J, Ishitani R. Structure and function of Zucchini endoribonuclease in piRNA biogenesis. Nature. 2012 Nov 8;491(7423):284.

Noordermeer D, Leleu M, Schorderet P, Joye E, Chabaud F, Duboule D. Temporal dynamics and developmental memory of 3D chromatin architecture at Hox gene loci. Elife. 2014 Apr 29;3:e02557.

Nora EP, Goloborodko A, Valton AL, Gibcus JH, Uebersohn A, Abdennur N, Dekker J, Mirny LA, Bruneau BG. Targeted degradation of CTCF decouples local insulation of chromosome domains from genomic compartmentalization. Cell. 2017 May 18;169(5):930-44.

Nora EP, Lajoie BR, Schulz EG, *et al*. Spatial partitioning of the regulatory landscape of the X-inactivation center. Nature. 2012;485(7398):381-385. doi:10.1038/nature11049.

Ohno R, Nakayama M, Naruse C, Okashita N, Takano O, Tachibana M, Asano M, Saitou M, Seki Y. A replication-dependent passive mechanism modulates DNA demethylation in mouse primordial germ cells. Development. 2013 Jul;140(14):2892-903. doi: 10.1242/dev.093229.

Okano M, Bell DW, Haber DA, and Li E (1999). DNA Methyltransferases Dnmt3a and Dnmt3b Are Essential for De Novo Methylation and Mammalian Development. Cell 99: 247-257.

Okashita N, Kumaki Y, Ebi K, Nishi M, Okamoto Y, Nakayama M, Hashimoto S, Nakamura T, Sugasawa K, Kojima N, Takada T, Okano M, Seki Y. PRDM14 promotes active DNA demethylation through the ten-eleven translocation (TET)-mediated base excision repair pathway in embryonic stem cells. Development. 2014 Jan;141(2):269-80. doi: 10.1242/dev.099622.

Okashita N, Kumaki Y, Ebi K, Nishi M, Okamoto Y, Nakayama M, Hashimoto S, Nakamura T, Sugasawa K, Kojima N, Takada T, Okano M, Seki Y. PRDM14 promotes active DNA demethylation through the ten-eleven translocation (TET)-mediated base excision repair pathway in embryonic stem cells. Development. 2014 Jan;141(2):269-80. doi: 10.1242/dev.099622.

Olivieri D, Sykora MM, Sachidanandam R, Mechtler K, Brennecke J. An in vivo RNAi assay identifies major genetic and cellular requirements for primary piRNA biogenesis in Drosophila. The EMBO journal. 2010 Oct 6;29(19):3301-17.

Ooi SKT, Qiu C, Bernstein E, *et al*. DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. Nature. 2007;448(7154):714-717. doi:10.1038/nature05987.

Otani J, Nankumo T, Arita K, Inamoto S, Ariyoshi M, Shirakawa M. Structural basis for recognition of H3K4 methylation status by the DNA methyltransferase 3A ATRX–DNMT3–DNMT3L domain. EMBO reports. 2009 Nov 1;10(11):1235-41.

Pandey RR, Tokuzawa Y, Yang Z, *et al*. Tudor domain containing 12 (TDRD12) is essential for secondary PIWI interacting RNA biogenesis in mice. Proceedings of the National Academy of Sciences of the United States of America. 2013;110(41):16492-16497. doi:10.1073/pnas.1316316110.

Pane A, Wehr K, Schüpbach T. zucchini and squash encode two putative nucleases required for rasiRNA production in the Drosophila germline. Developmental cell. 2007 Jun 5;12(6):851-62.

Pant V, Kurkuti S, Pugacheva E, Shamsuddin S, Mariano P, Renkawitz R, *et al*. Mutation of a single CTCF target site within the H19 imprinting control region leads to loss of Igf2 imprinting and complex patterns of de novo methylation upon maternal inheritance. Mol Cell Biol 2004, 24(8): 3497-504

Pant V, Mariano P, Kanduri C, Mattsson A, Lobanenkov V, Heuchel R, and Ohlsson R. The nucleotides responsible for the direct physical contact between the chromatin insulator protein CTCF and the H19 imprinting control region manifest parent of origin-specific long-distance insulation and methylation-free domains. Genes and Development 2003, 17: 586-90.

Papamichos-Chronakis M, Watanabe S, Rando OJ, Peterson CL. Global regulation of H2A. Z localization by the INO80 chromatin-remodeling enzyme is essential for genome integrity. Cell. 2011 Jan 21;144(2):200-13.

Park YJ, Herman H, Gao Y, Lindroth AM, Hu BY, Murphy PJ, Putnam JR, and Soloway PD. Sequences Sufficient for Programming Imprinted Germline DNA Methylation Defined. PLoS ONE 2012, 7(3): e33024.

Pastor WA, Aravind L, Rao A. TETonic shift: biological roles of TET proteins in DNA demethylation and transcription. Nature reviews Molecular cell biology. 2013;14(6):341-356. doi:10.1038/nrm3589.

Pastor WA, Stroud H, Nee K, *et al*. MORC1 represses transposable elements in the mouse male germline. Nature Communications. 2014;5:5795. doi:10.1038/ncomms6795.

Payer B, Saitou M, Barton SC, Thresher R, Dixon JP, Zahn D, Colledge WH, Carlton MB, Nakano T, Surani MA. Stella is a maternal effect gene required for normal early development in mice. Curr Biol. 2003 Dec 2;13(23):2110-7.

Pearsall RS, Plass C, Romano MA, Garrick MD, Shibata H, Hayashizaki Y, Held WA. A Direct Repeat Sequence at the*Rasgrf1*Locus and Imprinted Expression. Genomics. 1999 Jan 15;55(2):194-201.

Peaston AE, Evsikov AV, Graber JH, de Vries WN, Holbrook AE, Solter D, Knowles BB. Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos. Dev Cell 2004 7: 597–606.

Peat JR, Dean W, Clark SJ, *et al*. Genome-wide Bisulfite Sequencing in Zygotes Identifies Demethylation Targets and Maps the Contribution of TET3 Oxidation. Cell Reports. 2014;9(6):1990-2000. doi:10.1016/j.celrep.2014.11.034.

Pezic D, Manakov SA, Sachidanandam R, Aravin AA. piRNA pathway targets active LINE1 elements to establish the repressive H3K9me3 mark in germ cells. Genes & development. 2014 Jul 1;28(13):1410-28.

Plass C, Shibata H, Kalcheva I, Mullins L, Kotelevtseva N, Mullins J, Kato R, Sasaki H, Hirotsune S, Okazaki Y, Held WA. Identification of Grf1 on mouse chromosome 9 as an imprinted gene by RLGS–M. Nature genetics. 1996 Sep 1;14(1):106-9.

Pope BD, Ryba T, Dileep V, Yue F, Wu W, Denas O, Vera DL, Wang Y, Hansen RS, Canfield TK, Thurman RE. Topologically-associating domains are stable units of replication-timing regulation. Nature. 2014 Nov 20;515(7527):402.

Popp C, Dean W, Feng S, Cokus SJ, Andrews S, Pellegrini M, *et al*. Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. Nature 2010, 63: 1101-1105.

Prickett AR, Oakey RJ. A survey of tissue-specific genomic imprinting in mammals. Molecular genetics and genomics. 2012 Aug 1;287(8):621-30.

Qiao D, Zeeman AM, Deng W, Looijenga LH, Lin H. Molecular characterization of hiwi, a human member of the piwi gene family whose overexpression is correlated to seminomas. Oncogene. 2002 Jun 6;21(25):3988.

Qin H, Von Arnim AG. Epigenetic history of an Arabidopsis trans-silencer locus and a test for relay of trans-silencing activity. BMC plant biology. 2002 Dec 11;2(1):11.

Qiu W, Baccarelli A, Carey VJ, *et al*. Variable DNA Methylation Is Associated with Chronic Obstructive Pulmonary Disease and Lung Function. American Journal of Respiratory and Critical Care Medicine. 2012;185(4):373-381. doi:10.1164/rccm.201108-1382OC.

Quenneville S, Verde G, Corsinotti A, Kapopoulou A, Jakobsson J, Offner 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. 2011 Nov 4;44(3):361-72. doi: 10.1016/j.molcel.2011.08.032

Rakyan VK, Down TA, Thorne NP, *et al*. An integrated resource for genome-wide identification and analysis of human tissue-specific differentially methylated regions (tDMRs). Genome Research. 2008;18(9):1518-1529. doi:10.1101/gr.077479.108.

Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, Aiden EL. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell. 2014 Dec 18;159(7):1665-80.

Rassoulzadegan M, Grandjean V, Gounon P, Vincent S, Gillot I, and Cuzin F. RNA-mediated non-mendelian inheritance of an epigenetic change in the mouse. Nature 2006, 441: 469-74.

Rassoulzadegan M, Magliano M, Cuzin F. Transvection effects involving DNA methylation during meiosis in the mouse. The EMBO journal. 2002 Feb 1;21(3):440-50.

Ratnam S, Mertineit C, Ding F, Howell CY, Clarke HJ, Bestor TH, Chaillet JR, Trasler JM. Dynamics of Dnmt1 methyltransferase expression and intracellular localization

during oogenesis and preimplantation development. Dev Biol. 2002 May 15;245(2):304-14.

Rea S, Eisenhaber F, O'Carroll D, Strahl BD, Sun ZW, Schmid M, Opravil S, Mechtler K, Ponting CP, Allis CD, Jenuwein T. Regulation of chromatin structure by site-specific histone H3 methyltransferases. Nature. 2000 Aug 10;406(6796):593-9.

Reddington JP, Perricone SM, Nestor CE, *et al*. Redistribution of H3K27me3 upon DNA hypomethylation results in de-repression of Polycomb target genes. Genome Biology. 2013;14(3):R25. doi:10.1186/gb-2013-14-3-r25.

Reuter M, Berninger P, Chuma S, Shah H, Hosokawa M, Funaya C, Antony C, Sachidanandam R, Pillai RS. Miwi catalysis is required for piRNA amplification-independent LINE1 transposon silencing. Nature. 2011 Dec 8;480(7376):264.

Reuter M, Chuma S, Tanaka T, Franz T, Stark A, and Pillai R. Loss of the Mili-interacting Tudor domain–containing protein-1 activates transposons and alters the Mili-associated small RNA profile. Nature Structural & Molecular Biology 16, 639 - 646 (2009) doi:10.1038/nsmb.1615

Reuter M, Chuma S, Tanaka T, Franz T, Stark A, Pillai RS. Loss of the Mili-interacting Tudor domain–containing protein-1 activates transposons and alters the Mili-associated small RNA profile. Nature structural & molecular biology. 2009 Jun 1;16(6):639-46.

Reynolds N, Salmon-Divon M, Dvinge H, *et al*. NuRD-mediated deacetylation of H3K27 facilitates recruitment of Polycomb Repressive Complex 2 to direct gene repression. The EMBO Journal. 2012;31(3):593-605. doi:10.1038/emboj.2011.431.

Rodriguez C, Borgel J, Court F, Cathala G, Forné T, Piette J. CTCF is a DNA methylation-sensitive positive regulator of the INK/ARF locus. Biochemical and biophysical research communications. 2010 Feb 5;392(2):129-34.

Rothbart SB, Krajewski K, Nady N, *et al*. Association of UHRF1 with H3K9 methylation directs the maintenance of DNA methylation. Nature structural & molecular biology. 2012;19(11):1155-1160. doi:10.1038/nsmb.2391.

Rowe HM, Kapopoulou A, Corsinotti A, *et al*. TRIM28 repression of retrotransposon-based enhancers is necessary to preserve transcriptional dynamics in embryonic stem cells. Genome Research. 2013;23(3):452-461. doi:10.1101/gr.147678.112.

Rozhkov NV, Hammell M, Hannon GJ. Multiple roles for Piwi in silencing Drosophila transposons. Genes & development. 2013 Feb 15;27(4):400-12.

Rubio ED, Reiss DJ, Welcsh PL, *et al*. CTCF physically links cohesin to chromatin. Proceedings of the National Academy of Sciences of the United States of America. 2008;105(24):8309-8314. doi:10.1073/pnas.0801273105.

Rugg-Gunn PJ, Cox BJ, Ralston A, Rossant J. Distinct histone modifications in stem cell lines and tissue lineages from the early mouse embryo. Proceedings of the National Academy of Sciences of the United States of America. 2010;107(24):10783-10790. doi:10.1073/pnas.0914507107.

Ryba T, Hiratani I, Lu J, Itoh M, Kulik M, Zhang J, Schulz TC, Robins AJ, Dalton S, Gilbert DM. Evolutionarily conserved replication timing profiles predict long-range chromatin interactions and distinguish closely related cell types. Genome research. 2010 Jun 1;20(6):761-70.

Santi DV, Garrett CE, Barr PJ. On the mechanism of inhibition of DNA-cytosine methyltransferases by cytosine analogs. Cell Volume 33, Issue 1, p9–10, May 1983. Doi: 10.1016/0092-8674(83)90327-6

Sapetschnig A, Sarkies P, Lehrbach NJ, Miska EA (2015) Tertiary siRNAs Mediate Paramutation in C. elegans. PLoS Genet 11(3): e1005078. https://doi.org/10.1371/journal.pgen.1005078

Sato M1, Kimura T, Kurokawa K, Fujita Y, Abe K, Masuhara M, Yasunaga T, Ryo A, Yamamoto M, Nakano T. Identification of PGC7, a New Gene Expressed Specifically in Preimplantation Embryos and Germ Cells. Mech Dev. 2002 Apr;113(1):91-4.

Saxe JP, Chen M, Zhao H, Lin H. Tdrkh is essential for spermatogenesis and participates in primary piRNA biogenesis in the germline. The EMBO Journal. 2013;32(13):1869-1885. doi:10.1038/emboj.2013.121.

Scheid OM, Afsar K, Paszkowski J. Formation of stable epialleles and their paramutation-like interaction in tetraploid Arabidopsis thaliana. Nature genetics. 2003 Aug 1;34(4):450.

Scheid OM, Afsar K, Paszkowski J. Formation of stable epialleles and their paramutation-like interaction in tetraploid Arabidopsis thaliana. Nature genetics. 2003 Aug 1;34(4):450.

Schoenherr CJ, Levorse JM, Tilghman SM. CTCF maintains differential methylation at the Igf2/H19 locus. Nat Genet. 2003 Jan;33(1):66-9.

Schotta G, Lachner M, Sarma K, Ebert A, Sengupta R, Reuter G, Reinberg D, Jenuwein T. A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. Genes Dev 2004 18: 1251–1262.

Schotta G, Sengupta R, Kubicek S, Malin S, Kauer M, Callen E, Celeste A, Pagani M, Opravil S, De La Rosa-Velazquez IA, *et al*. A chromatin-wide transition to H4K20 monomethylation impairs genome integrity and programmed DNA rearrangements in the mouse. Genes Dev 2008 22: 2048–2061.

Schultz DC, Ayyanathan K, Negorev D, Maul GG, Rauscher FJ. SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins. Genes & Development. 2002;16(8):919-932. doi:10.1101/gad.973302.

Schwarzer W, Abdennur N, Goloborodko A, Pekowska A, Fudenberg G, Loe-Mie Y, Fonseca NA, Huber W, Haering C, Mirny L, Spitz F. Two independent modes of chromosome organization are revealed by cohesin removal. bioRxiv. 2016 Jan 1:094185.

Seisenberger S, Andrews S, Krueger F, *et al*. The Dynamics of Genome-wide DNA Methylation Reprogramming in Mouse Primordial Germ Cells. Molecular Cell. 2012;48(6):849-862. doi:10.1016/j.molcel.2012.11.001.

Sexton T, Yaffe E, Kenigsberg E, Bantignies F, Leblanc B, Hoichman M, Parrinello H, Tanay A, Cavalli G. Three-dimensional folding and functional organization principles of the Drosophila genome. Cell. 2012 Feb 3;148(3):458-72.

Sharif J, Muto M, Takebayashi S, Suetake I, Iwamatsu A, Endo TA, Shinga J, Mizutani-Koseki Y, Toyoda T, Okamura K, Tajima S, Mitsuya K, Okano M, Koseki H. The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. Nature. 2007 Dec 6;450(7171):908-12.

Shin T, Kraemer D, Pryor J, Liu L, Rugila J, Howe L, Buck S, Murphy K, Lyons L, Westhusin M. Cell biology: a cat cloned by nuclear transplantation. Nature. 2002 Feb 21;415(6874):859.

Shibata M, Blauvelt KE, Liem KF, García-García MJ. TRIM28 is required by the mouse KRAB domain protein ZFP568 to control convergent extension and morphogenesis of extra-embryonic tissues. Development 2011, 138(24):5333-5343. doi:10.1242/dev.072546.

Shirayama M, Seth M, Lee H-C, *et al*. piRNAs initiate an epigenetic memory of non-self RNA in the C. elegans germline. Cell. 2012;150(1):65-77. doi:10.1016/j.cell.2012.06.015.

Shoji M, Tanaka T, Hosokawa M, Reuter M, Stark A, Kato Y, Kondoh G, Okawa K, Chujo T, Suzuki T, Hata K, Martin SL, Noce T, Kuramochi-Miyagawa S, Nakano T, Sasaki H, Pillai RS, Nakatsuji N, Chuma S. The TDRD9-MIWI2 complex is essential for piRNA-mediated retrotransposon silencing in the mouse male germline. Dev Cell. 2009 Dec;17(6):775-87. doi: 10.1016/j.devcel.2009.10.012.

Sidorenko LV, Peterson T. Transgene-induced silencing identifies sequences involved in the establishment of paramutation of the maize p1 gene. The Plant Cell. 2001 Feb 1;13(2):319-35.

Sidorenko LV, Peterson T. Transgene-induced silencing identifies sequences involved in the establishment of paramutation of the maize p1 gene. The Plant Cell. 2001 Feb 1;13(2):319-35.

Sienski G, Dönertas D, Brennecke J. Transcriptional silencing of transposons by Piwi and maelstrom and its impact on chromatin state and gene expression. Cell. 2012 Nov 21;151(5):964-80.

Simon B, Kirkpatrick JP, Eckhardt S, Reuter M, Rocha EA, Andrade-Navarro MA, Sehr P, Pillai RS, Carlomagno T. Recognition of 2′-O-methylated 3′-end of piRNA by the PAZ domain of a Piwi protein. Structure. 2011 Feb 9;19(2):172-80.

Slotkin RK, Martienssen R. Transposable elements and the epigenetic regulation of the genome. Nature reviews. Genetics. 2007 Apr 1;8(4):272.

Smallwood SA, Tomizawa S, Krueger F, *et al*. Dynamic CpG island methylation landscape in oocytes and preimplantation embryos. Nature genetics. 2011;43(8):811-814. doi:10.1038/ng.864.

Smith CL, Peterson CL. ATP-dependent chromatin remodeling. Curr Top Dev Biol. 2005;65:115-48.

Smith FM, Garfield AS, Ward A. Regulation of growth and metabolism by imprinted genes. Cytogenetic and genome research. 2006;113(1-4):279-91.

Smith RJ, Dean W, Konfortova G, Kelsey G. Identification of novel imprinted genes in a genome-wide screen for maternal methylation. Genome research. 2003 Apr 1;13(4):558-69.

Song, F., Smith, J.F., Kimura, M.T., Morrow, A.D., Matsuyama, T., Nagase, H. and Held, W.A., 2005. Association of tissue-specific differentially methylated regions (TDMs) with differential gene expression. Proceedings of the National Academy of Sciences of the United States of America, 102(9), pp.3336-3341.

Spencer HG, Clark AG. Non-conflict theories for the evolution of genomic imprinting. Heredity. 2014 Aug;113(2):112.

Spilianakis CG, Lalioti MD, Town T, Lee GR, Flavell RA. Interchromosomal associations between alternatively expressed loci. Nature. 2005 Jun 2;435(7042):637.

Stack SM, Brown DB, Dewey WC. Visualization of interphase chromosomes. Journal of cell science. 1977 Aug 1;26(1):281-99.

Stam M, Belele C, Dorweiler JE, Chandler VL. Differential chromatin structure within a tandem array 100 kb upstream of the maize b1 locus is associated with paramutation. Genes & development. 2002 Aug 1;16(15):1906-18.

Stam M, Belele C, Dorweiler JE, Chandler VL. Differential chromatin structure within a tandem array 100 kb upstream of the maize b1 locus is associated with paramutation. Genes & Development. 2002;16(15):1906-1918. doi:10.1101/gad.1006702.

Steger DJ, Lefterova MI, Ying L, *et al*. DOT1L/KMT4 Recruitment and H3K79 Methylation Are Ubiquitously Coupled with Gene Transcription in Mammalian Cells . Molecular and Cellular Biology. 2008;28(8):2825-2839. doi:10.1128/MCB.02076-07.

Stelzer Y, Ronen D, Bock C, Boyle P, Meissner A, Benvenisty N. Identification of novel imprinted differentially methylated regions by global analysis of human-parthenogenetic-induced pluripotent stem cells. Stem Cell Reports. 2013 Jun 4;1(1):79-89.

Stöger R, Kubička P, Liu CG, Kafri T, Razin A, Cedar H, Barlow DP. Maternal-specific methylation of the imprinted mouse Igf2r locus identifies the expressed locus as carrying the imprinting signal. Cell. 1993 Apr 9;73(1):61-71.

Strahl BD and Allis CD. The language of covalent histone modifications. Nature. 2000 Jan 6;403(6765):41-5.

Sundaram V, Cheng Y, Ma Z, Li D, Xing X, Edge P, Snyder MP, Wang T. Widespread contribution of transposable elements to the innovation of gene regulatory networks. Genome research. 2014 Dec 1;24(12):1963-76.

Surani MA, Barton SC, Norris ML. Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis. Nature. 1984 Apr 5;308(5959):548-50.

Surani MA. A conserved imprinting control region at the HYMAI/ZAC domain is implicated in transient neonatal diabetes mellitus. Human molecular genetics. 2001 Jul 1;10(14):1475-83.

Tachibana M, Sugimoto K, Nozaki M, *et al*. G9a histone methyltransferase plays a dominant role in euchromatic histone H3 lysine 9 methylation and is essential for early embryogenesis. Genes & Development. 2002;16(14):1779-1791. doi:10.1101/gad.989402.

Tahiliani M, Koh KP, Shen Y, *et al*. Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1. Science (New York, NY). 2009;324(5929):930-935. doi:10.1126/science.1170116.

Tanaka SS, Toyooka Y, Akasu R, Katoh-Fukui Y, Nakahara Y, Suzuki R, Yokoyama M, Noce T. The mouse homolog of Drosophila Vasa is required for the development of male germ cells. Genes & development. 2000 Apr 1;14(7):841-53.

Tang W, Tu S, Lee H-C, Weng Z, Mello CC. The ribonuclease PARN-1 trims piRNA 3′ ends to promote transcriptome surveillance in C. elegans. Cell. 2016;164(5):974-984. doi:10.1016/j.cell.2016.02.008.

Tang W, Tu S, Lee HC, Weng Z, Mello CC. The RNase PARN-1 trims piRNA 3′ ends to promote transcriptome surveillance in C. elegans. Cell. 2016 Feb 25;164(5):974-84.

Taylor DH, McLean CM, Wu WL, Wang AB, and Soloway PD. Imprinted DNA methylation reconstituted at a non-imprinted locus. Epigenetics & Chromatin 2016, 9:41.

Tedeschi A, Wutz G, Huet S, Jaritz M, Wuensche A, Schirghuber E, Davidson IF, Tang W, Cisneros DA, Bhaskara V, Nishiyama T. Wapl is an essential regulator of chromatin structure and chromosome segregation. Nature. 2013 Sep 26;501(7468):564.

Thompson TC, Southgate J, Kitchener G, Land H. Multistage carcinogenesis induced by ras and myc oncogenes in a reconstituted organ. Cell. 1989 Mar 24;56(6):917-30.

Thornburg BG, Gotea V, Makałowski W. Transposable elements as a significant source of transcription regulating signals. Gene. 2006 Jan 3;365:104-10.

Thorvaldsen JL, Duran KL, Bartolomei MS. Deletion of the H19 differentially methylated domain results in loss of imprinted expression of H19 and Igf2. Genes & Development. 1998;12(23):3693-3702.

Tomizawa SI, Sasaki H. Genomic imprinting and its relevance to congenital disease, infertility, molar pregnancy and induced pluripotent stem cell. Journal of human genetics. 2012 Feb 1;57(2).

Toperoff G, Aran D, Kark JD, *et al*. Genome-wide survey reveals predisposing diabetes type 2-related DNA methylation variations in human peripheral blood. Human Molecular Genetics. 2012;21(2):371-383. doi:10.1093/hmg/ddr472.

Torres-Padilla M, Parfitt D, Kouzarides T, and Zernicka-Goetz M. Histone arginine methylation regulates pluripotency in the early mouse embryo. Nature 11 January 2007 445, 214-218 doi:10.1038/nature05458

Troyer AF. Adaptedness and heterosis in corn and mule hybrids. Crop science. 2006 Mar 1;46(2):528-43.

Tunster SJ, Van de Pette M, John RM. Fetal overgrowth in the Cdkn1c mouse model of Beckwith-Wiedemann syndrome. Disease models & mechanisms. 2011 Nov 1;4(6):814-21.

Turelli P, Castro-Diaz N, Marzetta F, *et al*. Interplay of TRIM28 and DNA methylation in controlling human endogenous retroelements. Genome Research. 2014;24(8):1260-1270. doi:10.1101/gr.172833.114.

Ueda T, Abe K, Miura A, Yuzuriha M, Zubair M, Noguchi M, Niwa K, Kawase Y, Kono T, Matsuda Y, Fujimoto H. The paternal methylation imprint of the mouse H19 locus is acquired in the gonocyte stage during foetal testis development. Genes to Cells. 2000 Aug 1;5(8):649-59.

Upadhyay AK and Cheng X. Dynamics of Histone Lysine Methylation: Structures of Methyl Writers and Erasers. Epigenetics and Disease pp 107-124

Vagin VV, Wohlschlegel J, Qu J, Jonsson Z, Huang X, Chuma S, Girard A, Sachidanandam R, Hannon GJ, Aravin AA. Proteomic analysis of murine Piwi proteins reveals a role for arginine methylation in specifying interaction with Tudor family members. Genes & development. 2009 Aug 1;23(15):1749-62.

Varma G, Rawat P, Jalan M, Vinayak M, Srivastava M. Influence of a CTCF-dependent insulator on multiple aspects of enhancer-mediated chromatin organization. Molecular and cellular biology. 2015 Oct 15;35(20):3504-16.

Vasileva A, Tiedau D, Firooznia A, Müller-Reichert T, Jessberger R. Tudor domain protein Tdrd6 is required for spermiogenesis, chromatoid body architecture and regulation of miRNA expression. Current biology : CB. 2009;19(8):630-639. doi:10.1016/j.cub.2009.02.047.

Vastenhouw NL, Zhang Y, Woods IG, Imam F, Regev A, Liu XS, Rinn J, Schier AF. Chromatin signature of embryonic pluripotency is established during genome activation. Nature. 2010;464:922–926.

Villar-Garea A, Fraga MF, Espada J, Esteller M. Procaine is a DNA-demethylating agent with growth-inhibitory effects in human cancer cells. Cancer Res. 2003 Aug 15;63(16):4984-9.

Visel A, Blow MJ, Li Z, *et al*. ChIP-seq accurately predicts tissue-specific activity of enhancers. Nature. 2009;457(7231):854-858. doi:10.1038/nature07730.

Volkmar M, Dedeurwaerder S, Cunha DA, *et al*. DNA methylation profiling identifies epigenetic dysregulation in pancreatic islets from type 2 diabetic patients. The EMBO Journal. 2012;31(6):1405-1426. doi:10.1038/emboj.2011.503.

Vourekas A, Zheng K, Fu Q, Maragkakis M, Alexiou P, Ma J, Pillai RS, Mourelatos Z, Wang PJ. The RNA helicase MOV10L1 binds piRNA precursors to initiate piRNA processing. Genes & development. 2015 Mar 15;29(6):617-29.

Vucic EA, Chari R, Thu KL, *et al*. DNA Methylation Is Globally Disrupted and Associated with Expression Changes in Chronic Obstructive Pulmonary Disease

Small Airways. American Journal of Respiratory Cell and Molecular Biology. 2014;50(5):912-922. doi:10.1165/rcmb.2013-0304OC.

Wade PA, Gegonne A, Jones PL, Ballestar E, Aubry F, Wolffe AP. Mi-2 complex couples DNA methylation to chromatin remodelling and histone deacetylation. Nature genetics. 1999 Sep 1;23(1).

Wagner KD, Wagner N, Ghanbarian H, Grandjean V, Gounon P, Cuzin F, *et al*. RNA induction and inheritance of epigenetic cardiac hypertrophy in the mouse. Dev Cell. 2008;14: 962–969. pmid:18539123

Walker EL. Paramutation of the r1 locus of maize is associated with increased cytosine methylation. Genetics. 1998 Apr 1;148(4):1973-81.

Walter J, Hutter B, Khare T, Paulsen M. Repetitive elements in imprinted genes. Cytogenetic and genome research. 2006;113(1-4):109-15.

Walter M, Teissandier A, Pérez-Palacios R, Bourc'his D. An epigenetic switch ensures transposon repression upon dynamic loss of DNA methylation in embryonic stem cells. Ferguson-Smith AC, ed. eLife. 2016;5:e11418. doi:10.7554/eLife.11418.

Walter M, Teissandier A, Pérez-Palacios R, Bourc'his D. An epigenetic switch ensures transposon repression upon dynamic loss of DNA methylation in embryonic stem cells. Ferguson-Smith AC, ed. eLife. 2016;5:e11418. doi:10.7554/eLife.11418.

Wang X, Sun Q, McGrath SD, Mardis ER, Soloway PD, Clark AG. Transcriptome-wide identification of novel imprinted genes in neonatal mouse brain. PloS one. 2008 Dec 4;3(12):e3839.

Wang X, Soloway PD, Clark AG. A survey for novel imprinted genes in the mouse placenta by mRNA-seq. Genetics. 2011 Sep 1;189(1):109-22.

Wang H, Maurano MT, Qu H, Varley KE, Gertz J, Pauli F, Lee K, Canfield T, Weaver M, Sandstrom R, Thurman RE. Widespread plasticity in CTCF occupancy linked to DNA methylation. Genome research. 2012 Sep 1;22(9):1680-8.

Wang J, Hevi S, Kurash JK, Lei H, Gay F, Bajko J, Su H, Sun W, Chang H, Xu G, Gaudet F, Li E, Chen T. The lysine demethylase LSD1 (KDM1) is required for maintenance of global DNA methylation. Nat Genet. 2009 Jan;41(1):125-9. doi: 10.1038/ng.268.

Wang J, Saxe JP, Tanaka T, Chuma S, Lin H. Mili interacts with tudor domain-containing protein 1 in regulating spermatogenesis. Current Biology. 2009 Apr 28;19(8):640-4.

Wang L, Zhang J, Duan J, Gao X, Zhu W, Lu X, Yang L, Zhang J, Li G, Ci W, Li W. Programming and inheritance of parental DNA methylomes in mammals. Cell. 2014 May 8;157(4):979-91.

Wang Z, Zang C, Rosenfeld JA, *et al*. Combinatorial patterns of histone acetylations and methylations in the human genome. Nature genetics. 2008;40(7):897-903. doi:10.1038/ng.154.

Watanabe T, Tomizawa S, Mitsuya K, Totoki Y, Yamamoto Y, Kuramochi-Miyagawa S, *et al*.Role for piRNAs and noncoding RNA in de novo DNA methylation of the imprinted mouse *Rasgrf1* locus. Science 2011, 332(6031):848-52.

Weiser NE, Yang DX, Feng S, Kalinava N, Brown KC, Khanikar J, Freeberg MA, Snyder MJ, Csankovszki G, Chan RC, Gu SG. MORC-1 Integrates Nuclear RNAi and Transgenerational Chromatin Architecture to Promote Germline Immortality. Developmental Cell. 2017 May 22;41(4):408-23.

Wenda JM, Homolka D, Yang Z, *et al*. Distinct Roles of RNA Helicases MVH and TDRD9 in PIWI Slicing-Triggered Mammalian piRNA Biogenesis and Function. Developmental Cell. 2017;41(6):623-637.e9. doi:10.1016/j.devcel.2017.05.021.

Whitelaw NC, Chong S, Morgan DK, *et al*. Reduced levels of two modifiers of epigenetic gene silencing, Dnmt3a and Trim28, cause increased phenotypic noise. Genome Biology. 2010;11(11):R111. doi:10.1186/gb-2010-11-11-r111.

Wilkins JF, Haig D. What good is genomic imprinting: the function of parent-specific gene expression. Nature reviews. Genetics. 2003 May 1;4(5):359.

Wolf D, Goff SP. Embryonic stem cells use ZFP809 to silence retroviral DNAs. Nature. 2009;458(7242):1201-1204. doi:10.1038/nature07844.

Woods GL, White KL, Vanderwall DK, Li GP, Aston KI, Bunch TD, Meerdo LN, Pate BJ. A mule cloned from fetal cells by nuclear transfer. Science. 2003 Aug 22;301(5636):1063-.

Worch S, Hansmann I, Schlote D. Paramutation-like effects at the mouse scapinin (Phactr3) locus. J Mol Biol. 2008 Mar 28;377(3):605-8. doi: 10.1016/j.jmb.2008.01.062.

Wossidlo M, Nakamura T, Lepikhov K, Marques CJ, Zakhartchenko V, Boiani M, Arand J, Nakano T, Reik W, Walter J. 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. Nat Commun. 2011 2:241. doi: 10.1038/ncomms1240.

Wu H, Coskun V, Tao J, *et al*. Dnmt3a-Dependent Nonpromoter DNA Methylation Facilitates Transcription of Neurogenic Genes. Science (New York, NY). 2010;329(5990):444-448. doi:10.1126/science.1190485.

Wu J, Huang B, Chen H, Yin Q, Liu Y, Xiang Y, *et al*. The landscape of accessible chromatin in mammalian preimplantation embryos. Nature. 2016 Jun 30;534(7609):652-7

Wutz A, Smrzka OW, Schweifer N, Schellander K, Wagner EF, Barlow DP. Imprinted expression of the Igf2r gene depends on an intronic CpG island. Nature. 1997 Oct 16;389(6652):745-9.

Xu GL, Bestor TH, Bourc'his D, Hsieh CL, Tommerup N, Bugge M, Hulten M, Qu X, Russo JJ, Viegas-Péquignot E. Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. Nature. 1999 Nov 11;402(6758):187-91.

Xue Y, Wong J, Moreno GT, Young MK, Côté J, Wang W. NURD, a novel complex with both ATP-dependent chromatin-remodeling and histone deacetylase activities. Molecular cell. 1998 Dec 31;2(6):851-61.

Yamaguchi S, Hong K, Liu R, *et al*. Tet1 controls meiosis by regulating meiotic gene expression. Nature. 2012;492(7429):443-447. doi:10.1038/nature11709.

Yamaguchi S, Shen L, Liu Y, Sendler D, Zhang Y. Role of Tet1 in genomic imprinting erasure. Nature. 2013;504(7480):460-464. doi:10.1038/nature12805.

Yang F, Deng X, Ma W, Berletch JB, Rabaia N, Wei G, Moore JM, Filippova GN, Xu J, Liu Y, Noble WS. The lncRNA Firre anchors the inactive X chromosome to the nucleolus by binding CTCF and maintains H3K27me3 methylation. Genome biology. 2015 Mar 12;16(1):52.

Yang T, Adamson TE, Resnick JL, Leff S, Wevrick R, Francke U, Jenkins NA, Copeland NG, Brannan CI. A mouse model for Prader-Willi syndrome imprinting-centre mutations. Nature genetics. 1998 May 1;19(1):25-31.

Yang Z, Chen K-M, Pandey RR, *et al*. PIWI Slicing and EXD1 Drive Biogenesis of Nuclear piRNAs from Cytosolic Targets of the Mouse piRNA Pathway. Molecular Cell. 2016;61(1):138-152. doi:10.1016/j.molcel.2015.11.009.

Yoon BJ, Herman H, Sikora A, Smith LT, Plass C, and Soloway PD. Regulation of DNA methylation at *Rasgrf1*. Nature Genetics 2002, 30(1): 92-6.

Yoon B, Herman H, Hu B, Park YJ, Lindroth A, Bell A, West AG, Chang Y, Stablewski A, Piel JC, Loukinov DI. *Rasgrf1* imprinting is regulated by a CTCF-dependent methylation-sensitive enhancer blocker. Molecular and cellular biology. 2005 Dec 15;25(24):11184-90.

Zaina S, Heyn H, Carmona FJ, *et al*. DNA methylation map of human atherosclerosis. Circ Cardiovasc Genet. 2014;7:692–700. doi: 10.1161/CIRCGENETICS.113.000441

Zampieri M, Guastafierro T, Calabrese R, Ciccarone F, Bacalini MG, Reale A, Perilli M, Passananti C, Caiafa P. ADP-ribose polymers localized on Ctcf–Parp1–Dnmt1 complex prevent methylation of Ctcf target sites. Biochemical Journal. 2012 Jan 15;441(2):645-52.

Zhang B, Zheng H, Huang B, Li W, Xiang Y, Peng X, *et al*. Allelic reprogramming of the histone modification H3K4me3 in early mammalian development. Nature. 2016 Sep 22;537(7621):553-557. doi: 10.1038/nature19361

Zhang J, Wang Q, Wang M, Jiang M, Wang Y, Sun Y, Wang J, Xie T, Tang C, Tang N, Song H. GASZ and mitofusin-mediated mitochondrial functions are crucial for spermatogenesis. EMBO reports. 2016 Feb 1;17(2):220-34.

Zhang S, Kubota C, Yang L, Zhang Y, Page R, O'Neill M, Yang X, Tian XC. Genomic imprinting of H19 in naturally reproduced and cloned cattle. Biology of reproduction. 2004 Nov 1;71(5):1540-4.

Zhang W, Xia W, Wang Q, Towers AJ, Chen J, Gao R, Zhang Y, Yen CA, Lee AY, Li Y, Zhou C, Liu K, Zhang J, Gu TP, Chen X, Chang Z, Leung D, Gao S, Jiang YH, Xie W. Isoform Switch of TET1 Regulates DNA Demethylation and Mouse Development. Mol Cell. 2016 Dec 15;64(6):1062-1073. doi: 10.1016/j.molcel.2016.10.030.

Zhang Y, Ng H-H, Erdjument-Bromage H, Tempst P, Bird A, Reinberg D. Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. Genes & Development. 1999;13(15):1924-1935.

Zhang Y, Ng H-H, Erdjument-Bromage H, Tempst P, Bird A, Reinberg D. Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. Genes & Development. 1999;13(15):1924-1935.

Zheng H, Huang B, Zhang B, Xiang Y, Du Z, Li Y, *et al*. Resetting Epigenetic Memory by Reprogramming of Histone Modifications in Mammals. Mol Cell. 2016 Sep 15;63(6):1066-79. doi: 10.1016/j.molcel.2016.08.032.

Zheng K, Wang PJ (2012) Blockade of Pachytene piRNA Biogenesis Reveals a Novel Requirement for Maintaining Post-Meiotic Germline Genome Integrity. PLoS Genet 8(11): e1003038. https://doi.org/10.1371/journal.pgen.1003038

Zheng Z, Yu H, Miki D, Jin D, Zhang Q, Ren Z, *et al*. Involvement of Multiple Gene-Silencing Pathways in a Paramutation-like Phenomenon in Arabidopsis. Cell Reports Volume 11, Issue 8, 26 May 2015, 1160-1167 doi: 10.1016/j.celrep.2015.04.034

Zhu S, Goldschmidt-Clermont PJ, Dong C. Inactivation of monocarboxylate transporter MCT3 by DNA methylation in atherosclerosis. Circulation. 2005 Aug 30;112(9):1353-61.

Ziller MJ, Gu H, Müller F, *et al*. Charting a dynamic DNA methylation landscape of the human genome. Nature. 2013;500(7463):477-481. doi:10.1038/nature12433.

Zuin J, Dixon JR, van der Reijden MI, Ye Z, Kolovos P, Brouwer RW, van de Corput MP, van de Werken HJ, Knoch TA, van IJcken WF, Grosveld FG. Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. Proceedings of the National Academy of Sciences. 2014 Jan 21;111(3):996-1001.

## CHAPTER II

Alexander KA, Wang X, Shibata M, Clark AG, García-García MJ. TRIM28 controls genomic imprinting through distinct mechanisms during and after early genome-wide reprogramming. Cell Reports 2015;13(6):1194-1205. doi:10.1016/j.celrep.2015.09.078.

Alleman M, Sidorenko L, McGinnis K, Seshadri V, Dorweiler JE, White J, et al. An RNA-dependent RNA polymerase is required for paramutation in maize. Nature 2006, 442: 295-8.

Amândio AR, Necsulea A, Joye E, Mascrez B, Duboule D (2016). Hotair Is Dispensible for Mouse Development. PLoS Genet 12(12): e1006232.

Barau J, Teissandier A, Zamudio N, Roy S, Nalesso V, Herault Y, et al. The DNA methyltransferase DNMT3C protects male germ cells from transposon activity. Science 2016, 354(6314): 909-912.

Baron U, Gossen M, and Bujard H. Co-regulation of two gene activities by tetracycline via a bidirectional promoter. Nucleic Acids Research 1997, 25: 2723-2729.

Bengani H, Ganapathi M, Singh GP and Brahmachari V. Mining of Putative cis-acting Elements for Chromatin Mediated Regulation of Hox Genesin Mammals by in-silico Analysis. Journal of Experimental Zoology (Mol Dev Evol) 2007, 308B:384-395.

Bhutani N, Brady JJ, Damian M, Sacco A, Corbel SY, and Blau HM. Reprogramming towards pluripotency requires AID-dependent DNA demethylation. Nature 2010,463(7284): 1042-1047.

Bourc'his D, Xu GL, Lin CS, Bollman B, and Bestor TH. Dnmt3L and the establishment of maternal genomic imprints. Science 2001, 294(5551): 2536-9.

Brideau C, Kauppinen KP, Holmes R, and Soloway PD. A Non-Coding RNA Within the Rasgrf1 Locus in Mouse Is Imprinted and Regulated by Its Homologous Chromosome in Trans. PLoS ONE 2010: e13784. https://doi.org/10.1371/journal.pone.0013784

Brown JL, Grau DJ, DeVido SK, and Kassis JA. Sp1/KLF binding site is important for the activity of a Polycomb group response element from the Drosophila engrailed gene. Nucleic Acids Research 2005 33(16): 5181-5189. doi:10.1093/nar/gki827

Chi W, Wu E, and Morgan BA. Dermal papilla cell number specifies hair size, shape and cycling and its reduction causes follicular decline. Development 2013, 140(8): 1676-1683. doi: 10.1242/dev.090662

Chotalia M, Smallwod SA, Ruf N, Dawson C, Lucifero D, Frontera M, et al. Transcription is required for establishment of germline methylation marks at imprinted genes. Genes and Development 2008, 23: 105-117.

Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, Zhang F. Multiplex Genome Engineering Using CRISPR/Cas Systems. Science 2013, 339(6121):819-23.

Cortellino S, Xu J, Sannai M, Moore R, Carette E, Cigliano A, et al. Thymine DNA Glycosylase Is Essential for Active DNA Demethylation by Linked Deamination-Base Excision Repair. Cell 2011, 146(1): 67-79.

de Vanssay A, Boug AL, Boivin A, Hermant C, Teysset L, Delmarre V, et al. Paramutation in Drosophila linked to emergence of a piRNA-producing locus. Nature 2012, 490: 112-5.

Deshane J, Kim J, Bolisetty S, Hock TD, Hill-Kapturczak N, and Agarwal A. Sp1 regulates chromatin looping between an intronic enhancer and distal promoter of the human heme oxygenase-1 gene in renal cells. Journal of Biological Chemistry 2010, 285(22): 16476-86.

Dowen JM, Fan ZP, Hnisz D, Ren G, Abraham BJ, Zhang LN, et al. Control of Cell Identity Genes Occurs in Insulated Neighborhoods in Mammalian Chromosomes. Cell 2014, 159(2): 374–387.

Drake NM, Park YJ, Shirali AS, Cleland TA, and Soloway PD. Imprint switch mutations at *Rasgrf1* support conflict hypothesis of imprinting and define a growth control mechanism upstream of IGF1. Mamm Genome 2009, 20(9-10):654-63. doi: 10.1007/s00335-009-9192-7

Duselis, A. R., Vrana, P. B. Retrieval of Mouse Oocytes. J. Vis. Exp. 2007, (3), e185, doi:10.3791/185.

Eggan K, Akutsu H, Loring J, Jackson-Grusby L, Klemm M, Rideout WM, et al. Hybrid vigor, fetal overgrowth, and viability of mice derived by nuclear cloning and tetraploid embryo complementation. PNAS 2001 May 22, 98(11): 6209-6214.

Engreitz JM, Haines JE, Perez EM, Munson G, Chen J, Kane M, McDonel PE, Guttman M, and Lander E (2016). Local regulation of gene expression by lncRNA promoters, transcription and splicing. Nature 539, 452–455.

Frohlich LF, Mrakovcic M, Steinborn R, Chung U, Bastepe M, and Juppner H. Targeted deletion of the Nesp55 DMR defines another Gnas imprinting control region and provides a mouse model of autosomal dominant PHP-Ib. PNAS 2010, 107(20): 9275-9280.

Fukuhara M, Ma Y, Nagasawa K, and Toyosihma F. A G-quadruplex structure at the 5′ end of the H19 coding region regulates H19 transcription. Scientific Reports 2017, 7, Article number: 45815. doi:10.1038/srep45815

Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, et al. Bioconductor: open software development for computational biology and bioinformatics. Genome Biology 2004, 5:R80. doi: https://doi.org/10.1186/gb-2004-5-10-r80

Goff LA and Rinn JL. Linking RNA biology to lncRNAs. Genome Research 2015, 25: 1456-1465.

Goodrich R, Johnson G, and Krawetz SA. The preparation of human spermatozoal RNA for clinical analysis. Archives of Andrology 2007, 53(3): 161-7.

Gossen M and Bujard H. Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. Proceedings of the National Academy of Science 1992, 89: 5547-5551.

Hayashi S, Lewis P, Pevny L, and McMahan AP. Efficient gene modulation in mouse epiblast using a Sox2Cre transgenic mouse strain. Mechanisms of Development 2002, 119 Suppl 1:S97-S101.

Herman H, Lu M, Anggraini M, Sikora A, Chang Y, Yoon BJ, and Soloway PD. Trans allele methylation and paramutation-like effects in mice. Nature Genetics 2003, 34(2): 199-292.

Holmes R, Chang Y, and Soloway PD. Timing and sequence requirements defined for embryonic maintenance of imprinted DNA methylation at *Rasgrf1*. Molecular and Cell Biology 2006, 26(24):9564-70.

Howell CY, Bestor TH, Ding F, Latham KE, Mertineit C, Trasler JM, and Chaillet JR. Genomic imprinting disrupted by a maternal effect mutation in the Dnmt1 gene. Cell 2001, 104(6):829-38.

Ito S, Shen L, Dai Q, Wu SC, Collins LB, Swenberg J, He C, et al. Tet Proteins Can

Convert 5-Methylcytosine to 5-Formylcytosine and 5-Carboxylcytosine. Science 2011, 333(6047): 1300-1303.

Jain D, Meydan C, Lange J, Bouuaert CC, Mason CE, Anderson K, et al. rahu is a mutant allele of Dnmt3c, encoding a DNA methyltransferase homolog required for meiosis and transposon repression in the mouse male germline. bioRxiv 2017, https://doi.org/10.1101/121822.

Kantor B, Kaufman Y, Makedonski K, Razin A, and Shemer R. Establishing the epigenetic status of the Prader–Willi/Angelman imprinting center in the gametes and embryo. Human Molecular Genetics 2004, 12(22): 2767-2779.

Kato Y, Kaneda M, Hata K, Kumaki K, Hisano M, Kohara 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. Human Molecular Genetics 2007, 16(19): 2272-80.

Kumaki Y, Oda M, and Okano M. QUMA: quantification tool for methylation analysis. Nucleic Acids Research 2008, 36(Web Server issue), W170–W175.

Kuramochi-Miyagawa S, Watanabe T, Gotoh K, Totoki Y, Toyoda A, Ikawa M, et al. DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. Genes and Development 2008: 22: 908-917. doi: 10.1101/gad.1640708

Laird PW, Zijderveld A, Linders K, Rudnicki MA, Jaenisch R, and Berns A. Simplified mammalian DNA isolation procedure. Nucleic Acids Res 1991, 19(15): 4293.

Latos P, Pauler FM, Koerner MV, Senergin HB, Hudson QJ, Stocsits RR, et al. Airn Transcriptional Overlap, But Not Its lncRNA Products, Induces Imprinted Igf2r Silencing. Science 2012, 1469-1472.

Lee J. Epigenetic Regulation by Long Noncoding RNAs. Science 2012, 338(6113): 1435-1439.

Li E, Bestor TH, and Jaenisch R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell 1992, 69: 915-926

Li L and Chang HY. Physiological roles of long noncoding RNAs: Insights from knockout mice. Trends in Cell Biology 2014, 24(10): 594–602.

Lindroth AM, Park YJ, McLean CM, Dokshin GA, Parker JM, Pasini D, et al. Antagonism between DNA and H3K27 Methylation at the Imprinted *Rasgrf1* Locus. PLoS Genetics 2008, 4(8): e1000145. https://doi.org/10.1371/journal.pgen.1000145.

Liu Z-W, Shao C-R, Zhang C-J, Zhou J-X, Zhang S-W, Li L, et al. The SET Domain Proteins SUVH2 and SUVH9 Are Required for Pol V Occupancy at RNA-Directed DNA Methylation Loci. PLoS Genet 2014, 10(1): e1003948.

Nagano T, Mitchell JA, Sanz LA, Pauler FM, Ferguson-Smith AC, Feil F, et al. The Air Noncoding RNA Epigenetically Silences Transcription by Targeting G9a to Chromatin. Science 2008, 322(5908): 1717-1720. DOI: 10.1126/science.1163802

Nakamura T, Arai Y, Umehara H, Masuhara M, Kimura T, Taniguchi H, et al. PGC7/Stella protects against DNA demethylation in early embryogenesis. Nature Cell Biology 2006, 9, 64 – 71.

Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, et al. Spatial partitioning of the regulatory landscape of the X inactivation center. Nature 2012; 485(7398): 381–385. doi:10.1038/nature11049.

O'Connor L, Gilmour J, and Bonifer C. The Role of the Ubiquitously Expressed Transcription Factor Sp1 in Tissue-specific Transcriptional Regulation and in Disease. Yale Journal of Biology and Medicine 2016, 89: 413-525.

Ogawa Y and Lee JT (2003). Xite, X-inactivation intergenic transcription elements that regulate the probability of choice. Molecular Cell 11(3):731-43.

Okano M, Bell DW, Haber DA, and Li E. DNA Methyltransferases Dnmt3a and Dnmt3b Are Essential for De Novo Methylation and Mammalian Development. Cell 1999, 99: 247-257.

Orom UA, Derrien T, Beringer M, Gumireddy K, Gardini A, Bussotti G, et al. Long Noncoding RNAs with Enhancer-like Function in Human Cells. Cell 2010, 143(1): 46-58.

Pant V, Kurkuti S, Pugacheva E, Shamsuddin S, Mariano P, Renkawitz R, et al. Mutation of a single CTCF target site within the H19 imprinting control region leads to loss of Igf2 imprinting and complex patterns of de novo methylation upon maternal inheritance. Mol Cell Biol 2004, 24(8): 3497-504.

Pant V, Mariano P, Kanduri C, Mattsson A, Lobanenkov V, Heuchel R, and Ohlsson R. The nucleotides responsible for the direct physical contact between the chromatin insulator protein CTCF and the H19 imprinting control region manifest parent of origin-specific long-distance insulation and methylation-free domains. Genes and Development 2003, 17: 586-90.

Paralkar VR, Taborda CC, Huang P, Yao Y, Kossenkov AV, Prasad R, Luan J, Davies JOJ, Hughes JR, Hardison RC, Blobel GA, and Weiss MJ. Unlinking a lncRNA from its associated cis element. Molecular Cell 2016, 62(1), 104–110.

Park YJ, Herman H, Gao Y, Lindroth AM, Hu BY, Murphy PJ, Putnam JR, and Soloway PD. Sequences Sufficient for Programming Imprinted Germline DNA Methylation Defined. PLoS ONE 2012, 7(3): e33024.

Pastor WA, Stroud H, Nee K, Liu W, Pezic D, Lee SA, et al. MORC1 represses transposable elements in the mouse male germline. Nat Commun. 2014 Dec 12; 5: 5795. doi: 10.1038/ncomms6795

Payer B, Saitou M, Baron SC, Thresher R, Dixon JP, Zahn D, Colledge WH, et al. Curr Biol 2003, 13(23):2110-7.

Pearsall RS, Plass C, Romano MA, Garrick MD, Shibata H, Hayashizaki Y, Held WA. A Direct Repeat Sequence at theRasgrf1Locus and Imprinted Expression. Genomics 1999 Jan 15;55(2):194-201.

Plath K, Fang J, Mlynarczyk-Evans SK, Cao R, Worringer KA, Wang H, et al. Role of histone H3 lysine 27 methylation in X inactivation. Science 2003, 300: 131-135.

Popp C, Dean W, Feng S, Cokus SJ, Andrews S, Pellegrini M, et al. Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. Nature 2010, 63: 1101-1105.

Quenneville S, Verde G, Corsinotti A, Kapopoulou A, Jakobsson J, Offner 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 2011 Nov 4;44(3):361-72. doi: 10.1016/j.molcel.2011.08.032.

Quilez J, Guilmatre A, Garg P, Highnam G, Gymrek M, Erlich Y, et al. Polymorphic tandem repeats within gene promoters act as modifiers of gene expression and DNA methylation in humans. Nucleic Acids Research 2016 May 5; 44(8): 3750-3762. doi: 10.1093/nar/gkw219

Raiber E, Kranaster R, Lam E, Nikan M, and Balasubramanian S. A non-canonical DNA structure is a binding motif for the transcription factor SP1 in vitro. Nucleic Acids Research 2012, 40(4): 1499-1508.

Rassoulzadegan M, Grandjean V, Gounon P, Vincent S, Gillot I, and Cuzin F. RNA-mediated non-mendelian inheritance of an epigenetic change in the mouse. Nature 2006, 441: 469-74.

Rose SD and MacDonald RJ. Integration of Tetracycline Regulation into a Cell-specific Transcriptional Enhancer. Journal of Biological Chemistry 1997, 272(8): 4375-4379.

Shibata M, Blauvelt KE, Liem KF, García-García MJ. TRIM28 is required by the

mouse KRAB domain protein ZFP568 to control convergent extension and morphogenesis of extra-embryonic tissues. Development 2011, 138(24):5333-5343. doi:10.1242/dev.072546.

Sleutels F, Zwart R, Barlow DP. The non-coding Air RNA is required for silencing autosomal imprinted genes. Nature 2002, Feb 7;415(6873):810-3.

Stavropoulos N, Rowntree RK, and Lee JT (2005). Identification of developmentally specific enhancers for Tsix in the regulation of X chromosome inactivation. Molecular Cell Biology, 25(7):2757-69.

Sumimoto Y, Higashitsuji H, Higashitsuji H, Liu Y, Fujita T, Kajurai T, Candeias MM, Itoh K, Chiba T, and Fujita J. Identification of a novel enhancer that binds Sp1 and contributes to induction of cold-inducible RNA-binding protein (cirp) expression in mammalian cells. BMC Biotechnology 2012, 12:72.

Suzuki T, Kimura A, Nagai R, Horikoshi M. Regulation of interaction of the acetyltransferase region of p300 and the DNA-binding domain of Sp1 on and through DNA binding. Genes Cells 2000, 5(1): 29-41.

Taylor DH, Chu ET, Spektor R, and Soloway PD. Long non-coding RNA regulation of reproduction and development. Molecular Reproduction and Development 2015, 82(12):932-56.

Taylor DH, McLean CM, Wu WL, Wang AB, and Soloway PD. Imprinted DNA methylation reconstituted at a non-imprinted locus. Epigenetics & Chromatin 2016, 9:41.

Todd AK and Neidle S. The relationship of potential G-quadruplex sequences in cis-upstream regions of the human genome to SP1-binding elements. Nucleic Acids Research 2008, 36(8): 2700-4.

Tsai C, Rowntree RK, Cohen DE, and Lee JT. Higher order chromatin structure at the X-inactivation center via looping DNA. Dev Biol. 2008 Jul 15; 319(2): 416–425. doi: 10.1016/j.ydbio.2008.04.010

Turelli P, Castro-Diaz N, Marzetta F, Kapopoulou A, Raclot C, Duc J, et al. Interplay of TRIM28 and DNA methylation in controlling human endogenous retroelements. Genome Res. 2014 Aug; 24(8): 1260–1270.

Urlinger S, Baron U, Thellmann M, Hasan MT, Bujard H, and Hillen W. Exploring the sequence space for tetracycline dependent transcriptional activators: Novel mutations yield expanded range and sensitivity. Proceedings of the National Academy of Science 2001, 97(14): 7963-7968.

Watanabe T, Tomizawa S, Mitsuya K, Totoki Y, Yamamoto Y, Kuramochi-Miyagawa S, et al. Role for piRNAs and noncoding RNA in de novo DNA methylation of the imprinted mouse Rasgrf1 locus. Science 2011, 332(6031):848-52.

Wiznerowicz M, Jakobsson J, Szulc J, Liao S, Quazzola A, Beermann F, et al. The Kru¨ ppel-associated Box Repressor Domain Can Trigger de Novo Promoter Methylation during Mouse Early Embryogenesis. J Biol Chem 2007, 282(47): 34535-34541.

Wu H, de Gannes MK, Luchetti G, and Pilsner JR. Rapid method for the isolation of mammalian sperm DNA. Biotechniques 2014, 59(6): 293-300.

Wutz A, Smrzka OW, Schweifer N, Schellander K, Wagner EF, and Barlow DP. Imprinted expression of the Igf2r gene depends on an intronic CpG island. Nature 1997, 389(6652): 745-9.

Yin Y, Yan P, Lu J, Song G, Zhu Y, Li Z2, Zhao Y, Shen B, Huang X, Zhu H, Orkin SH, Shen X. Opposing Roles for the lncRNA Haunt and Its Genomic Locus in Regulating HOXA Gene Activation during Embryonic Stem Cell Differentiation. Cell Stem Cell 2015, 16(5):504-16.

Yoon B, Herman H, Hu B, Park YJ, Lindroth A, Bell A, West AG, Chang Y, Stablewski A, Piel JC, Loukinov DI. Rasgrf1 imprinting is regulated by a CTCF-dependent methylation-sensitive enhancer blocker. Molecular and cellular biology. 2005 Dec 15;25(24):11184-90.

Yoon BJ, Herman H, Sikora A, Smith LT, Plass C, and Soloway PD. Regulation of DNA methylation at Rasgrf1. Nature Genetics 2002, 30(1): 92-6.

Zhang Y, Shu J, Si J, Shen L, Estecia MRH, and Issa JJ. Repetitive elements and enforced transcriptional repression co-operate to enhance DNA methylation spreading into a promoter CpG-island. Nucleic Acids Research 2012 Aug, 40(15): 7257-7268.

Zhao S, Venkatasubbarao K, Li S, and Freeman JW. Requirement of a Specific Sp1 Site for Histone Deacetylase-mediated Repression of Transforming Growth Factor β Type II Receptor Expression in Human Pancreatic Cancer Cells. Cancer Research 2003, 63, 2624-2630.

## CHAPTER III

Alam T, Medvedeva YA, Jia H, Brown JB, Lipovich L, Bajic VB (2014) Promoter Analysis of Human Long Non-Coding RNA and Protein-Coding Genes. PLoS ONE 9(10): e10944 https://doi.org/10.1371/journal.pone.0109443

Antequera, F. & Bird, A. Number of CpG islands and genes in human and

mouse. *Proc. Natl Acad. Sci. USA* **90**, 11995–11999 (1993)

Billon N, Carlisi D, Datto MB, van Grunsven LA, Watt A, Wang XF, Rudkin BB. Cooperation of Sp1 and p300 in the induction of the CDK inhibitor p21WAF1/CIP1 during NGF-mediated neuronal differentiation. Oncogene. 1999 May 6;18(18):2872-82.

Bochman ML, Paeschke K, Zakian VA. DNA secondary structures: stability and function of G-quadruplex structures. Nat Rev Genet. 2012 Nov;13(11):770-80. doi: 10.1038/nrg3296.

Borsani G, Tonlorenzi R, Simmler MC, Dandolo L, Arnaud D, Capra V, Grompe M, Pizzuti A, Muzny D, Lawrence C, Willard HF, Avner P, Ballabio A. Characterization of a murine gene expressed from the inactive X chromosome. Nature. 1991 May 23;351(6324):325-9.

Bouman P and Philepsen S. Regulation of the activity of Sp1-related transcription factors. Mol Cell Endocrinol. 2002 Sep 30;195(1-2):27-38.

Brown CJ, Ballabio A, Rupert JL, Lafreniere RG, Grompe M, Tonlorenzi R, Willard HF. A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. Nature. 1991 Jan 3;349(6304):38-44.

Caiafa P and Zampieri M. DNA methylation and chromatin structure: the puzzling CpG islands. J Cell Biochem. 2005 Feb 1;94(2):257-65.

Calo E, Wysocka J. Modification of enhancer chromatin: what, how and why? *Molecular cell*. 2013;49(5):10.1016/j.molcel.2013.01.038. doi:10.1016/j.molcel.2013.01.038.

Chu S and Ferro TJ. Sp1: regulation of gene expression by phosphorylation. Gene. 2005 Mar 28;348:1-11.

Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, Zhang F. Multiplex Genome Engineering Using CRISPR/Cas Systems. Science. 2013 Jan 3. 10.1126/science.1231143 PubMed 23287718

Cooper, S. J., Trinklein, N. D., Anton, E. D., Nguyen, L. & Myers, R. M. Comprehensive analysis of transcriptional promoter structure and function in 1% of the human genome. *Genome Res.* **16**, 1–10 (2006).

David AP, Margarit E, Domizi P, Banchio C, Armas P, Calcaterra NB. G-quadruplexes as novel cis-elements controlling transcription during embryonic development. Nucleic Acids Research. 2016;44(9):4163-4173. doi:10.1093/nar/gkw011.

Eissenberg JC, Shilatifard A. Histone H3 Lysine 4 (H3K4) Methylation in Development

and Differentiation. Developmental biology. 2010;339(2):240-249. doi:10.1016/j.ydbio.2009.08.017.

Fischer KD, Haese A, and Nowock J. Cooperation of GATA-1 and Sp1 can result in synergistic transcriptional activation or interference. *J Biol Chem* 1993, 268, 23915-23923.

Gershenzon, N. I. & Ioshikhes, I. P. Synergy of human Pol II core promoter elements revealed by statistical sequence analysis. *Bioinformatics* **21**, 1295–1300 (2005).

Guttman M, Amit I, Garber M, *et al*. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature*. 2009;458(7235):223-227. doi:10.1038/nature07672.

He S, Sun J-M, Li L, Davie JR. Differential Intranuclear Organization of Transcription Factors Sp1 and Sp3. Tansey W, ed. Molecular Biology of the Cell. 2005;16(9):4073-4083. doi:10.1091/mbc.E05-05-0388.

Isomura H, Stinski MF, Kudoh A, Daikoku T, Shirata N, and Tsurumi T. Two Sp1/Sp3 Binding Sites in the Major Immediate-Early Proximal Enhancer of Human Cytomegalovirus Have a Significant Role in Viral Replication. J. Virol. August 2005 79(15)9597-9607.

Janson L and Pettersson U. Cooperative interactions between transcription factors Sp1 and OTF-1. Proc Natl Acad Sci U S A. 1990 Jun;87(12):4732-6.

Javahery R, Khachi A, Lo K, Zenzie-Gergor B, and Smale ST. DNA sequence requirements for transcriptional initiator activity in mammalian cells. Mol Cell Biol. 1994 Jan;14(1):116-27.

Laird P *et al*. “Simplified Mammalian DNA Isolation Procedure”, Nucleic Acids Research 1991. 19: 4293.

Lanz RB, McKenna NJ, Onate SA, Albrecht U, Wong J, Tsai SY, Tsai MJ, O'Malley BW. A steroid receptor coactivator, SRA, functions as an RNA and is present in an SRC-1 complex. Cell. 1999 Apr 2;97(1):17-27.

Lewis, B. A. & Reinberg, D. The mediator coactivator complex: functional and physical roles in transcriptional regulation. *J. Cell Sci.* **116**, 3667–3675 (2003).

Lindroth AM, Park YJ, McLean CM, Dokshin GA, Persson JM, Herman H, *et al*. (2008) Antagonism between DNA and H3K27 Methylation at the Imprinted *Rasgrf1* Locus. PLoS Genet 4(8): e1000145. https://doi.org/10.1371/journal.pgen.1000145
Macleod D, Charlton J, Mullins J, and Bird AP. Sp1 sites in the mouse aprt gene promoter are required to prevent methylation of the CpG island. Genes Dev. 1994 Oct

1;8(19):2282-92.

Maizels N, Gray LT. The G4 Genome. PLoS Genet 2013, 9(4): e1003468. https://doi.org/10.1371/journal.pgen.1003468

Marin M, Karis A, Visser P, Grosveld F, and Philipsen S. Transcription Factor Sp1 Is Essential for Early Embryonic Development but Dispensable for Cell Growth and Differentiation, Cell, Volume 89, Issue 4, 1997, Pages 619-628, ISSN 0092-8674, http://dx.doi.org/10.1016/S0092-8674(00)80243-3.

Merchant JL, Du M, and Todisco A. Sp1 phosphorylation by Erk 2 stimulates DNA binding. Biochem Biophys Res Commun. 1999 Jan 19;254(2):454-61.

Mummaneni P, Yates P, Simpson J, Rose J, and Turker MS. The primary function of a redundant Sp1 binding site in the mouse aprt gene promoter is to block epigenetic gene inactivation. Nucleic Acids Research 1998 26(22) 5163–5169, https://doi.org/10.1093/nar/26.22.5163

Nam E, Lee Y, Park K, Lee JW, and Kim S. ZEB2 upregulates integrin α5 expression through cooperation with Sp1 to induce invasion during epithelial–mesenchymal transition of human cancer cells, Carcinogenesis, Volume 33, Issue 3, 1 March 2012, Pages 563–571, https://doi.org/10.1093/carcin/bgs005

Necsulea A, Soumillon M, Warnefors M, Liechti A, Diash T, Zeller U, Backer JC, Grutzner F, and Kaessman H. The evolution of lncRNA repertoires and expression patterns in tetrapods. Nature 505, 635–640 (30 January 2014) doi:10.1038/nature12943

Nolis IK, McKay DJ, Mantouvalou E, Lomvardas S, Merika M, and Thanos D. Transcription factors mediate long-range enhancer–promoter interactions. PNAS December 1, 2009 106(48) 20222–20227, doi: 10.1073/pnas.0902454106

Pasini, D., A. P. Bracken, M. R. Jensen, E. L. Denchi, and K. Helin. 2004. Suz12 is essential for mouse development and for EZH2 histone methyltransferase activity. EMBO J. 23:4061-4071.

Philipsen S and Suske G. A tale of three fingers: the family of mammalian Sp/XKLF transcription factors, Nucleic Acids Research, Volume 27, Issue 15, 1 August 1999, Pages 2991–3000, https://doi.org/10.1093/nar/27.15.2991

Raiber E, Kranaster R, Lam E, Nikan M, and Balasubramanian S. A non-canonical DNA structure is a binding motif for the transcription factor SP1 in vitro. Nucleic Acids Research 2012, 40(4): 1499-1508.

Reynard LN, Bui C, Syddall CM, and Loughlin J. CpG methylation regulates allelic

expression of GDF5 by modulating binding of SP1 and SP3 repressor proteins to the osteoarthritis susceptibility SNP rs143383. Hum Genet. 2014 Aug;133(8):1059-73. doi: 10.1007/s00439-014-1447-z.

Saxonov, S., Berg, P. & Brutlag, D. L. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proc. Natl Acad. Sci. USA* **103**, 1412–1417 (2006).

Solomon SS, Majumdar G, Martinez-Hernandez A, Raghow R. A critical role of Sp1 transcription factor in regulating gene expression in response to insulin and other hormones. Life Sci. 2008 Aug 29;83(9-10):305-12. doi: 10.1016/j.lfs.2008.06.024. Epub 2008 Jul 9.

Sumimoto Y, Higashitsuji H, Higashitsuji H, Liu Y, Fujita T, Sakurai T, Candeias MM, Itoh K, Chiba T, and Fujita J. Identification of a novel enhancer that binds Sp1 and contributes to induction of cold-inducible RNA-binding protein (cirp) expression in mammalian cells. BMC Biotechnology 2012, 12:72-81.

Wang HB, Liu GH, Zhang H, Xing S, Hu LJ, Zhao WF, Xie B, Li MZ, Zeng BH, Li Y, Zeng MS. Sp1 and c-Myc regulate transcription of BMI1 in nasopharyngeal carcinoma. FEBS J. 2013 Jun;280(12):2929-44. doi: 10.1111/febs.12299.

Watanabe T, Tomizawa S, Mitsuya K, *et al*. Role for piRNAs and Noncoding RNA in de Novo DNA Methylation of the Imprinted Mouse *Rasgrf1* Locus. *Science (New York, N.y)*. 2011;332(6031):848-852. doi:10.1126/science.1203919.

Zelko IN, Mueller MR, and Folz RJ. CpG methylation attenuates Sp1 and Sp3 binding to the human extracellular superoxide dismutase promoter and regulates its cell-specific expression. Free Radic Biol Med. 2010 Apr 1;48(7):895-904. doi: 10.1016/j.freeradbiomed.2010.01.007.

Zu X, Yu L, Sun Q, *et al*. SP1 enhances Zbtb7A gene expression via direct binding to GC box in HePG2 cells. *BMC Research Notes*. 2009;2:175. doi:10.1186/1756-0500-2-175.

## CHAPTER IV

Agarwala SD, Blitzblau HG, Hochwagen A, Fink GR. 2012. RNA methylation by the MIS complex regulates a cell fate decision in yeast. PLoS Genet 8(6):e1002732.

Alexandrova O, Schade M, Bottger A, David CN. 2005. Oogenesis in Hydra: nurse cells transfer cytoplasm directly to the growing oocyte. Dev Biol 281(1):91-101.

Alleman M, Sidorenko L, McGinnis K, Seshadri V, Dorweiler JE, White J, Sikkink K, Chandler VL. 2006. An RNA-dependent RNA polymerase is required for paramutation in maize. Nature 442(7100):295-298.

Allen L, Kloc M, Etkin LD. 2003. Identification and characterization of the Xlsirt cis-acting RNA localization element. Differentiation; research in biological diversity 71(6):311-321.

Amin V, Harris RA, Onuchic V, Jackson AR, Charnecki T, Paithankar S, Lakshmi Subramanian S, Riehle K, Coarfa C, Milosavljevic A. 2015. Epigenomic footprints across 111 reference epigenomes reveal tissue-specific epigenetic regulation of lincRNAs. Nature communications 6:6370.

Anguera MC, Ma W, Clift D, Namekawa S, Kelleher RJ, 3rd, Lee JT. 2011. Tsx produces a long noncoding RNA and has general functions in the germline, stem cells, and brain. PLoS Genet 7(9):e1002248.

Arteaga-Vazquez M, Sidorenko L, Rabanal FA, Shrivistava R, Nobuta K, Green PJ, Meyers BC, Chandler VL. 2010. RNA-mediated trans-communication can establish paramutation at the b1 locus in maize. Proc Natl
Acad Sci U S A 107(29):12986-12991.

Arun G, Akhade VS, Donakonda S, Rao MR. 2012. mrhl RNA, a long noncoding RNA, negatively regulates Wnt signaling through its protein partner Ddx5/p68 in mouse spermatogonial cells. Mol Cell Biol 32(15):3140-3152.
Autuoro JM, Pirnie SP, Carmichael GG. 2014. Long noncoding RNAs in imprinting and X chromosome inactivation. Biomolecules 4(1):76-100.

Banfai B, Jia H, Khatun J, Wood E, Risk B, Gundling WE, Kundaje A, Gunawardena HP, Yu YB, Xie L, Krajewski K, Strahl BD, Chen X, Bickel P, Giddings MC, Brown JB, Lipovich L. 2012. Long noncoding RNAs are rarely translated in two human cell lines. Genome Research 22(9):1646-1657.

Bao J, Wu J, Schuster AS, Hennig GW, Yan W. 2013. Expression profiling reveals developmentally regulated lncRNA repertoire in the mouse male germline. Biol Reprod 89(5):107.

Batista PJ, Chang HY. 2013. Long noncoding RNAs: cellular address codes in development and disease. Cell 152(6):1298-1307.
Batista PJ, Molinie B, Wang J, Qu K, Zhang J, Li L, Bouley DM, Lujan E,

Haddad B, Daneshvar K, Carter AC, Flynn RA, Zhou C, Lim KS, Dedon P, Wernig M, Mullen AC, Xing Y, Giallourakis CC, Chang HY. 2014. m(6)A RNA modification controls cell fate transition in mammalian embryonic stem cells. Cell Stem Cell 15(6):707-719.

Bertone P, Stolc V, Royce TE, Rozowsky JS, Urban AE, Zhu X, Rinn JL, Tongprasit W, Samanta M, Weissman S, Gerstein M, Snyder M. 2004. Global identification of

human transcribed sequences with genome tiling arrays. Science 306(5705):2242-2246.

Bliek J, Terhal P, van den Bogaard MJ, Maas S, Hamel B, Salieb-Beugelaar G, Simon M, Letteboer T, van der Smagt J, Kroes H, Mannens M. 2006. Hypomethylation of the H19 gene causes not only Silver-Russell syndrome (SRS) but also isolated asymmetry or an SRS-like phenotype. Am J Hum Genet 78(4):604-614.

Borsani G, Tonlorenzi R, Simmler MC, Dandolo L, Arnaud D, Capra V, Grompe M, Pizzuti A, Muzny D, Lawrence C, Willard HF, Avner P, Ballabio A. 1991. Characterization of a murine gene expressed from the inactive X chromosome. Nature 351(6324):325-329.

Brannan CI, Dees EC, Ingram RS, Tilghman SM. 1990. The product of the H19 gene may function as an RNA. Mol Cell Biol 10(1):28-36.

Briggs SF, Reijo Pera RA. 2014. X chromosome inactivation: recent advances and a look forward. Curr Opin Genet Dev 28:78-82.

Brockdorff N, Ashworth A, Kay GF, Cooper P, Smith S, McCabe VM, Norris DP, Penny GD, Patel D, Rastan S. 1991. Conservation of position and exclusive expression of mouse Xist from the inactive X chromosome. Nature 351(6324):329-331.

Brook M, Smith JW, Gray NK. 2009. The DAZL and PABP families: RNA-binding proteins with interrelated roles in translational control in oocytes. Reproduction 137(4):595-617.

Brown CJ, Ballabio A, Rupert JL, Lafreniere RG, Grompe M, Tonlorenzi R, Willard HF. 1991. A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. Nature 349(6304):38-44.

Brown JA, Bulkley D, Wang J, Valenstein ML, Yario TA, Steitz TA, Steitz JA. 2014a. Structural insights into the stabilization of MALAT1 noncoding RNA by a bipartite triple helix. Nat Struct Mol Biol 21(7):633-640.

Brown JB, Boley N, Eisman R, May GE, Stoiber MH, Duff MO, Booth BW, Wen J, Park S, Suzuki AM, Wan KH, Yu C, Zhang D, Carlson JW, Cherbas L, Eads BD, Miller D, Mockaitis K, Roberts J, Davis CA, Frise E, Hammonds AS, Olson S, Shenker S, Sturgill D, Samsonova AA, Weiszmann R, Robinson G, Hernandez J, Andrews J, Bickel PJ, Carninci P, Cherbas P, Gingeras TR, Hoskins RA, Kaufman TC, Lai EC, Oliver B, Perrimon N, Graveley BR, Celniker SE. 2014b. Diversity and dynamics of the Drosophila transcriptome. Nature 512(7515):393-399.

Bruce S, Hannula-Jouppi K, Peltonen J, Kere J, Lipsanen-Nyman M. 2009. Clinically

distinct epigenetic subgroups in Silver-Russell syndrome: the degree of H19 hypomethylation associates with phenotype severity and genital and skeletal anomalies. The Journal of clinical endocrinology and metabolism 94(2):579-587.

Caballero J, Gilbert I, Fournier E, Gagne D, Scantland S, Macaulay A, Robert C. 2014. Exploring the function of long non-coding RNA in the development of bovine early embryos. Reproduction, fertility, and development 27(1):40-52.

Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A, Rinn JL. 2011. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes Dev 25(18):1915-1927.

Carninci P, Kasukawa T, Katayama S, Gough J, Frith MC, Maeda N, Oyama R, *et al*. 2005. The transcriptional landscape of the mammalian genome. Science 309(5740):1559-1563.

Carrieri C, Cimatti L, Biagioli M, Beugnet A, Zucchelli S, Fedele S, Pesce E, Ferrer I, Collavin L, Santoro C, Forrest AR, Carninci P, Biffo S, Stupka E, Gustincich S. 2012. Long non-coding antisense RNA controls Uchl1 translation through an embedded SINEB2 repeat. Nature 491(7424):454-457.

Cech TR, Steitz JA. 2014. The noncoding RNA revolution-trashing old rules to forge new ones. Cell 157(1):77-94.

Cesana M, Cacchiarelli D, Legnini I, Santini T, Sthandier O, Chinappi M, Tramontano A, Bozzoni I. 2011. A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. Cell 147(2):358-369.

Cha BJ, Koppetsch BS, Theurkauf WE. 2001. In vivo analysis of Drosophila bicoid mRNA localization reveals a novel microtubule-dependent axis specification pathway. Cell 106(1):35-46.

Chakraborty S, Deb A, Maji RK, Saha S, Ghosh Z. 2014. LncRBase: An Enriched Resource for lncRNA Information. PLoS ONE 9(9):e108010.

Chakravarty D, Sboner A, Nair SS, Giannopoulou E, Li R, Hennig S, Mosquera JM, Pauwels J, Park K, Kossai M, MacDonald TY, Fontugne J, Erho N, Vergara IA, Ghadessi M, Davicioni E, Jenkins RB, Palanisamy N, Chen Z, Nakagawa S, Hirose T, Bander NH, Beltran H, Fox AH, Elemento O, Rubin MA. 2014. The oestrogen receptor alpha-regulated lncRNA NEAT1 is a critical modulator of prostate cancer. Nature communications 5:5383.

Chalmel F, Lardenois A, Evrard B, Rolland AD, Sallou O, Dumargne MC, Coiffec I, Collin O, Primig M, Jegou B. 2014. High-resolution profiling of novel transcribed regions during rat spermatogenesis. Biol Reprod 91(1):5.

Chen G, Wang Z, Wang D, Qiu C, Liu M, Chen X, Zhang Q, Yan G, Cui Q. 2013. LncRNADisease: a database for long-non-coding RNA-associated diseases. Nucleic Acids Res 41(Database issue):D983-986.

Cheng J, Kapranov P, Drenkow J, Dike S, Brubaker S, Patel S, Long J, Stern D, Tammana H, Helt G, Sementchenko V, Piccolboni A, Bekiranov S, Bailey DK, Ganesh M, Ghosh S, Bell I, Gerhard DS, Gingeras TR. 2005. Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. Science 308(5725):1149-1154.

Chew GL, Pauli A, Rinn JL, Regev A, Schier AF, Valen E. 2013. Ribosome profiling reveals resemblance between long non-coding RNAs and 5 ' leaders of coding RNAs. Development 140(13):2828-2834.

Chocu S, Evrard B, Lavigne R, Rolland AD, Aubry F, Jegou B, Chalmel F, Pineau C. 2014. Forty-four novel protein-coding loci discovered using a proteomics informed by transcriptomics (PIT) approach in rat male germ cells. Biol Reprod 91(5):123.

Chu C, Qu K, Zhong FL, Artandi SE, Chang HY. 2011. Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. Mol Cell 44(4):667-678.

Chu C, Spitale RC, Chang HY. 2015. Technologies to probe functions and mechanisms of long noncoding RNAs. Nat Struct Mol Biol 22(1):29-35.

Clemson CM, Hutchinson JN, Sara SA, Ensminger AW, Fox AH, Chess A, Lawrence JB. 2009. An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles. Mol Cell 33(6):717-726.

Collier B, Gorgoni B, Loveridge C, Cooke HJ, Gray NK. 2005. The DAZL family proteins are PABP-binding proteins that regulate translation in germ cells. Embo J 24(14):2656-2666.

Cossetti C, Lugini L, Astrologo L, Saggio I, Fais S, Spadafora C. 2014. Soma-to-Germline Transmission of RNA in Mice Xenografted with Human Tumour Cells: Possible Transport by Exosomes. PLoS ONE 9(7):e101629.

Davidovich C, Wang X, Cifuentes-Rojas C, Goodrich KJ, Gooding AR, Lee JT, Cech TR. 2015. Toward a consensus on the binding specificity and promiscuity of PRC2 for RNA. Mol Cell 57(3):552-558.

Davidovich C, Zheng L, Goodrich KJ, Cech TR. 2013. Promiscuous R
NA binding by Polycomb repressive complex 2. Nat Struct Mol Biol 20(11):1250-1257.

De Santa F, Barozzi I, Mietton F, Ghisletti S, Polletti S, Tusi BK, Muller H, Ragoussis J, Wei CL, Natoli G. 2010. A large fraction of extragenic RNA pol II transcription sites overlap enhancers. PLoS Biol 8(5):e1000384.

Deng X, Berletch JB, Nguyen DK, Disteche CM. 2014. X chromosome regulation: diverse patterns in development, tissues and disease. Nat Rev Genet 15(6):367-378.

Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, Guernec G, Martin D, Merkel A, Knowles DG, Lagarde J, Veeravalli L, Ruan X, Ruan Y, Lassmann T, Carninci P, Brown JB, Lipovich L, Gonzalez JM, Thomas M, Davis CA, Shiekhattar R, Gingeras TR, Hubbard TJ, Notredame C, Harrow J, Guigo R. 2012. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. Genome Res 22(9):1775-1789.

Di Ruscio A, Ebralidze AK, Benoukraf T, Amabile G, Goff LA, Terragni J, Figueroa ME, De Figueiredo Pontes LL, Alberich-Jorda M, Zhang P, Wu M, D/'Alo F, Melnick A, Leone G, Ebralidze KK, Pradhan S, Rinn JL, Tenen DG. 2013. DNMT1-interacting RNAs block gene-specific DNA methylation. Nature 503(7476):371-376.

Ding J, Lu Q, Ouyang Y, Mao H, Zhang P, Yao J, Xu C, Li X, Xiao J, Zhang Q. 2012. A long noncoding RNA regulates photoperiod-sensitive male sterility, an essential component of hybrid rice. Proc Natl Acad Sci U S A 109(7):2654-2659.

Dorweiler JE, Carey CC, Kubo KM, Hollick JB, Kermicle JL, Chandler VL. 2000. mediator of paramutation1 is required for establishment and maintenance of paramutation at multiple maize loci. Plant Cell 12(11):2101-2118.

Eggers S, DeBoer KD, van den Bergen J, Gordon L, White SJ, Jamsai D, McLachlan RI, Sinclair AH, O'Bryan MK. 2015. Copy number variation associated with meiotic arrest in idiopathic male infertility. Fertility and sterility 103(1):214-219.

Engreitz J, Lander ES, Guttman M. 2015. RNA antisense purification (RAP) for mapping RNA interactions with chromatin. Methods Mol Biol 1262:183-197.

Engreitz JM, Sirokman K, McDonel P, Shishkin AA, Surka C, Russell P, Grossman SR, Chow AY, Guttman M, Lander ES. 2014. RNA-RNA interactions enable specific targeting of noncoding RNAs to nascent Pre-mRNAs and chromatin sites. Cell 159(1):188-199.

Fatica A, Bozzoni I. 2014. Long non-coding RNAs: new players in cell differentiation and development. Nat Rev Genet 15(1):7-21.

Fitzpatrick GV, Soloway PD, Higgins MJ. 2002. Regional loss of imprinting and growth deficiency in mice with a targeted deletion of KvDMR1. Nat Genet 32:426 - 431.

Flynn RA, Chang HY. 2014. Long noncoding RNAs in cell-fate programming and reprogramming. Cell Stem Cell 14(6):752-761.

Franco-Zorrilla JM, Valli A, Todesco M, Mateos I, Puga MI, Rubio-Somoza I, Leyva A, Weigel D, Garcia JA, Paz-Ares J. 2007. Target mimicry provides a new mechanism for regulation of microRNA activity. Nat Genet 39(8):1033-1037.

Frost RJ, Hamra FK, Richardson JA, Qi X, Bassel-Duby R, Olson EN. 2010. MOV10L1 is necessary for protection of spermatocytes against retrotransposons by Piwi-interacting RNAs. Proc Natl Acad Sci U S A 107(26):11847-11852.

Fu Y, Dominissini D, Rechavi G, He C. 2014. Gene expression regulation mediated through reversible m(6)A RNA methylation. Nat Rev Genet 15(5):293-306.

Galupa R, Heard E. 2015. X-chromosome inactivation: new insights into cis and trans regulation. Curr Opin Genet Dev 31:57-66.

Gao WL, Liu M, Yang Y, Yang H, Liao Q, Bai Y, Li YX, Li D, Peng C, Wang YL. 2012. The imprinted H19 gene regulates human placental trophoblast cell proliferation via encoding miR-675 that targets Nodal Modulator 1 (NOMO1). RNA Biol 9(7):1002-1010.

Gapp K, Jawaid A, Sarkies P, Bohacek J, Pelczar P, Prados J, Farinelli L, Miska E, Mansuy IM. 2014. Implication of sperm RNAs in transgenerational inheritance of the effects of early trauma in mice. Nat Neurosci 17:667-669.

Geisler S, Coller J. 2013. RNA in unexpected places: long non-coding RNA functions in diverse cellular contexts. Nat Rev Mol Cell Biol 14(11):699-712.

Gelfand B, Mead J, Bruning A, Apostolopoulos N, Tadigotla V, Nagaraj V, Sengupta AM, Vershon AK. 2011. Regulated antisense transcription controls expression of cell-type-specific genes in yeast. Mol Cell Biol 31(8):1701-1709.

Gezer U, Ozgur E, Cetinkaya M, Isin M, Dalay N. 2014. Long non-coding RNAs with low expression levels in cells are enriched in secreted exosomes. Cell biology international 38(9):1076-1079.

Ghosal S, Das S, Chakrabarti J. 2013. Long noncoding RNAs: new players in the molecular mechanism for maintenance and differentiation of pluripotent stem cells. Stem cells and development 22(16):2240-2253.

Ginno PA, Lott PL, Christensen HC, Korf I, Chedin F. 2012. R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. Mol Cell 45(6):814-825.

Ginsburg M, Snow MH, McLaren A. 1990. Primordial germ cells in the mouse embryo during gastrulation. Development 110(2):521-528.

Golestaneh N, Beauchamp E, Fallen S, Kokkinaki M, Uren A, Dym M. 2009. Wnt signaling promotes proliferation and stemness regulation of spermatogonial stem/progenitor cells. Reproduction 138(1):151-162.

Gonzalez I, Munita R, Agirre E, Dittmer TA, Gysling K, Misteli T, Luco RF. 2015. A lncRNA regulates alternative splicing via establishment of a splicing-specific chromatin signature. Nat Struct Mol Biol 22(5):370-376.

Grandjean V, Gounon P, Wagner N, Martin L, Wagner KD, Bernex F, Cuzin F, Rassoulzadegan M. 2009. The miR-124-Sox9 paramutation: RNA-mediated epigenetic control of embryonic and adult growth. Development 136(21):3647-3655.

Gremlich S, Damnon F, Reymondin D, Braissant O, Schittny JC, Baud D, Gerber S, Roth-Kleiner M. 2014. The long non-coding RNA NEAT1 is increased in IUGR placentas, leading to potential new hypotheses of IUGR origin/development. Placenta 35(1):44-49.

Guttman M, Amit I, Garber M, French C, Lin MF, Feldser D, Huarte M, Zuk O, Carey BW, Cassady JP, Cabili MN, Jaenisch R, Mikkelsen TS, Jacks T, Hacohen N, Bernstein BE, Kellis M, Regev A, Rinn JL, Lander ES. 2009. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature 458(7235):223-227.

Guttman M, Donaghey J, Carey BW, Garber M, Grenier JK, Munson G, Young G, Lucas AB, Ach R, Bruhn L, Yang X, Amit I, Meissner A, Regev A, Rinn JL, Root DE, Lander ES. 2011. lincRNAs act in the circuitry controlling pluripotency and differentiation. Nature 477(7364):295-300.

Guttman M, Garber M, Levin JZ, Donaghey J, Robinson J, Adiconis X, Fan L, Koziol MJ, Gnirke A, Nusbaum C, Rinn JL, Lander ES, Regev A. 2010. Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. Nat Biotechnol 28(5):503-510.

Guttman M, Russell P, Ingolia NT, Weissman JS, Lander ES. 2013. Ribosome Profiling Provides Evidence that Large Noncoding RNAs Do Not Encode Proteins. Cell 154(1):240-251.

Gyory I, Fejer G, Ghosh N, Seto E, Wright KL. 2003. Identification of a functionally impaired positive regulatory domain I binding factor 1 transcription repressor in myeloma cell lines. J Immunol 170(6):3125-3133.

Hacisuleyman E, Goff LA, Trapnell C, Williams A, Henao-Mejia J, Sun L, McClanahan P, Hendrickson DG, Sauvageau M, Kelley DR, Morse M, Engreitz J, Lander ES, Guttman M, Lodish HF, Flavell R, Raj A, Rinn JL. 2014. Topological organization of multichromosomal regions by the long intergenic noncoding RNA Firre. Nat Struct Mol Biol 21(2):198-206.

Hall IM, Shankaranarayana GD, Noma K, Ayoub N, Cohen A, Grewal SI. 2002. Establishment and maintenance of a heterochromatin domain. Science 297(5590):2232-2237.

Hamazaki N, Uesaka M, Nakashima K, Agata K, Imamura T. 2015. Gene activation-associated long noncoding RNAs function in mouse preimplantation development. Development 142(5):910-920.

Han P, Li W, Lin CH, Yang J, Shang C, Nurnberg ST, Jin KK, Xu W, Lin CY, Lin CJ, Xiong Y, Chien HC, Zhou B, Ashley E, Bernstein D, Chen PS, Chen HS, Quertermous T, Chang CP. 2014. A long noncoding RNA protects the heart from pathological hypertrophy. Nature 514(7520):102-106.

Hansen TB, Jensen TI, Clausen BH, Bramsen JB, Finsen B, Damgaard CK, Kjems J. 2013. Natural RNA circles function as efficient microRNA sponges. Nature 495(7441):384-388.

Harigaya Y, Tanaka H, Yamanaka S, Tanaka K, Watanabe Y, Tsutsumi C, Chikashige Y, Hiraoka Y, Yamashita A, Yamamoto M. 2006. Selective elimination of messenger RNA prevents an incidence of untimely meiosis. Nature 442(7098):45-50.

Hatchell EC, Colley SM, Beveridge DJ, Epis MR, Stuart LM, Giles KM, Redfern AD, Miles LE, Barker A, MacDonald LM, Arthur PG, Lui JC, Golding JL, McCulloch RK, Metcalf CB, Wilce JA, Wilce MC, Lanz RB, O'Malley BW, Leedman PJ. 2006. SLIRP, a small SRA binding protein, is a nuclear receptor corepressor. Mol Cell 22(5):657-668.

He X, He Y, Xi B, Zheng J, Zeng X, Cai Q, Ouyang Y, Wang C, Zhou X, Huang H, Deng W, Xin S, Huang Q, Liu H. 2013. LncRNAs Expression in Preeclampsia Placenta Reveals the Potential Role of LncRNAs Contributing to Preeclampsia Pathogenesis. PLoS ONE 8(11):e81437.

Herman H, Lu M, Anggraini M, Sikora A, Chang Y, Yoon BJ, Soloway PD. 2003. Trans allele methylation and paramutation-like effects in mice. Nat Genet 34(2):199-202.

Herzog VA, Lempradl A, Trupke J, Okulski H, Altmutter C, Ruge F, Boidol B, Kubicek S, Schmauss G, Aumayr K, Ruf M, Pospisilik A, Dimond A, Senergin HB, Vargas ML, Simon JA, Ringrose L. 2014. A strand-specific switch in noncoding transcription

switches the function of a Polycomb/Trithorax response element. Nat Genet 46(9):973-981.

Holdt LM, Hoffmann S, Sass K, Langenberger D, Scholz M, Krohn K, Finstermeier K, Stahringer A, Wilfert W, Beutner F, Gielen S, Schuler G, Gabel G, Bergert H, Bechmann I, Stadler PF, Thiery J, Teupser D. 2013. Alu elements in ANRIL non-coding RNA at chromosome 9p21 modulate atherogenic cell functions through trans-regulation of gene networks. PLoS Genet 9(7):e1003588.

Holoch D, Moazed D. 2015. RNA-mediated epigenetic regulation of gene expression. Nat Rev Genet 16(2):71-84.

Hongay CF, Grisafi PL, Galitski T, Fink GR. 2006. Antisense transcription controls cell fate in Saccharomyces cerevisiae. Cell 127(4):735-745.

Hongay CF, Orr-Weaver TL. 2011. Drosophila Inducer of MEiosis 4 (IME4) is required for Notch signaling during oogenesis. Proc Natl Acad Sci U S A 108(36):14855-14860.

Hsieh CL, Fei T, Chen Y, Li T, Gao Y, Wang X, Sun T, Sweeney CJ, Lee GS, Chen S, Balk SP, Liu XS, Brown M, Kantoff PW. 2014. Enhancer RNAs participate in androgen receptor-driven looping that selectively enhances gene activation. Proc Natl Acad Sci U S A 111(20):7319-7324.

Hu W, Alvarez-Dominguez JR, Lodish HF. 2012. Regulation of mammalian cell differentiation by long non-coding RNAs. EMBO Rep 13(11):971-983.

Hung T, Wang Y, Lin MF, Koegel AK, Kotake Y, Grant GD, Horlings HM, Shah N, Umbricht C, Wang P, Wang Y, Kong B, Langerod A, Borresen-Dale AL, Kim SK, van de Vijver M, Sukumar S, Whitfield ML, Kellis M, Xiong Y, Wong DJ, Chang HY. 2011. Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. Nat Genet 43(7):621-629.

Hussein-Fikret S, Fuller PJ. 2005. Expression of nuclear receptor coregulators in ovarian stromal and epithelial tumours. Molecular and cellular endocrinology 229(1-2):149-160.

Ingolia NT, Ghaemmaghami S, Newman JR, Weissman JS. 2009. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science 324(5924):218-223.

Iyer MK, Niknafs YS, Malik R, Singhal U, Sahu A, Hosono Y, Barrette TR, Prensner JR, Evans JR, Zhao S, Poliakov A, Cao X, Dhanasekaran SM, Wu YM, Robinson DR, Beer DG, Feng FY, Iyer HK, Chinnaiyan AM. 2015. The landscape of long noncoding RNAs in the human transcriptome. Nat Genet 47(3):199-208.

Jenny A, Hachet O, Zavorszky P, Cyrklaff A, Weston MD, Johnston DS, Erdelyi M, Ephrussi A. 2006. A translation-independent role of oskar RNA in early Drosophila oogenesis. Development 133(15):2827-2833.

Jeon Y, Lee JT. 2011. YY1 tethers Xist RNA to the inactive X nucleation center. Cell 146(1):119-133.

Johnson GD, Mackie P, Jodar M, Moskovtsev S, Krawetz SA. 2015. Chromatin and extracellular vesicle associated sperm RNAs. Nucleic Acids Res.

Kallen AN, Zhou XB, Xu J, Qiao C, Ma J, Yan L, Lu L, Liu C, Yi JS, Zhang H, Min W, Bennett AM, Gregory RI, Ding Y, Huang Y. 2013. The imprinted H19 lncRNA antagonizes let-7 microRNAs. Mol Cell 52(1):101-112.

Kampa D, Cheng J, Kapranov P, Yamanaka M, Brubaker S, Cawley S, Drenkow J, Piccolboni A, Bekiranov S, Helt G, Tammana H, Gingeras TR. 2004. Novel RNAs identified from an in-depth analysis of the transcriptome of human chromosomes 21 and 22. Genome Res 14(3):331-342.

Kaneko S, Bonasio R, Saldana-Meyer R, Yoshida T, Son J, Nishino K, Umezawa A, Reinberg D. 2014. Interactions between JARID2 and Noncoding RNAs Regulate PRC2 Recruitment to Chromatin. Mol Cell 53:290-300.

Kaneko S, Son J, Shen SS, Reinberg D, Bonasio R. 2013. PRC2 binds active promoters and contacts nascent RNAs in embryonic stem cells. Nat Struct Mol Biol 20(11):1258-1264.

Kapranov P, Drenkow J, Cheng J, Long J, Helt G, Dike S, Gingeras TR. 2005. Examples of the complex architecture of the human transcriptome revealed by RACE and high-density tiling arrays. Genome Res 15(7):987-997.

Kapusta A, Kronenberg Z, Lynch VJ, Zhuo X, Ramsay L, Bourque G, Yandell M, Feschotte C. 2013. Transposable elements are major contributors to the origin, diversification, and regulation of vertebrate long noncoding RNAs. PLoS Genet 9(4):e1003470.

Kashimada K, Koopman P. 2010. Sry: the master switch in mammalian sex determination. Development 137(23):3921-3930.

Kawashima H, Takano H, Sugita S, Takahara Y, Sugimura K, Nakatani T. 2003. A novel steroid receptor co-activator protein (SRAP) as an alternative form of steroid receptor RNA-activator gene: expression in prostate cancer cells and enhancement of androgen receptor activity. Biochemical Journal 369(Pt 1):163-171.
Kee K, Angeles VT, Flores M, Nguyen HN, Reijo Pera RA. 2009. Human DAZL, DAZ

and BOULE genes modulate primordial germ-cell and haploid gamete formation. Nature 462(7270):222-225.

Keller AD, Maniatis T. 1991. Identification and characterization of a novel repressor of beta-interferon gene expression. Genes Dev 5(5):868-879.
Keller C, Kulasegaran-Shylini R, Shimada Y, Hotz HR, Buhler M. 2013. Noncoding RNAs prevent spreading of a repressive histone mark. Nat Struct Mol Biol 20(8):994-1000.

Kelley D, Rinn J. 2012. Transposable elements reveal a stem cell-specific class of long noncoding RNAs. Genome Biology 13(11):R107.

Keniry A, Oxley D, Monnier P, Kyba M, Dandolo L, Smits G, Reik W. 2012. The H19 lincRNA is a developmental reservoir of miR-675 that suppresses growth and Igf1r. Nat Cell Biol 14(7):659-665.

Khalil AM, Guttman M, Huarte M, Garber M, Raj A, Rivea Morales D, Thomas K, Presser A, Bernstein BE, van Oudenaarden A, Regev A, Lander ES, Rinn JL. 2009. Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. Proc Natl Acad Sci U S A 106(28):11667-11672.

Kiani J, Grandjean V, Liebers R, Tuorto F, Ghanbarian H, Lyko F, Cuzin F, Rassoulzadegan M. 2013. RNA-Mediated Epigenetic Heredity Requires the Cytosine Methyltransferase Dnmt2. PLoS Genet 9(5):e1003498.

Kim T, Xu Z, Clauder-Munster S, Steinmetz LM, Buratowski S. 2012. Set3 HDAC mediates effects of overlapping noncoding transcription on gene induction kinetics. Cell 150(6):1158-1169.

Kim TK, Hemberg M, Gray JM, Costa AM, Bear DM, Wu J, Harmin DA, Laptewicz M, Barbara-Haley K, Kuersten S, Markenscoff-Papadimitriou E, Kuhl D, Bito H, Worley PF, Kreiman G, Greenberg ME. 2010. Widespread transcription at neuronal activity-regulated enhancers. Nature 465(7295):182-187.

Kino T, Hurt DE, Ichijo T, Nader N, Chrousos GP. 2010. Noncoding RNA gas5 is a growth arrest- and starvation-associated repressor of the glucocorticoid receptor. Science signaling 3(107):ra8.

Klattenhoff CA, Scheuermann JC, Surface LE, Bradley RK, Fields PA, Steinhauser ML, Ding H, Butty VL, Torrey L, Haas S, Abo R, Tabebordbar M, Lee RT, Burge CB, Boyer LA. 2013. Braveheart, a long noncoding RNA required for cardiovascular lineage commitment. Cell 152(3):570-583.

Kloc M, Wilk K, Vargas D, Shirato Y, Bilinski S, Etkin LD. 2005. Potential structural

role of non-coding and coding RNAs in the organization of the cytoskeleton at the vegetal cortex of Xenopus oocytes. Development 132(15):3445-3457.

Kofron M, Demel T, Xanthos J, Lohr J, Sun B, Sive H, Osada S, Wright C, Wylie C, Heasman J. 1999. Mesoderm induction in Xenopus is a zygotic event regulated by maternal VegT via TGFbeta growth factors. Development 126(24):5759-5770.

Koltunow AM, Grossniklaus U. 2003. Apomixis: a developmental perspective. Annual review of plant biology 54:547-574.

Koopman P, Gubbay J, Vivian N, Goodfellow P, Lovell-Badge R. 1991. Male development of chromosomally female mice transgenic for Sry. Nature 351(6322):117-121.

Kotov AA, Akulenko NV, Kibanov MV, Olenina LV. 2014. DEAD-Box RNA helicases in animal gametogenesis. Molecular Biology 48(1):16-28.

Kretz M, Siprashvili Z, Chu C, Webster DE, Zehnder A, Qu K, Lee CS, Flockhart RJ, Groff AF, Chow J, Johnston D, Kim GE, Spitale RC, Flynn RA, Zheng GX, Aiyer S, Raj A, Rinn JL, Chang HY, Khavari PA. 2013. Control of somatic tissue differentiation by the long non-coding RNA TINCR. Nature 493(7431):231-235.

Kugler JM, Lasko P. 2009. Localization, anchoring and translational control of oskar, gurken, bicoid and nanos mRNA during Drosophila oogenesis. Fly 3(1):15-28.

Lai F, Orom UA, Cesaroni M, Beringer M, Taatjes DJ, Blobel GA, Shiekhattar R. 2013. Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. Nature 494(7438):497-501.

Laiho A, Kotaja N, Gyenesei A, Sironen A. 2013. Transcriptome profiling of the murine testis during the first wave of spermatogenesis. PLoS ONE 8(4):e61558.

Lam MT, Cho H, Lesch HP, Gosselin D, Heinz S, Tanaka-Oishi Y, Benner C, Kaikkonen MU, Kim AS, Kosaka M, Lee CY, Watt A, Grossman TR, Rosenfeld MG, Evans RM, Glass CK. 2013. Rev-Erbs repress macrophage gene expression by inhibiting enhancer-directed transcription. Nature 498(7455):511-515.

Lanz RB, McKenna NJ, Onate SA, Albrecht U, Wong J, Tsai SY, Tsai MJ, O'Malley BW. 1999. A steroid receptor coactivator, SRA, functions as an RNA and is present in an SRC-1 complex. Cell 97(1):17-27.

Lanz RB, Razani B, Goldberg AD, O'Malley BW. 2002. Distinct RNA motifs are important for coactivation of steroid hormone receptors by steroid receptor RNA activator (SRA). Proc Natl Acad Sci U S A 99(25):16081-16086.

Lardenois A, Liu Y, Walther T, Chalmel F, Evrard B, Granovskaia M, Chu A, Davis RW, Steinmetz LM, Primig M. 2011. Execution of the meiotic noncoding RNA expression program and the onset of gametogenesis in yeast require the conserved exosome subunit Rrp6. Proc Natl Acad Sci U S A 108(3):1058-1063.

Latos PA, Pauler FM, Koerner MV, Senergin HB, Hudson QJ, Stocsits RR, Allhoff W, Stricker SH, Klement RM, Warczok KE, Aumayr K, Pasierbek P, Barlow DP. 2012. Airn transcriptional overlap, but not its lncRNA products, induces imprinted Igf2r silencing. Science 338(6113):1469-1472.

Lee JT, Bartolomei MS. 2013. X-inactivation, imprinting, and long noncoding RNAs in health and disease. Cell 152(6):1308-1323.

Lee MP, DeBaun MR, Mitsuya K, Galonek HL, Brandenburg S, Oshimura M, Feinberg AP. 1999. Loss of imprinting of a paternally expressed transcript, with antisense orientation to KVLQT1, occurs frequently in Beckwith-Wiedemann syndrome and is independent of insulin-like growth factor II imprinting. Proc Natl Acad Sci U S A 96(9):5203-5208.

Leucci E, Patella F, Waage J, Holmstrom K, Lindow M, Porse B, Kauppinen S, Lund AH. 2013. microRNA-9 targets the long non-coding RNA MALAT1 for degradation in the nucleus. Scientific reports 3:2535.

Leveille N, Melo CA, Rooijers K, Diaz-Lagares A, Melo SA, Korkmaz G, Lopes R, Akbari Moqadam F, Maia AR, Wijchers PJ, Geeven G, den Boer ML, Kalluri R, de Laat W, Esteller M, Agami R. 2015. Genome-wide profiling of p53-regulated enhancer RNAs uncovers a subset of enhancers controlled by a lncRNA. Nature communications 6:6520.

Leygue E, Dotzlaw H, Watson PH, Murphy LC. 1999. Expression of the steroid receptor RNA activator in human breast tumors. Cancer Res 59(17):4190-4193.

Li S, Mason CE. 2014. The pivotal regulatory landscape of RNA modifications. Annual review of genomics and human genetics 15:127-150.

Li W, Notani D, Ma Q, Tanasa B, Nunez E, Chen AY, Merkurjev D, Zhang J, Ohgi K, Song X, Oh S, Kim H-S, Glass CK, Rosenfeld MG. 2013. Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. Nature 498(7455):516-520.

Liang M, Li W, Tian H, Hu T, Wang L, Lin Y, Li Y, Huang H, Sun F. 2014. Sequential expression of long noncoding RNA as mRNA gene expression in specific stages of mouse spermatogenesis. Scientific reports 4:5966.

Liang Q, Xu Z, Xu R, Wu L, Zheng S. 2012. Expression patterns of non-coding

spliced transcripts from human endogenous retrovirus HERV-H elements in colon cancer. PLoS ONE 7(1):e29950.

Liu N, Dai Q, Zheng G, He C, Parisien M, Pan T. 2015a. N(6)-methyladenosine-dependent RNA structural switches regulate RNA-protein interactions. Nature 518(7540):560-564.

Liu Z, Hao C, Huang X, Zhang N, Bao H, Qu Q. 2015b. Peripheral blood leukocyte expression level of lncRNA steroid receptor RNA activator (SRA) and its association with polycystic ovary syndrome: a case control study. Gynecological endocrinology : the official journal of the International Society of Gynecological Endocrinology 31(5):363-368.

Liu Z, Hao C, Song D, Zhang N, Bao H, Qu Q. 2014. Androgen Receptor Coregulator CTBP1-AS Is Associated With Polycystic Ovary Syndrome in Chinese Women: A Preliminary Study. Reprod Sci.

Luo X, Pan J, Wang L, Wang P, Zhang M, Liu M, Dong Z, Meng Q, Tao X, Zhao X, Zhong J, Ju W, Gu Y, Jenkins EC, Brown WT, Shi Q, Zhong N. 2015. Epigenetic regulation of lncRNA connects ubiquitin-proteasome system with infection-inflammation in preterm births and preterm premature rupture of membranes. BMC pregnancy and childbirth 15:35.

Luo X, Shi Q, Gu Y, Pan J, Hua M, Liu M, Dong Z, Zhang M, Wang L, Gu Y, Zhong J, Zhao X, Jenkins EC, Brown WT, Zhong N. 2013. LncRNA pathway involved in premature preterm rupture of membrane (PPROM): an epigenomic approach to study the pathogenesis of reproductive disorders. PLoS ONE 8(11):e79897.

Lyle R, Watanabe D, te Vruchte D, Lerchner W, Smrzka OW, Wutz A, Schageman J, Hahner L, Davies C, Barlow DP. 2000. The imprinted antisense RNA at the Igf2r locus overlaps but does not imprint Mas1. Nat Genet 25(1):19-21.

Ma L, Li A, Zou D, Xu X, Xia L, Yu J, Bajic VB, Zhang Z. 2015. LncRNAWiki: harnessing community knowledge in collaborative curation of human long non-coding RNAs. Nucleic Acids Res 43(Database issue):D187-192.

Macaulay AD, Gilbert I, Caballero J, Barreto R, Fournier E, Tossou P, Sirard MA, Clarke HJ, Khandjian EW, Richard FJ, Hyttel P, Robert C. 2014. The gametic synapse: RNA transfer to the bovine oocyte. Biol Reprod 91(4):90.

Magnusdottir E, Dietmann S, Murakami K, Gunesdogan U, Tang F, Bao S, Diamanti E, Lao K, Gottgens B, Azim Surani M. 2013. A tripartite transcription factor network regulates primordial germ cell specification in mice. Nat Cell Biol 15(8):905-915.

Marahrens Y, Panning B, Dausman J, Strauss W, Jaenisch R. 1997. Xist-deficient

mice are defective in dosage compensation but not spermatogenesis. Genes Dev 11(2):156-166.

Margolin G, Khil PP, Kim J, Bellani MA, Camerini-Otero RD. 2014. Integrated transcriptome analysis of mouse spermatogenesis. BMC Genomics 15:39.
Margueron R, Reinberg D. 2011. The Polycomb complex PRC2 and its mark in life. Nature 469(7330):343-349.

Martianov I, Ramadass A, Serra Barros A, Chow N, Akoulitchev A. 2007. Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript. Nature 445:666-670.

Matson CK, Zarkower D. 2012. Sex and the singular DM domain: insights into sexual regulation, evolution and plasticity. Nat Rev Genet 13(3):163-174.
Mattick JS, Rinn JL. 2015. Discovery and annotation of long noncoding RNAs. Nat Struct Mol Biol 22(1):5-7.

Mau M, Corral JM, Vogel H, Melzer M, Fuchs J, Kuhlmann M, de Storme N, Geelen D, Sharbel TF. 2013. The Conserved Chimeric Transcript UPGRADE2 Is Associated with Unreduced Pollen Formation and Is Exclusively Found in Apomictic Boechera Species. Plant physiology 163(4):1640-1659.

McHugh CA, Chen CK, Chow A, Surka CF, Tran C, McDonel P, Pandya-Jones A, Blanco M, Burghard C, Moradian A, Sweredoski MJ, Shishkin AA, Su J, Lander ES, Hess S, Plath K, Guttman M. 2015. The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. Nature 521(7551):232-236.

Medrano JV, Ramathal C, Nguyen HN, Simon C, Reijo Pera RA. 2012. Divergent RNA-binding proteins, DAZL and VASA, induce meiotic progression in human germ cells derived in vitro. Stem cells (Dayton, Ohio) 30(3):441-451.

Memczak S, Jens M, Elefsinioti A, Torti F, Krueger J, Rybak A, Maier L, Mackowiak SD, Gregersen LH, Munschauer M, Loewer A, Ziebold U, Landthaler M, Kocks C, le Noble F, Rajewsky N. 2013. Circular RNAs are a large class of animal RNAs with regulatory potency. Nature 495(7441):333-338.

Mercer TR, Wilhelm D, Dinger ME, Solda G, Korbie DJ, Glazov EA, Truong V, Schwenke M, Simons C, Matthaei KI, Saint R, Koopman P, Mattick JS. 2011. Expression of distinct RNAs from 3' untranslated regions. Nucleic Acids Res 39(6):2393-2403.

Monnier P, Martinet C, Pontis J, Stancheva I, Ait-Si-Ali S, Dandolo L. 2013. H19 lncRNA controls gene expression of the Imprinted Gene Network by recruiting MBD1. Proc Natl Acad Sci U S A 110(51):20693-20698.
Morris KV, Mattick JS. 2014. The rise of regulatory RNA. Nat Rev Genet 15:423-437.

Mousavi K, Zare H, Dell'orso S, Grontved L, Gutierrez-Cruz G, Derfoul A, Hager GL, Sartorelli V. 2013. eRNAs promote transcription by establishing chromatin accessibility at defined genomic loci. Mol Cell 51(5):606-617.

Mulvey BB, Olcese U, Cabrera JR, Horabin JI. 2014. An interactive network of long non-coding RNAs facilitates the Drosophila sex determination decision. Biochim Biophys Acta 1839(9):773-784.

Murphy LC, Simon SL, Parkes A, Leygue E, Dotzlaw H, Snell L, Troup S, Adeyinka A, Watson PH. 2000. Altered expression of estrogen receptor coregulators during human breast tumorigenesis. Cancer Res 60(22):6266-6271.

Nagamori I, Cruickshank VA, Sassone-Corsi P. 2011. Regulation of an RNA granule during spermatogenesis: acetylation of MVH in the chromatoid body of germ cells. J Cell Sci 124:4346-4355.

Nagano T, Mitchell JA, Sanz LA, Pauler FM, Ferguson-Smith AC, Feil R, Fraser P. 2008. The Air Noncoding RNA Epigenetically Silences Transcription by Targeting G9a to Chromatin. Science 322:1717-1720.

Nakagawa S, Shimada M, Yanaka K, Mito M, Arai T, Takahashi E, Fujita Y, Fujimori T, Standaert L, Marine JC, Hirose T. 2014. The lncRNA Neat1 is required for corpus luteum formation and the establishment of pregnancy in a subpopulation of mice. Development 141(23):4618-4627.

Nicolas E, Chenouard N, Olivo-Marin JC, Guichet A. 2009. A dual role for actin and microtubule cytoskeleton in the transport of Golgi units from the nurse cells to the oocyte across ring canals. Mol Biol Cell 20(1):556-568.

Nolasco S, Bellido J, Goncalves J, Tavares A, Zabala JC, Soares H. 2012. The expression of tubulin cofactor A (TBCA) is regulated by a noncoding antisense Tbca RNA during testis maturation. PLoS ONE 7(8):e42536.

Orom UA, Derrien T, Beringer M, Gumireddy K, Gardini A, Bussotti G, Lai F, Zytnicki M, Notredame C, Huang Q, Guigo R, Shiekhattar R. 2010. Long noncoding RNAs with enhancer-like function in human cells. Cell 143(1):46-58.

Pandey RR, Mondal T, Mohammad F, Enroth S, Redrup L, Komorowski J, Nagano T, Mancini-Dinardo D, Kanduri C. 2008. Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. Mol Cell 32(2):232-246.

Paranjpe SS, Jacobi UG, van Heeringen SJ, Veenstra GJ. 2013. A genome-wide survey of maternal and embryonic transcripts during Xenopus tropicalis development.

BMC Genomics 14:762.

Park YJ, Herman H, Gao Y, Lindroth AM, Hu BY, Murphy PJ, Putnam JR, Soloway PD. 2012. Sequences sufficient for programming imprinted germline DNA methylation defined. PLoS ONE 7(3):e33024.

Pefanis E, Wang J, Rothschild G, Lim J, Kazadi D, Sun J, Federation A, Chao J, Elliott O, Liu ZP, Economides AN, Bradner JE, Rabadan R, Basu U. 2015. RNA exosome-regulated long non-coding RNA transcription controls super-enhancer activity. Cell 161(4):774-789.

Penny GD, Kay GF, Sheardown SA, Rastan S, Brockdorff N. 1996. Requirement for Xist in X chromosome inactivation. Nature 379(6561):131-137.

Petkovic S, Muller S. 2015. RNA circularization strategies in vivo and in vitro. Nucleic Acids Research 43(4):2454-2465.

Prensner JR, Iyer MK, Sahu A, Asangani IA, Cao Q, Patel L, Vergara IA, Davicioni E, Erho N, Ghadessi M, Jenkins RB, Triche TJ, Malik R, Bedenis R, McGregor N, Ma T, Chen W, Han S, Jing X, Cao X, Wang X, Chandler B, Yan W, Siddiqui J, Kunju LP, Dhanasekaran SM, Pienta KJ, Feng FY, Chinnaiyan AM. 2013. The long noncoding RNA SChLAP1 promotes aggressive prostate cancer and antagonizes the SWI/SNF complex. Nat Genet 45(11):1392-1398.

Quek XC, Thomson DW, Maag JL, Bartonicek N, Signal B, Clark MB, Gloss BS, Dinger ME. 2015. lncRNAdb v2.0: expanding the reference database for functional long noncoding RNAs. Nucleic Acids Res 43(Database issue):D168-173.

Quenet D, Dalal Y. 2014. A long non-coding RNA is required for targeting centromeric protein A to the human centromere. eLife 3:e03254.

Quinodoz S, Guttman M. 2014. Long noncoding RNAs: an emerging link between gene regulation and nuclear organization. Trends in Cell Biology 24(11):651-663.

Rakyan V, Whitelaw E. 2003. Transgenerational epigenetic inheritance. Curr Biol 13(1):R6.

Rassoulzadegan M, Grandjean V, Gounon P, Vincent S, Gillot I, Cuzin F. 2006. RNA-mediated non-mendelian inheritance of an epigenetic change in the mouse. Nature 441(7092):469-474.

Reijo RA, Dorfman DM, Slee R, Renshaw AA, Loughlin KR, Cooke H, Page DC. 2000. DAZ family proteins exist throughout male germ cell development and transit from nucleus to cytoplasm at meiosis in humans and mice. Biol Reprod 63(5):1490-1496.

Rinn JL, Euskirchen G, Bertone P, Martone R, Luscombe NM, Hartman S, Harrison PM, Nelson FK, Miller P, Gerstein M, Weissman S, Snyder M. 2003. The transcriptional activity of human Chromosome 22. Genes Dev 17(4):529-540.

Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Brugmann SA, Goodnough LH, Helms JA, Farnham PJ, Segal E, Chang HY. 2007. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell 129(7):1311-1323.

RNAcentral-Consortium. 2015. RNAcentral: an international database of ncRNA sequences. Nucleic Acids Res 43(Database issue):D123-129.
Ruiz-Orera J, Messeguer X, Subirana JA, Alba MM. 2014. Long non-coding RNAs as a source of new peptides. eLife 3:e03523.

Sandbacka M, Bruce S, Halttunen M, Puhakka M, Lahermo P, Hannula-Jouppi K, Lipsanen-Nyman M, Kere J, Aittomaki K, Laivuori H. 2011. Methylation of H19 and its imprinted control region (H19 ICR1) in Mullerian aplasia. Fertility and sterility 95(8):2703-2706.

Santoro F, Mayer D, Klement RM, Warczok KE, Stukalov A, Barlow DP, Pauler FM. 2013. Imprinted Igf2r silencing depends on continuous Airn lncRNA expression and is not restricted to a developmental window. Development 140(6):1184-1195.

Sarma K, Cifuentes-Rojas C, Ergun A, Del Rosario A, Jeon Y, White F, Sadreyev R, Lee JT. 2014. ATRX directs binding of PRC2 to Xist RNA and Polycomb targets. Cell 159(4):869-883.

Sauvageau M, Goff LA, Lodato S, Bonev B, Groff AF, Gerhardinger C, Sanchez-Gomez DB, Hacisuleyman E, Li E, Spence M, Liapis SC, Mallard W, Morse M, Swerdel MR, D'Ecclessis MF, Moore JC, Lai V, Gong G, Yancopoulos GD, Frendewey D, Kellis M, Hart RP, Valenzuela DM, Arlotta P, Rinn JL. 2013. Multiple knockout mouse models reveal lincRNAs are required for life and brain development. eLife 2:e01749.

Schaukowitch K, Joo JY, Liu X, Watts JK, Martinez C, Kim TK. 2014. Enhancer RNA Facilitates NELF Release from Immediate Early Genes. Mol Cell 56(1):29-42.

Schmitz KM, Mayer C, Postepska A, Grummt I. 2010. Interaction of noncoding RNA with the rDNA promoter mediates recruitment of DNMT3b and silencing of rRNA genes. Genes Dev 24(20):2264-2269.

Schutt C, Nothiger R. 2000. Structure, function and evolution of sex-determining systems in Dipteran insects. Development 127(4):667-677.

Schwartz S, Agarwala SD, Mumbach MR, Jovanovic M, Mertins P, Shishkin A, Tabach Y, Mikkelsen TS, Satija R, Ruvkun G, Carr SA, Lander ES, Fink GR, Regev A. 2013. High-resolution mapping reveals a conserved, widespread, dynamic mRNA methylation program in yeast meiosis. Cell 155(6):1409-1421.

Schwartz S, Bernstein DA, Mumbach MR, Jovanovic M, Herbst RH, Leon-Ricardo BX, Engreitz JM, Guttman M, Satija R, Lander ES, Fink G, Regev A. 2014. Transcriptome-wide mapping reveals widespread dynamic-regulated pseudouridylation of ncRNA and mRNA. Cell 159(1):148-162.

Shah JC, Clancy MJ. 1992. IME4, a gene that mediates MAT and nutritional control of meiosis in Saccharomyces cerevisiae. Mol Cell Biol 12(3):1078-1086.

Shi Y, Downes M, Xie W, Kao HY, Ordentlich P, Tsai CC, Hon M, Evans RM. 2001. Sharp, an inducible cofactor that integrates nuclear receptor repression and activation. Genes Dev 15(9):1140-1151.

Shichino Y, Yamashita A, Yamamoto M. 2014. Meiotic long non-coding meiRNA accumulates as a dot at its genetic locus facilitated by Mmi1 and plays as a decoy to lure Mmi1. Open biology 4(6):140022.

Shimada T, Yamashita A, Yamamoto M. 2003. The fission yeast meiotic regulator Mei2p forms a dot structure in the horse-tail nucleus in association with the sme2 locus on chromosome II. Mol Biol Cell 14(6):2461-2469.

Shiraki T, Kondo S, Katayama S, Waki K, Kasukawa T, Kawaji H, Kodzius R, Watahiki A, Nakamura M, Arakawa T, Fukuda S, Sasaki D, Podhajska A, Harbers M, Kawai J, Carninci P, Hayashizaki Y. 2003. Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage. Proc Natl Acad Sci U S A 100(26):15776-15781.

Skourti-Stathaki K, Kamieniarz-Gdula K, Proudfoot NJ. 2014. R-loops induce repressive chromatin marks over mammalian gene terminators. Nature 516(7531):436-439.

Sleutels F, Zwart R, Barlow DP. 2002. The non-coding Air RNA is required for silencing autosomal imprinted genes. Nature 415(6873):810-813.

Smith MA, Gesell T, Stadler PF, Mattick JS. 2013. Widespread purifying selection on RNA structure in mammals. Nucleic Acids Res 41(17):8220-8236.

Smorag L, Xu X, Engel W, Pantakani DV. 2014. The roles of DAZL in RNA biology and development. Wiley interdisciplinary reviews RNA 5(4):527-535.

Somarowthu S, Legiewicz M, Chillon I, Marcia M, Liu F, Pyle AM. 2015. HOTAIR

Forms an Intricate and Modular Secondary Structure. Mol Cell 58(2):353-361.

Soumillon M, Necsulea A, Weier M, Brawand D, Zhang X, Gu H, Barthes P, Kokkinaki M, Nef S, Gnirke A, Dym M, de Massy B, Mikkelsen TS, Kaessmann H. 2013. Cellular source and mechanisms of high transcriptome complexity in the mammalian testis. Cell Rep 3(6):2179-2190.

Stam M, Belele C, Dorweiler JE, Chandler VL. 2002. Differential chromatin structure within a tandem array 100 kb upstream of the maize b1 locus is associated with paramutation. Genes Dev 16(15):1906-1918.

Sun M, Kraus WL. 2013. Minireview: Long noncoding RNAs: new "links" between gene expression and cellular outcomes in endocrinology. Mol Endocrinol 27(9):1390-1402.

Tay Y, Rinn J, Pandolfi PP. 2014. The multilayered complexity of ceRNA crosstalk and competition. Nature 505(7483):344-352.

Trimarchi T, Bilal E, Ntziachristos P, Fabbri G, Dalla-Favera R, Tsirigos A, Aifantis I. 2014. Genome-wide mapping and characterization of Notch-regulated long noncoding RNAs in acute leukemia. Cell 158(3):593-606.

Tripathi V, Ellis JD, Shen Z, Song DY, Pan Q, Watt AT, Freier SM, Bennett CF, Sharma A, Bubulya PA, Blencowe BJ, Prasanth SG, Prasanth KV. 2010. The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. Mol Cell 39(6):925-938.

Troy A, Sharpless NE. 2012. Genetic "lnc"-age of noncoding RNAs to human disease. J Clin Invest 122(11):3837-3840.

Tsai MC, Manor O, Wan Y, Mosammaparast N, Wang JK, Lan F, Shi Y, Segal E, Chang HY. 2010. Long noncoding RNA as modular scaffold of histone modification complexes. Science 329(5992):689-693.

Tuck AC, Tollervey D. 2013. A transcriptome-wide atlas of RNP composition reveals diverse classes of mRNAs and lncRNAs. Cell 154(5):996-1009.

Ulitsky I, Bartel DP. 2013. lincRNAs: genomics, evolution, and mechanisms. Cell 154(1):26-46.

van Dijk M, Thulluru HK, Mulders J, Michel OJ, Poutsma A, Windhorst S, Kleiverda G, Sie D, Lachmeijer AM, Oudejans CB. 2012. HELLP babies link a novel lincRNA to the trophoblast cell cycle. J Clin Invest 122(11):4003-4011.

van Werven FJ, Neuert G, Hendrick N, Lardenois A, Buratowski S, van Oudenaarden

A, Primig M, Amon A. 2012. Transcription of two long noncoding RNAs mediates mating-type control of gametogenesis in budding yeast. Cell 150(6):1170-1181.

Vicent GP, Nacht AS, Zaurin R, Font-Mateu J, Soronellas D, Le Dily F, Reyes D, Beato M. 2013. Unliganded progesterone receptor-mediated targeting of an RNA-containing repressive complex silences a subset of hormone-inducible genes. Genes Dev 27(10):1179-1197.

Volders PJ, Verheggen K, Menschaert G, Vandepoele K, Martens L, Vandesompele J, Mestdagh P. 2015a. An update on LNCipedia: a database for annotated human lncRNA sequences. Nucleic Acids Res 43(8):4363-4364.

Volders PJ, Verheggen K, Menschaert G, Vandepoele K, Martens L, Vandesompele J, Mestdagh P. 2015b. An update on LNCipedia: a database for annotated human lncRNA sequences. Nucleic Acids Res 43(Database issue):D174-180.

Volpe TA, Kidner C, Hall IM, Teng G, Grewal SI, Martienssen RA. 2002. Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. Science 297(5588):1833-1837.

Vourekas A, Zheng K, Fu Q, Maragkakis M, Alexiou P, Ma J, Pillai RS, Mourelatos Z, Wang PJ. 2015. The RNA helicase MOV10L1 binds piRNA precursors to initiate piRNA processing. Genes Dev 29(6):617-629.

Wagner KD, Wagner N, Ghanbarian H, Grandjean V, Gounon P, Cuzin F, Rassoulzadegan M. 2008. RNA induction and inheritance of epigenetic cardiac hypertrophy in the mouse. Dev Cell 14(6):962-969.

Wang J, Liu X, Wu H, Ni P, Gu Z, Qiao Y, Chen N, Sun F, Fan Q. 2010. CREB up-regulates long non-coding RNA, HULC expression through interaction with microRNA-372 in liver cancer. Nucleic Acids Res 38(16):5366-5383.

Wang KC, Chang HY. 2011. Molecular mechanisms of long noncoding RNAs. Mol Cell 43(6):904-914.

Wang KC, Yang YW, Liu B, Sanyal A, Corces-Zimmerman R, Chen Y, Lajoie BR, Protacio A, Flynn RA, Gupta RA, Wysocka J, Lei M, Dekker J, Helms JA, Chang HY. 2011. A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. Nature 472(7341):120-124.

Wang X, Arai S, Song X, Reichart D, Du K, Pascual G, Tempst P, Rosenfeld MG, Glass CK, Kurokawa R. 2008. Induced ncRNAs allosterically modify RNA-binding proteins in cis to inhibit transcription. Nature 454(7200):126-130.

Wang X, He C. 2014. Dynamic RNA modifications in posttranscriptional regulation.

Mol Cell 56(1):5-12.

Wang Y, Xu Z, Jiang J, Xu C, Kang J, Xiao L, Wu M, Xiong J, Guo X, Liu H. 2013. Endogenous miRNA sponge lincRNA-RoR regulates Oct4, Nanog, and Sox2 in human embryonic stem cell self-renewal. Dev Cell 25(1):69-80.

Watanabe M, Yanagisawa J, Kitagawa H, Takeyama K, Ogawa S, Arao Y, Suzawa M, Kobayashi Y, Yano T, Yoshikawa H, Masuhiro Y, Kato S. 2001. A subfamily of RNA-binding DEAD-box proteins acts as an estrogen receptor alpha coactivator through the N-terminal activation domain (AF-1) with an RNA coactivator, SRA. Embo J 20(6):1341-1352.

Watanabe T, Tomizawa S, Mitsuya K, Totoki Y, Yamamoto Y, Kuramochi-Miyagawa S, Iida N, Hoki Y, Murphy PJ, Toyoda A, Gotoh K, Hiura H, Arima T, Fujiyama A, Sado T, Shibata T, Nakano T, Lin H, Ichiyanagi K, Soloway PD, Sasaki H. 2011. Role for piRNAs and noncoding RNA in de novo DNA methylation of the imprinted mouse *Rasgrf1* locus. Science 332(6031):848-852.

Watanabe Y, Shinozaki-Yabana S, Chikashige Y, Hiraoka Y, Yamamoto M. 1997. Phosphorylation of RNA-binding protein controls cell cycle switch from mitotic to meiotic in fission yeast. Nature 386(6621):187-190.

Watanabe Y, Yamamoto M. 1994. S. pombe mei2+ encodes an RNA-binding protein essential for premeiotic DNA synthesis and meiosis I, which cooperates with a novel RNA species meiRNA. Cell 78(3):487-498.

West JA, Davis CP, Sunwoo H, Simon MD, Sadreyev RI, Wang PI, Tolstorukov MY, Kingston RE. 2014. The long noncoding RNAs NEAT1 and MALAT1 bind active chromatin sites. Mol Cell 55(5):791-802.

Xanthos JB, Kofron M, Wylie C, Heasman J. 2001. Maternal VegT is the initiator of a molecular network specifying endoderm in Xenopus laevis. Development 128(2):167-180.

Xie C, Yuan J, Li H, Li M, Zhao G, Bu D, Zhu W, Wu W, Chen R, Zhao Y. 2014. NONCODEv4: exploring the world of long non-coding RNA genes. Nucleic Acids Res 42(Database issue):D98-103.

Xing Z, Lin A, Li C, Liang K, Wang S, Liu Y, Park PK, Qin L, Wei Y, Hawke DH, Hung MC, Lin C, Yang L. 2014. lncRNA directs cooperative epigenetic regulation downstream of chemokine signals. Cell 159(5):1110-1125.

Xu N, Tsai CL, Lee JT. 2006. Transient homologous chromosome pairing marks the onset of X inactivation. Science 311(5764):1149-1152.

Xu XF, Li J, Cao YX, Chen DW, Zhang ZG, He XJ, Ji DM, Chen BL. 2014. Differential Expression of Long Noncoding RNAs in Human Cumulus Cells Related to Embryo Developmental Potential: A Microarray Analysis. Reprod Sci 22:672-678.

Yamashita A, Watanabe Y, Nukina N, Yamamoto M. 1998. RNA-assisted nuclear transport of the meiotic regulator Mei2p in fission yeast. Cell 95(1):115-123.

Yan L, Yang M, Guo H, Yang L, Wu J, Li R, Liu P, Lian Y, Zheng X, Yan J, Huang J, Li M, Wu X, Wen L, Lao K, Li R, Qiao J, Tang F. 2013. Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. Nat Struct Mol Biol 20(9):1131-1139.

Yang F, Deng X, Ma W, Berletch JB, Rabaia N, Wei G, Moore JM, Filippova GN, Xu J, Liu Y, Noble WS, Shendure J, Disteche CM. 2015. The lncRNA Firre anchors the inactive X chromosome to the nucleolus by binding CTCF and maintains H3K27me3 methylation. Genome Biol 16:52.

Yang L, Lin C, Jin C, Yang JC, Tanasa B, Li W, Merkurjev D, Ohgi KA, Meng D, Zhang J, Evans CP, Rosenfeld MG. 2013. lncRNA-dependent mechanisms of androgen-receptor-regulated gene activation programs. Nature 500(7464):598-602.

Yang YW, Flynn RA, Chen Y, Qu K, Wan B, Wang KC, Lei M, Chang HY. 2014. Essential role of lncRNA binding for WDR5 maintenance of active chromatin and embryonic stem cell pluripotency. eLife 3:e02046.

Yao B, Jin P. 2014. Unlocking epigenetic codes in neurogenesis. Genes Dev 28(12):1253-1271.

Yeh JR, Zhang X, Nagano MC. 2011. Wnt5a is a cell-extrinsic factor that supports self-renewal of mouse spermatogonial stem cells. J Cell Sci 124(Pt 14):2357-2366.

Yerushalmi GM, Salmon-Divon M, Yung Y, Maman E, Kedem A, Ophir L, Elemento O, Coticchio G, Dal Canto M, Mignini Renzinu M, Fadini R, Hourvitz A. 2014. Characterization of the human cumulus cell transcriptome during final follicular maturation and ovulation. Mol Hum Reprod 20(8):719-735.

Yin QF, Yang L, Zhang Y, Xiang JF, Wu YW, Carmichael GG, Chen LL. 2012. Long noncoding RNAs with snoRNA ends. Mol Cell 48(2):219-230.

Yokoo M, Sato E. 2004. Cumulus-oocyte complex interactions during oocyte maturation. International review of cytology 235:251-291.

Yoon JH, Abdelmohsen K, Srikantan S, Yang X, Martindale JL, De S, Huarte M, Zhan M, Becker KG, Gorospe M. 2012. LincRNA-p21 suppresses target mRNA translation. Mol Cell 47(4):648-655.

Yuan S, Oliver D, Schuster A, Zheng H, Yan W. 2015. Breeding scheme and maternal small RNAs affect the efficiency of transgenerational inheritance of a paramutation in mice. Scientific reports 5:9266.

Zhang B, Arun G, Mao YS, Lazar Z, Hung G, Bhattacharjee G, Xiao X, Booth CJ, Wu J, Zhang C, Spector DL. 2012. The lncRNA Malat1 is dispensable for mouse development but its transcription plays a cis-regulatory role in the adult. Cell Rep 2(1):111-123.

Zhang K, Huang K, Luo Y, Li S. 2014a. Identification and functional analysis of long non-coding RNAs in mouse cleavage stage embryonic development based on single cell transcriptome data. BMC Genomics 15:845.

Zhang L, Lu H, Xin D, Cheng H, Zhou R. 2010. A novel ncRNA gene from mouse chromosome 5 trans-splices with Dmrt1 on chromosome 19. Biochem Biophys Res Commun 400(4):696-700.

Zhang Y, Zhang XO, Chen T, Xiang JF, Yin QF, Xing YH, Zhu S, Yang L, Chen LL. 2013. Circular intronic long noncoding RNAs. Mol Cell 51(6):792-806.

Zhang YC, Liao JY, Li ZY, Yu Y, Zhang JP, Li QF, Qu LH, Shu WS, Chen YQ. 2014b. Genome-wide screening and functional analysis identify a large number of long noncoding RNAs involved in the sexual reproduction of rice. Genome Biol 15(12):512.

Zhao J, Sun BK, Erwin JA, Song JJ, Lee JT. 2008. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. Science 322(5902):750-756.

Zheng G, Dahl JA, Niu Y, Fedorcsak P, Huang CM, Li CJ, Vagbo CB, Shi Y, Wang WL, Song SH, Lu Z, Bosmans RP, Dai Q, Hao YJ, Yang X, Zhao WM, Tong WM, Wang XJ, Bogdan F, Furu K, Fu Y, Jia G, Zhao X, Liu J, Krokan HE, Klungland A, Yang YG, He C. 2013a. ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. Mol Cell 49(1):18-29.

Zheng G, Dahl JA, Niu Y, Fu Y, Klungland A, Yang YG, He C. 2013b. Sprouts of RNA epigenetics: the discovery of mammalian RNA demethylases. RNA Biol 10(6):915-918.

Zheng K, Wang PJ. 2012. Blockade of Pachytene piRNA Biogenesis Reveals a Novel Requirement for Maintaining Post-Meiotic Germline Genome Integrity. PLoS Genet 8(11):e1003038.

Zheng K, Xiol J, Reuter M, Eckardt S, Leu NA, McLaughlin KJ, Stark A, Sachidanandam R, Pillai RS, Wang PJ. 2010. Mouse MOV10L1 associates with Piwi

proteins and is an essential component of the Piwi-interacting RNA (piRNA) pathway. Proc Natl Acad Sci U S A 107(26):11841-11846.

Zhong S, Li H, Bodi Z, Button J, Vespa L, Herzog M, Fray RG. 2008. MTA is an Arabidopsis messenger RNA adenosine methylase and interacts with a homolog of a sex-specific splicing factor. Plant Cell 20(5):1278-1288.

Zou Y, Jiang Z, Yu X, Sun M, Zhang Y, Zuo Q, Zhou J, Yang N, Han P, Ge Z, De W, Sun L. 2013. Upregulation of Long Noncoding RNA SPRY4-IT1 Modulates Proliferation, Migration, Apoptosis, and Network Formation in Trophoblast Cells HTR-8SV/neo. PLoS ONE 8(11):e79598.

Zwart R, Sleutels F, Wutz A, Schinkel AH, Barlow DP. 2001. Bidirectional action of the Igf2r imprint control element on upstream and downstream imprinted genes. Genes Dev 15(18):2361-2366.

## CHAPTER V

Ambatipudi S, Horvath S, Perrier F, Cuenin C, Hernandez-Vargas H, Le Calvez-Kelm F, I identifies accelerated epigenetic ageing associated with postmenopausal breast cancer s Apr;75:299-307. doi: 10.1016/j.ejca.2017.01.014.

Bistner SI, Ford RB, Raffe MR. Hypoglycemia. Kirk & Bistner's Handbook of Veterinary P ed. WB Saunders Co, 2000;144-145.

Boyle P, Clement K, Gu H, *et al*. Gel-free multiplexed reduced representation bisulfite sec profiling. Genome Biology. 2012;13(10):R92. doi:10.1186/gb-2012-13-10-r92.

Cases O, Seif I, Grimsby J, Gaspar P, Chen K, Pournin S, Müller U, Aguet M, Babinet C, altered amounts of brain serotonin and norepinephrine in mice lacking MAOA. Science. 1

Chen D, Wen X, Song YS, *et al*. Associations and prognostic implications of Eastern Coo status and tumoral LINE-1 methylation status in stage III colon cancer patients. Clinical E doi:10.1186/s13148-016-0203-8.

Chen K, Holschneider DP, Wu W, Rebrin I, Shih JC. A spontaneous point mutation produ mice with greatly elevated monoamines and anxiety-like behavior. J Biol Chem. 2004 Sep

Chimonidou M, Kallergi G, Georgoulias V, Welch DR, Lianidou ES. Breast Cancer Metas in primary breast tumors and corresponding Circulating Tumor Cells. Molecular cancer re 2013;11(10):10.1158/1541-7786.MCR-13-0096. doi:10.1158/1541-7786.MCR-13-0096.

Christensen BC, Houseman EA, Marsit CJ, *et al*. Aging and Environmental Exposures Alt Dependent upon CpG Island Context. Schübeler D, ed. PLoS Genetics. 2009;5(8):e1000

Davidson SM, Papagiannakopoulos T, Olenchock BA, *et al*. Environment impacts the me
small cell lung cancer. Cell metabolism. 2016;23(3):517-528. doi:10.1016/j.cmet.2016.01.

Dobson JM. Breed-Predispositions to Cancer in Pedigree Dogs. *ISRN Veterinary Science*
doi:10.1155/2013/941275.

Doi A, Park I-H, Wen B, *et al*. Differential methylation of tissue- and cancer-specific CpG
induced pluripotent stem cells, embryonic stem cells and fibroblasts. *Nature genetics*. 20(

Eo J, Lee HE, Nam GH, Kwon YJ, Choi Y, Choi BH, Huh JW, Kim M, Lee SE, Seo B, Kim
monoamine oxidase A gene expression in the brains of different dog breeds. Gene. 2016
10.1016/j.gene.2016.01.022.

Fariello MI, Boitard S, Naya H, SanCristobal M, Servin B. Detecting Signatures of Selecti
Among Hierarchically Structured Populations. *Genetics*. 2013;193(3):929-941. doi:10.153

Fiebiger I. Hypoglycemia in puppies and young dogs, especially in toy breeds. Tierarztl P

Griffith OL, Montgomery SB, Bernier B, *et al*. ORegAnno: an open-access community-driv
annotation. *Nucleic Acids Research*. 2008;36(Database issue):D107-D113. doi:10.1093/r

Han L, Zhao Z. Contrast features of CpG islands in the promoter and other regions in the
Aug;94(2):117-24. doi: 10.1016/j.ygeno.2009.04.007.

Jeong YH, Lu H, Park CH, Li M, Luo H, Kim JJ, *et al*. Stochastic anomaly of methylome b
disorder of sex development in canine somatic cell nuclear transfer. Sci Rep. 2016 Aug 9

Karyadi DM, Karlins E, Decker B, *et al*. A Copy Number Variant at the KITLG Locus Likely
Carcinoma of the Digit. Horwitz MS, ed. PLoS Genetics. 2013;9(3):e1003409. doi:10.137

Koch IJ, Clark MM, Thompson MJ, *et al*. The concerted impact of domestication and trans
between dogs and gray wolves. Molecular ecology. 2016;25(8):1838-1855. doi:10.1111/m

Kumaki Y, Oda M, and Okano M. QUMA: quantification tool for methylation analysis. Nuc
Server issue), W170–W175.

Lingaas F, Comstock KE, Kirkness EF, Sørensen A, Aarskaug T, Hitte C, Nickerson ML,
Galibert F, Zbar B, Ostrander EA. A mutation in the canine BHD gene is associated with h
cystadenocarcinoma and nodular dermatofibrosis in the German Shepherd dog. Hum Mo

McDermott R, Tingley D, Cowden J, Frazzetto G, Johnson DD. Monoamine oxidase A ge
aggression following provocation. Proc Natl Acad Sci U S A. 2009 Feb 17;106(7):2118-23

Melin M, Rivera P, Arendt M, *et al*. Genome-Wide Analysis Identifies Germ-Line Risk Fac
Tumours. Leeb T, ed. PLoS Genetics. 2016;12(5):e1006029. doi:10.1371/journal.pgen.10

Mila H, Grellet A, Delebarre M, Mariani C, Feugier A, and Chastant-Maillard S. Monitoring neonatal mortality. Preventive Veterinary Medicine 1 August 2017, 143: 11-20.

Nickels S, Truong T, Hein R, *et al*. Evidence of Gene–Environment Interactions between Loci and Established Environmental Risk Factors. Horwitz MS, ed. PLoS Genetics. 2013; doi:10.1371/journal.pgen.1003284.

Ong C-T, Corces VG. Enhancer function: new insights into the regulation of tissue-specifi Genetics. 2011;12(4):283-293. doi:10.1038/nrg2957.

Parker N. Treating neonatal and pediatric hypoglycemia. 2011.

Ponnaluri VKC, Ehrlich KC, Zhang G, *et al*. Association of 5-hydroxymethylation and 5-m specific gene expression. Epigenetics. 2017;12(2):123-138. doi:10.1080/15592294.2016.

Rivera P, Melin M, Biagi T, Fall T, Häggström J, Lindblad-Toh K, von Euler H. Mammary with BRCA1 and BRCA2. Cancer Res. 2009 Nov 15;69(22):8770-4. doi: 10.1158/0008-5

Roos L, van Dongen J, Bell CG, *et al*. Integrative DNA methylome analysis of pan-cancer monozygotic twin-pairs. Clinical Epigenetics. 2016;8:7. doi:10.1186/s13148-016-0172-y.

Schlamp F, van der Made J, Stambler R, Chesebrough L, Boyko AR, Messer PW. Evalua to detect selective sweeps in domestic dogs. *Molecular ecology*. 2016;25(1):342-356. doi

Sloan CA, Chan ET, Davidson JM, *et al*. ENCODE data at the ENCODE portal. Nucleic A issue):D726-D732. doi:10.1093/nar/gkv1160.

Sutter NB, Bustamante CD, Chase K, *et al*. A Single IGF1 Allele Is a Major Determinant c NY). 2007;316(5821):112-115. doi:10.1126/science.1137045.

The ENCODE Project Consortium. An Integrated Encyclopedia of DNA Elements in the H 2012;489(7414):57-74. doi:10.1038/nature11247.

Thomas PD, Campbell MJ, Kejariwal A, *et al*. PANTHER: A Library of Protein Families ar Genome Research. 2003;13(9):2129-2141. doi:10.1101/gr.772403.

Tonomura N, Elvers I, Thomas R, Megquier K, Turner-Maier J, Howald C, *et al*. Genome- Risk Loci Common to Two Malignancies in Golden Retrievers. PLoS Genet 2015 11(2): e https://doi.org/10.1371/journal.pgen.1004922

Ulirsch J, Fan C, Knafl G, *et al*. Vimentin DNA methylation predicts survival in breast can treatment. 2013;137(2):10.1007/s10549-012-2353-5. doi:10.1007/s10549-012-2353-5.

Villar D, Berthelot C, Aldridge S, *et al*. Enhancer Evolution across 20 Mammalian Species

doi:10.1016/j.cell.2015.01.006.

Visel A, Akiyama JA, Shoukry M, Afzal V, Rubin EM, Pennacchio LA. Functional Autonom
Genomics. 2009;93(6):509-513. doi:10.1016/j.ygeno.2009.02.002.

Whiteley MH, Bell JS, Rothman DA. Novel Allelic Variants in the Canine Cyclooxgenase-2
Renal Dysplasia in Dogs. PLoS ONE. 2011 6(2): e16684. https://doi.org/10.1371/journal.p
Whiteley MH. Allelic variation in the canine Cox-2 promoter causes hypermethylation of th
of renal dysplasia. Clin Epigenetics. 2014 Apr 3;6(1):7. doi: 10.1186/1868-7083-6-7.

## CHAPTER VI

AVMA. U.S. Pet Ownership & Demographics Sourcebook, 2012. Retrieved from
https://ebusiness.avma.org/ProductCatalog/product.aspx?ID=454.

Bellumori TP, Famula TR, Bannasch DL, Belanger JM, Oberbauer AM. Prevalence of
inherited disorders among mixed-breed and purebred dogs: 27,254 cases (1995-
2010). J Am Vet Med Assoc. 2013 Jun 1;242(11):1549-55. doi:
10.2460/javma.242.11.1549.

Biffi G, Tannahill D, McCafferty J, Balasubramanian S. Quantitative Visualization of
DNA G-quadruplex Structures in Human Cells. Nature chemistry. 2013;5(3):182-186.
doi:10.1038/nchem.1548.

Blattler A, Farnham PJ. Cross-talk between site-specific transcription factors and DNA
methylation states. Journal of Biological Chemistry. 2013 Nov 29;288(48):34287-94.

Boyko AR, Quignon P, Li L, Schoenebeck JJ, Degenhardt JD, Lohmueller KE, Zhao K,
HG, vonHoldt BM, Cargill M, Auton A, Reynolds A, Elkahloun AG, Castelhano M, Mosh
Johnson GS, Novembre J, Hubisz MJ,

Burns KH, Viveiros MM, Ren Y, Wang P, DeMayo FJ, Frail DE, Eppig JJ, Matzuk MM. F
chromatin and nucleolar organization in oocytes and embryos. Science. 2003 Apr 25;30

Siepel A, Wayne RK, Bustamante CD, Ostrander EA. A simple genetic architecture und
morphological variation in dogs. PLoS Biol. 2010 Aug 10; 8(8):e1000451

Chan TA, Glockner S, Yi JM, *et al*. Convergence of Mutation and Epigenetic
Alterations Identifies Common Genes in Cancer That Predict for Poor Prognosis .
Gazdar AF, ed. PLoS Medicine. 2008;5(5):e114. doi:10.1371/journal.pmed.0050114.

Cogoi S and Xodo LE. G-quadruplex formation within the promoter of the KRAS
proto-oncogene and its effect on transcription. Nucleic Acids Res. 2006; 34(9): 2536–
2549. doi:10.1093/nar/gkl286

Cronin MA, Cánovas A, Bannasch DL, Oberbauer AM, Medrano JF. Single nucleotide polymorphism (SNP) variation of wolves (Canis lupus) in Southeast Alaska and comparison with wolves, dogs, and coyotes in North America. J Hered. 2015 Jan-Feb;106(1):26-36. doi: 10.1093/jhered/esu075.

David AP, Margarit E, Domizi P, Banchio C, Armas P, Calcaterra NB. G-quadruplexes as novel cis-elements controlling transcription during embryonic development. Nucleic Acids Research. 2016;44(9):4163-4173. doi:10.1093/nar/gkw011.

De La Fuente R, Eppig JJ. Transcriptional activity of the mouse oocyte genome: companion granulosa cells modulate transcription and chromatin remodeling. Developmental biology. 2001 Jan 1;229(1):224-36.

Dipankar S and Gilbert W. Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. Nature 28 July 1988, 334, 364 - 366; doi:10.1038/334364a0

Furumatsu T, Tsuda M, Yoshida K, Taniguchi N, Ito T, Hashimoto M, Ito T, Asahara H. Sox9 and p300 cooperatively regulate chromatin-mediated transcription. Journal of Biological Chemistry. 2005 Oct 21;280(42):35203-8.

Fukuhara M, Ma Y, Nagasawa K, Toyoshima F. A G-quadruplex structure at the 5′ end of the H19 coding region regulates H19 transcription. Scientific Reports. 2017;7:45815. doi:10.1038/srep45815.

Giraldo R, Suzuki M, Chapman L, Rhodes D. Promotion of parallel DNA quadruplexes by a yeast telomere binding protein: a circular dichroism study. Proceedings of the National Academy of Sciences of the United States of America. 1994;91(16):7658-7662.

Gu H, Smith ZD, Bock C, Boyle P, Gnirke A, Meissner A. Preparation of reduced representation bisulfite sequencing libraries for genome-scale DNA methylation profiling. Nat Protoc. 2011 Apr;6(4):468-81. doi: 10.1038/nprot.2010.190.

Halder R, Halder K, Sharma P, Garg G, Sengupta S, Chowdhury S. Guanine quadruplex DNA structure restricts methylation of CpG dinucleotides genome-wide. Mol Biosyst. 2010 Dec;6(12):2439-47. doi: 10.1039/c0mb00009d.

Hänsel-Hertsch R, Beraldi D, Lensing SV, Marsico G, Zyner K, Parry A, Di Antonio M, Pike J, Kimura H, Narita M, Tannahill D,
Balasubramanian S. G-quadruplex structures mark human regulatory chromatin. Nat Genet. 2016 Oct;48(10):1267-72. doi: 10.1038/ng.3662.

Huson HJ, vonHoldt BM, Rimbault M, *et al*. Breed-Specific Ancestry Studies and

Genome-Wide Association Analysis Highlight an Association Between the MYH9 Gene and Heat Tolerance in Alaskan Sprint Racing Sled Dogs. Mammalian Genome. 2012; 23(1-2):178-194. doi:10.1007/s00335-011-9374-y.

Koch IJ, Clark MM, Thompson MJ, *et al*. The concerted impact of domestication and transposon insertions on methylation patterns between dogs and gray wolves. Molecular ecology. 2016;25(8):1838-1855. doi:10.1111/mec.13480.
Lindblad-Toh K, Wade CM, Mikkelsen TS, Karlsson EK, Jaffe DB, Kamal M, *et al*. Genome sequence, comparative analysis and haplotype structure of the domestic dog. Nature. 2005 Dec 8;438(7069):803-19.

Moye AL, Porter KC, Cohen SB, *et al*. Telomeric G-quadruplexes are a substrate and site of localization for human telomerase. Nature Communications. 2015; 6:7643. doi:10.1038/ncomms8643.

Oertel BG, Doehring A, Roskam B, Kettner M, Hackmann N, Ferreirós N, Schmidt PH, Lötsch J. Genetic-epigenetic interaction modulates μ-opioid receptor regulation. Hum Mol Genet. 2012 Nov 1;21(21):4751-60. doi: 10.1093/hmg/dds314.

Parker HG, Dreger DL, Rimbault M, Davis BW, Mullen AB, Carpintero-Ramirez G, Ostrander EA. Genomic Analyses Reveal the Influence of Geographic Origin, Migration, and Hybridization on Modern Dog Breed Development. Cell Rep. 2017 Apr 25;19(4):697-708. doi: 10.1016/j.celrep.2017.03.079.

Rankin S, Reszka AP, Huppert J, *et al*. Putative DNA Quadruplex Formation within the Human c-kit Oncogene. Journal of the American Chemical Society. 2005;127(30):10584-10589. doi:10.1021/ja050823u.

Riou JF, Guittat L, Mailliet P, *et al*. Cell senescence and telomere shortening induced by a new series of specific G-quadruplex DNA ligands. Proceedings of the National Academy of Sciences of the United States of America. 2002;99(5):2672-2677. doi:10.1073/pnas.052698099.

Sekido R, Lovell-Badge R. Sex determination involves synergistic action of SRY and SF1 on a specific Sox9 enhancer. Nature. 2008 Jun 12;453(7197):930.

Siddiqui-Jain A, Grand CL, Bearss DJ, Hurley LH. Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. Proc Natl Acad Sci U S A. 2002 Sep 3;99(18):11593-8.

Stefan M, Wei C, Lombardi A, Li CW, Concepcion ES, Inabnet WB 3rd, Owen R, Zhang W, Tomer Y. Genetic-epigenetic dysregulation of thymic TSH receptor gene expression triggers thyroid autoimmunity. Proc Natl Acad Sci U S A. 2014 Aug 26;111(34):12562-7. doi: 10.1073/pnas.1408821111.

Thompson MJ, vonHoldt B, Horvath S, Pellegrini M. An epigenetic aging clock for dogs and wolves. Aging (Albany NY). 2017;9(3):1055-1068. doi:10.18632/aging.101211.

Vonholdt BM, Pollinger JP, Lohmueller KE, Han E, Parker HG, Quignon P, *et al*. Genome-wide SNP and haplotype analyses reveal a rich history underlying dog domestication. Nature. 2010 Apr 8;464(7290):898-902. doi: 10.1038/nature08837.

Zhao M, Shirley CR, Hayashi S, Marcon L, Mohapatra B, Suganuma R, Behringer RR, Boissonneault G, Yanagimachi R, Meistrich ML. Transition nuclear proteins are required for normal chromatin condensation and functional sperm development. Genesis. 2004 Apr 1;38(4):200-13.

Zhou G, Liu X, Li Y, *et al*. Telomere targeting with a novel G-quadruplex-interactive ligand BRACO-19 induces T-loop disassembly and telomerase displacement in human glioblastoma cells. Oncotarget. 2016;7(12):14925-14939. doi:10.18632/oncotarget.7483.