

Sperm histone methylation is implicated in paternal epigenetic inheritance

By
Keith Siklenka

Department of Pharmacology and Therapeutics

McGill University
Montreal, Quebec, Canada

November 2017

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree
of Doctor of Philosophy

Table of Contents

Abstract	4
Résumé.....	5
Acknowledgements.....	6
Preface and Contribution of Authors	7
Chapter 1: General Introduction.....	9
 1.1 Background.....	9
 1.2 The plasticity of development and its adaptability to the environment.....	9
 1.3 Rational for the Thesis	13
 1.4 An Introduction to The Epigenome	14
1.4.1 Background	14
1.4.2 The Nucleosome.....	15
1.4.3 DNA Methylation.....	21
1.4.4 RNA and its Subclasses	24
 1.5 Epigenetic Features of the Germline.....	27
1.5.1 Background	27
1.5.2 Germline Establishment and Epigenetic Reprogramming.....	27
1.5.3 Chromatin Remodeling in Spermatogenesis and Spermiogenesis.....	33
1.5.4 RNA Content in Mature Sperm.....	38
 1.6 Epigenetic Mechanisms of Inheritance	40
1.6.1 Background	40
1.6.2 Inheritance through DNA methylation	41
1.6.3 Inheritance Through RNA	44
1.6.4 Inheritance through Histone Modifications	46
 1.7 References.....	48
Chapter 2: Disruption of histone methylation in developing sperm impairs offspring health transgenerationally	73
 2.1 Abstract.....	74
 2.2 Introduction.....	74
 2.3 Results	77
 2.4 Discussion.....	87
 2.5 Conclusion.....	90
 2.6 Materials and Methods.....	90
 2.6 References	99
 2.7 Figure Legends	105
 2.8 Appendix I	119
 2.9 Connecting Text.....	139

Chapter 3: Sperm histone H3K4me3 is implicated in paternal epigenetic inheritance.	140
3.1 Abstract.....	140
3.2 Introduction.....	140
3.3 Results.....	143
3.4 Discussion.....	147
3.5 Methods.....	151
3.6 References	158
3.7 Figure Legends	164
3.8 Appendix II.....	174
3.9 Supplemental References	184
Chapter 4: General Discussion	185
4.1 Transgenerational Effects from KDM1A Overexpression in Mouse Testis	185
4.1.1 Male role on offspring survivability.....	185
4.2 Mechanisms Mediating Epigenetic Inheritance of Paternal Effects.....	187
4.2.1 The role of Transgenic KDM1A	187
4.2.2 Modes of intergenerational Inheritance	191
4.3 Regulatory Mechanisms for Epigenetic Information in Sperm	193
4.4 Next Generation Sequencing Techniques	197
4.4.1 Chromatin Immunoprecipitation Sequencing	197
4.4.2 Sample normalization and differential methylation analysis	199
4.5 Considerations of the Model	202
4.5.1 A True Example of Epigenetic Inheritance.....	202
4.5.2 Variability of Phenotype.....	202
4.5.3 Technical Limitations	204
4.6 Conclusions	206
4.7 References	207
END.....	212

Abstract

Paternal epigenetic inheritance has been described in human cohorts and animal models ranging from worm to mouse. Chemical or nutritional challenge during key developmental time points have been correlated with physiological change in future offspring that oftentimes span multiple generations. Despite these observations, the underlying molecular mechanisms of non-genetic (epigenetic) inheritance remain unknown. Recent studies have shown that sperm histones are retained at genomic regions high in CpG. Specifically, activating histone marks, such as histone H3 lysine 4 di-methylation (H3K4me2) and H3 lysine 4 tri-methylation (H3K4me3), are retained at promoters of genes with functional roles in embryonic development. Therefore, we hypothesize that the methylation of sperm histones is important to the health and development of future offspring and is involved in paternal epigenetic inheritance. To test this hypothesis, we designed a transgenic inbred mouse model that overexpressed the histone demethylase KDM1A in mouse testes with the goal of disrupting the sperm epigenome. This thesis shows that specific alterations of the sperm histones H3K4me2 and H3K4me3 in transgenic (TG) sires were related to severe abnormal phenotypes and gene expression changes in TG offspring and 2-cell embryos. Strikingly, abnormal embryonic development was inherited transgenerationally for two subsequent generations. Furthermore, we quantified consistent intergenerational differences in sperm H3K4me3 and RNA content, which for the first time, implicate mammalian sperm histone methylation in the mechanisms of paternal epigenetic inheritance.

Résumé

Le phénomène d'héritage épigénétique venant du père a été identifié lors de plusieurs études épidémiologiques chez l'humain ainsi que dans de nombreux modèles animaux allant des nématodes aux rongeurs. De plus, l'apport de certaines molécules ou nutriments à des moments clés du développement chez le père a été corrélé à des changements physiologiques dans la progéniture pouvant perdurer pendant de multiples générations. Malgré ces observations, les mécanismes de cet héritage non-génétique (épigénétique) demeurent inconnus. Des études récentes ont montré que des histones sont conservés dans le sperme au niveau des régions du génome riches en CpG. Plus spécifiquement, les modifications des histones activatrices de la transcription, telles que la di-méthylation de l'histone H3 à la lysine 4 (H3K4me2), sont présentes au niveau des promoteurs des gènes impliqués dans le développement embryonnaire. De ce fait, nous postulons l'hypothèse que la méthylation des histones spermatiques joue un rôle dans la santé et le développement de la progéniture et est impliquée dans le phénomène d'héritage épigénétique paternel. Afin de tester cette hypothèse, nous avons créé un modèle de souris transgéniques qui sur expriment la déméthylase d'histone KDM1A dans les testicules de souris dans le but d'altérer l'épigénome spermatique. Nous avons observé des altérations spécifiques de la di et tri méthylation de l'histone H3 au niveau de la lysine 4 (H3K4me2 et H3K4me3, respectivement) chez les animaux transgéniques, ainsi que de leurs ARNs spermatiques, une modification de l'expression génique dans leur embryons au stade 2 cellules mais pas de changement au niveau de la méthylation de l'ADN. Les souriceaux nés de pères transgéniques présentaient de nombreuses anomalies développementales au stade E18.5 et une survie réduite après la naissance. Ces phénotypes sont hérités par les deux générations suivantes,

ce qui démontre pour la première fois le rôle des histones présents dans le sperme dans l'héritage épigénétique venant du père.

Acknowledgements

I would like to thank my supervisor Dr. Sarah Kimmins for supporting my development and challenging me throughout my graduate career. You always gave me an opportunity to shine. I'm grateful for your consistent effort to provide the tools, resources, and connections that helped make this exciting project happen.

I was very fortunate to have additional mentorship from inspiring, and brilliant minds at McGill. Thank you to my thesis advisory committee, Dr. Dan Bernard, Dr. Jacquetta Trasler, and Dr. Jason Tanny for the thoughtful questions, insightful discussions and excellent support. I want to especially thank Dr. Dan Bernard for his guidance over the years. His passion for science and his commitment to the trainees has been inspiring. I also want to acknowledge and thank Dr. Jeff Xia for encouraging me to learn bioinformatics, and for supporting me in this arduous process. He has always believed in my potential and pushed me to keep going forward. A huge thank you to Dr. Vanessa Dumeaux who came to the Kimmins lab near the end of my program, but in just a few months had a profound effect on the way I approach computational biology.

I want to thank the RQR, CRRD, SSR, and ASA for awarding me funding for research and travel, and for giving me a platform to present my thesis work.

Thank you to everyone in the Kimmins lab for making each day more enjoyable. A special thank you to Dr. Romain Lambrot for always being willing to help discuss ideas, troubleshoot protocols, or lend a hand. And to Christine Lafleur for her support, expertise, and extra hands whenever it was needed.

Thank you to my mom, sister and friends who I left behind in beautiful British Columbia all these years. Thank you for your support, care packages, late night calls, messages, old fashioned snail mails and your occasional visits – although you could have visited more often. Also to the new family I made here. Judith and Nick, my “fake” parents and to Marion my rock. I couldn't have done this without you and your belief in me.

Preface and Contribution of Authors

This thesis is in manuscript format which conforms to the “Guidelines for Thesis Preparation” of the Faculty of Graduate Studies and Research at McGill University. The thesis consists of four chapters, two of which are in the form they were submitted for publication. Chapter 1 is a general introduction which provides the necessary background for the remainder of the thesis. Sections of the general introduction will be submitted as a review to the “Encyclopedia of Reproduction”. Chapter 2 is a manuscript published in the journal “Science”. Chapter 3 is a manuscript in preparation for submission to the journal “Nature Communications”. Appendices to these chapters contain the supplemental results which accompany the main figures and text. The connecting text between chapter 2 and 3 is to ensure continuity of the thesis. Chapter 4 contains an expanded discussion on the results and a commentary on the limitations and perspectives.

The work presented within this thesis was a collaborative effort with multiple co-authors. Below is a description of each authors’ contribution to the manuscripts within this thesis:

Chapter 2:

Disruption of histone methylation in developing sperm impairs offspring health transgenerationally

Keith Siklenka, Serap Erkek, Maren Godmann, Romain Lambrot, Serge McGraw, Christine Lafleur, Tamara Cohen, Jianguo Xia, Matthew Suderman, Michael Hallett, Jacquette Trasler, Antoine H. F. M. Peters, and Sarah Kimmins.

Science 2015; 350(6261), aab2006. doi:10.1126/science.aab2006

Sarah Kimmins and Antoine Peters directed the writing of the main text. I wrote the draft of the methods and figure legends and collaborated with Dr Kimmins and Dr Peters for the final

version. Maren Godmann created the mouse model and did preliminary characterization of the phenotype with Christine Lafleur (Figure 2.2, 2.3b-k). I was responsible for E18.5 data collection of the nonTG²⁻³ generations (Figure 2.1D) and analysis of all data associated with Figure 2.2 and SI Figure 2.4. I assisted in the development of the E18.5 skeletal analysis protocol in collaboration with Romain Lambrot and Tamara Cohen; images and results in Figure 2.4 and SI Figure 2.5 were generated by Tamara Cohen. The ChIP-sequencing results in Figure 2.5 was generated by Serap Erkek and Antoine Peters at FMI in Basel Switzerland; I was involved in data analysis and interpretation. I was responsible for generating microarrays to assess 2-cell gene expression and RNA sperm content and analyzed the results in collaboration with Jeff Xia (Figure 2.6). DNA methylation with RRBS was done by Serge McGraw and Jacquette Trasler and analyzed by Serge McGraw, Michael Hallet, Matt Sudermann, and Michael Stadler; I performed Epityper and Pyrosequencing to validate DNA methylation results (SI Figure 8).

Chapter 3

Sperm histone H3K4me3 is implicated in paternal epigenetic inheritance

Keith Siklenka, Christine Lafleur, Vanessa Dumeaux, Sarah Kimmins

Manuscript in preparation

I was responsible for performing all experiments and bioinformatics analysis. I prepared all figures, with guidance from Vanessa Dumeaux for Figure 3.2C and Figure 3.5B. I wrote the manuscript with guidance from Sarah Kimmins. Christine Lafleur assisted with colony maintenance and sample collection. Vanessa Dumeaux contributed bioinformatic expertise to complement my analysis.

Chapter 1: General Introduction

1.1 Background

Inheritance of acquired characteristics was first proposed by Jean-Baptiste Lamarck in 1801. Although his theories were overlooked compared to those of Charles Darwin's Theory of Evolution, Gregor Mendel's discovery of genetic inheritance, and August Weismann's germplasm barrier, some aspects of Lamarckian thinking have seen a resurgence since the 21st century. Major advancements in the fields of medicine and genetics have enabled comparative, genome-wide, approaches to track evolution, genetic variation, and mechanisms of disease across species (1-3). But rates of complex disease such as obesity, diabetes, cardiovascular disease, asthma, and ADHD are rising faster than can be explained by genetic contribution alone (4-6). This "missing heritability" is well demonstrated by the divergent phenotypes of monozygotic twins despite sharing identical DNA. These phenomena highlight the need to consider non-genetic mechanisms of inheritance as potential drivers of phenotypic variation. This chapter will present the current understanding of non-genetic inheritance with an emphasis on the male germ-line. It will outline evidence for multigenerational epigenetic inheritance in human and animal models and describe how epigenetic programming of mammalian sperm involves DNA methylation, RNA, and histone modifications.

1.2 The plasticity of development and its adaptability to the environment

Evidence that the developing embryo is susceptible to environmental cues stems from both experimental research and clinical observation conducted in the late 1950s (Reviewed in 7, 8, 9). The prescription compounds thalidomide and diethylstilbestrol (DES) remain the best-

known examples of an *in-utero* exposure resulting in congenital malformations, miscarriage, and death of the developing fetus with little reported effect on the mother (10-13). Epidemiologists David Barker *et al.*, expanded the idea of *in-utero* vulnerability in the late 1980s to develop the fetal origins of disease hypothesis (14). By comparing historical birth records with rates of adult cardiovascular disease, he found a strong anti-correlation between birth weight and cardiovascular disease, hypertension, diabetes mellitus (14-19). The inverse-relationship was strengthened when the population was stratified by geographical region and socio-economic status (14). These findings challenged the dogma of genetic predisposition to adult disease by supporting a model of maternal contribution via *in-utero* exposure.

A classical test of the fetal origins hypothesis came from human cohorts of the 1944-45 Dutch Hunger Winter (Reviewed by 20, 21-25). Near the end of World War II the Germans occupied the western-region of the Netherlands and blocked food supplies during what was an unusually cold winter. The rapid depletion of food stocks created a sudden famine across the region that impacted all social classes. At peak famine, the pregnant women and children were rationed less than 800 kcal / day (26). Upon liberation of the blockade in the following spring, the region's food supplies were rapidly returned. As a result, the population represents a human model of maternal-under nutrition with a well-defined window of exposure. Observation of the offspring born from mothers exposed to famine at any point during gestation related this maternal-effect with higher levels of adult onset glucose intolerance (14). Barker and others suggested that maternal under-nutrition induced an adaptive response that was predictive of an environment with limited food supply. These adaptations would then manifest as metabolic changes in the offspring (14, 27). If that environmental stress was later removed, as was the case with the cohort's sudden return to an abundance of food, the metabolic adaption could become

detrimental. The timing of starvation was shown to impact the effect of programming the onset of disease. Complete documentation of birth records allowed for many prenatal exposures to be grouped by trimester (Reviewed in 21). Consequently, under-nutrition at early gestation resulted in offspring who developed obesity and cardiovascular disease. Babies born with exposure in late development had reduced birth weight but lower rates of chronic disease.

Barker's hypotheses and the subsequent epidemiological studies it inspired have since been expanded to reflect the broad range of events that may contribute to the Developmental Origins of Health and Disease (DOHaD). For example, under- or over-nutrition, chemical exposure to toxicant and endocrine disruptors, stress, infections, and even changes to the microbiome have now been associated with the inheritance of chronic disease (7, 28-39). Manifestations of parental effects in generations immediately following an exposure are classified as intergenerational (Figure 1).

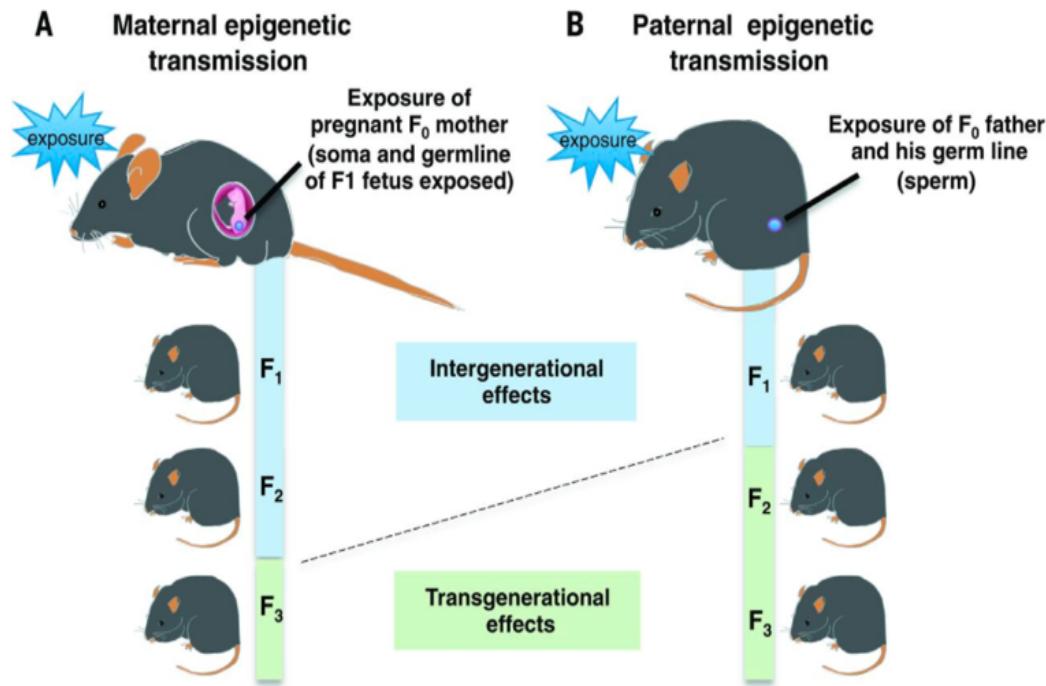


Figure 1.1 – Schematic outlining the definition of multigenerational effects in maternal and paternal lineages

In maternal models, like those described above, an intergenerational effect would span two familial generations (F1–F2) because the under-nutrition during gestation would have directly affected the embryo (F1) and its developing germline (F2). Parental effects observed in generations beyond any direct exposure are classified as transgenerational ($\geq F3$).

Intergenerational effects that arise through maternal exposure have become well established in the modern literature and are primarily focused on the mothers' contributions to the embryo prior to conception and through direct effects *in utero*. However, little is known about the role of the father despite contributing half of the genetic, and likely non-genetic, information to the next of kin. Any influence from father to the next generation must be transmitted through the sperm. But early research and public opinion have portrayed the sperm as a mere carrier of the haploid genome, with little ability to impart environmental effects on the developing zygote. Moreover, the absence of robust studies makes it difficult to assess the extent of paternal transmission. For example, reports of a correlation between the father's occupational exposure to chemicals and offspring health include a range of adverse outcomes such as preterm birth, congenital malformations, childhood leukemia, and death (40-46). But follow up studies often have trouble reproducing the observed effects (35, 46). As a result, there is a need for modern, reproducible studies able to describe paternal effects across generations.

The best evidence to show paternal transgenerational inheritance in humans remains from epidemiological studies of multiple cohorts from the Overkalix region of Sweden (47, 48). Historical records were kept from 1890 to 1995 and included the variation of food price and harvest statistics. Grandfathers who had high food availability during their pre-pubertal years had grandsons with increased rates of mortality by diabetes (47, 49). On the other hand, if food availability was poor during these years, grandsons had low probability of developing

cardiovascular disease. This sex specific response was also observed in the maternal lineage (47). More recently, modern cohorts have been designed specifically to test the paternal epigenetic inheritance hypothesis established in Overkalix (50, 51) (47, 48, 50, 52-54). For example, The Avon Longitudinal Study of Parents and Children (ALSPAC) has described the paternal transmission of childhood obesity to the offspring of fathers who smoked during mid-childhood (52). The obesity phenotype was greatest in offspring of fathers who started smoking before the onset of puberty (age < 11), and thus remains consistent with the idea of a developmentally sensitive window of exposure. However, these windows may not be limited to the slow-growth period. Tawianeese fathers who chewed the tobacco-like betel quid over the age of 18 conceived sons who developed early metabolic syndrome in an exposure dose and duration dependent manner (54, 55). Moreover, additional paternal-lifestyle related factors have been linked to paternal effects. The recent Framingham heart study showed that men who experienced early-onset obesity father children with higher serum alanine aminotransferase (ALT), a common marker for liver injury (56, 57).

1.3 Rational for the Thesis

These aforementioned epidemiological studies have expanded our understanding of parental effects and epigenetic inheritance. **We hypothesize that the mammalian sperm epigenome serves a direct role in offspring health and development and mediates the transmission of paternal effects across multiple generations.** The recognition of the father's contribution to offspring health has important social and biological implications; however, our current understanding is limited by the paucity of studies addressing mechanistic questions. Therefore, this thesis will attempt to close the gap in knowledge by functionally testing the effects of an altered sperm epigenome on offspring health. The manuscripts presented within will

examine how the epigenetic components of the sperm, such as histone methylation, DNA methylation, and RNA content, function together to regulate the health and development of the offspring. We will also attempt to identify the mechanisms involved in paternal transgenerational epigenetic inheritance. We find that sperm histone methylation is implicated in these processes.

1.4 An Introduction to The Epigenome

1.4.1 Background

The epigenome is made up of elements that modulate gene expression without changing the underlying DNA sequence. A true epigenetic factor must be stably maintained across cell divisions, and functions to pass on a memory of the parental state. The major epigenetic components include DNA methylation, histone modifications, and non-coding RNA. But the epigenome also involves the coordination of specific enzymes that control the spatial organization of chromosomes and the temporal establishment, maintenance, or removal of epigenetic marks. For the cell, this results in a molecular identity that directs fate-determination, differentiation, and biological function. The proper programming of cellular identity is influenced by environmental signals and allows for some phenotype plasticity throughout life. This flexibility may allow for rapid adaptation to changing cellular conditions. Moreover, acute environmental insults may induce changes to the cell's epigenetic landscape and prompt an adaptive response that's inherited by the daughter cells. If epigenetic alterations are established in the germline, parent of origin characteristics may be inherited by the offspring. Thus, the epigenome may contribute to phenotypic variation and heritability of health and disease in the next generation.

1.4.2 The Nucleosome

In eukaryotic cells, 147 bp of genomic DNA wraps around a protein complex known as the nucleosome (58, 59). The nucleosome is an octamer composed of two copies of core-histone protein-dimers: (H2A-H2B)₂ and (H3-H4)₂. The DNA-nucleosome complex forms repetitive, filamentous, arrays called chromatin that encompass the whole genome (58, 60). In the nucleus, chromatin functions to structure DNA into chromosomes through higher-order organization. At the gene level, chromatin remains a highly dynamic regulator of gene transcription through coordinated control of DNA accessibility for RNA Polymerase complex (61). The degree of compaction is influenced by the addition of post-translational modifications to the core histones. Unstructured N-terminal tails of each histone monomer extend outward from the DNA-wrapped globular core domain and influence the bio-physical interactions between DNA and nucleosome (59). The localized charge of the histone tail is altered when specific amino-acids are subject to post-translational modifications (PTM) such as acetylation, methylation, and phosphorylation (62, 63). In the context of these examples, chromatin undergoes expansion due to opposing charge between the histone tail and DNA backbone (64-66). This change in conformation generally results in the formation of euchromatin, an open and accessible chromatin state, that is most often associated with active promoters or intergenic regulatory elements such as enhancers (67). In contrast, specific variations of these histone modifications may instead induce chromatin condensation and restrict DNA accessibility and transcription at these sites. This repressive conformation is known as constitutive or facultative heterochromatin. Constitutive heterochromatin is static and provides stability to gene poor regions such as centromeres, telomeres, and repeats. Facultative heterochromatin, on the other hand, may exist at promoters or other responsive genomic elements that must be kept silent during specific points in

development. Additional posttranslational modifications such as the bulky sumoylation and ubiquitination moieties may also affect chromatin structure (68-70). With modern mass-spectrometry techniques, novel substrates have been appended to the classical list, including propionylation, crotonylation, and ADP-ribosylation, among others, but very little is known about their function (71-74).

One of the best studied examples of histone PTMs is the methylation of histone 3 at lysine 4 (H3K4me). The nitrogen on a lysine side-chain can be mono- di- or tri- methylated, each adding a different layer of nuanced functions. For example, di- and tri-methylation of H3K4 (H3K4me2, and H3K4me3) are most often associated with active gene transcription and found in a narrow distribution around transcriptional start sites (TSS) or gene promoters (75-78). While mono-methylation of H3K4 (H3K4me1) can localize to promoters, it is frequently associated with distal regulatory elements such as active or poised enhancers (79, 80). In contrast, methylation of alternative lysine residues on H3 can have an opposing effect. Tri-methylation of H3 lysine 9 (H3K9me3) and lysine 27 (H3K27me3) are most often associated with gene repression and heterochromatin formation (78, 81-83).

However, the binary classification of PTMs into an active or repressive state is difficult to apply on a general scale due to the added complexity of PTM combinations, distribution of modification, symmetry of modification between H3 tails, higher-order chromatin conformation, and chromatin-associated regulatory factors (84-86). Histone modifications such as H3K4me3 and H3K9me3 have been associated with both active and repressed genes that depends heavily on the surrounding biological context (87-89). A well-studied example of combinatorial effects of histone PTMs effects are promoters marked with both “active” H3K4me3 and “repressive” H3K27me3. This class of modification is defined as “bivalent” and is commonly found at

promoters of developmental genes in embryonic stem cells (ESC), developing embryos, and is retained throughout male germline development (90-96). Histone bivalency maintains a repressed transcriptional state while poising the gene for rapid activation at the appropriate time in development. A poised state is not limited to promoter regions and can be detected at enhancers containing H3K4me1, H3K4me2 and H3K27me3, with the exclusion of H3K27ac (97-100).

The strong correlations between biological processes and the specific class of histone modification have driven hypotheses for protein classes capable of recognizing these motifs and interpreting an encoded regulatory signal (101-107). The discovery of specialized protein domains tailored for the binding of acetylation, methylation and phosphorylation moieties suggests that histone modifications may serve as a scaffold by which regulatory proteins could bind (104-107). These theories have become established as the histone-code hypothesis (104). Advancements in chromatin immunoprecipitation and genome-wide sequencing have since helped to decipher the regulatory qualities of histone modifications by identifying their genomic locations, distribution, and profile in hundreds of cell types (67, 78, 91, 108)

Setting and interpretation of the histone code is carried out by a continuously growing list of enzymes, and chromatin binding proteins and complexes (109-111). The first descriptions of such structures were done in yeast with the discovery of COMPASS or Complex Of Proteins Associated with Set1 (112-115). The Set1 histone methyltransferase subunit of COMPASS shares a common motif with the trithorax (Tcx) gene, and has been described in *Drosophila* to cause patterning transformations in abdominal segments and wings when mutated (116). In yeast, a single Set1/COMPASS complex is sufficient for the generation of all lysine 4 methyl-states. However, plants and animals have evolved additional subtypes of Set1/COMPASS-like

protein families, each with subtle differences in their function (115). This diversity is illustrated by the 6 subtypes of the mammalian Mixed Lineage Leukemia (MLL) methyltransferase family (117-119). The MLL-1 subtype gains methyltransferase function when complexed with core COMPASS subunits WDR5, Ash2 and RBBP5 (119-121). Specifically, the WD40-domain in the WDR5 subunit recognizes H3K4me2 and directs MLL1 methyltransferase to add H3K4me3. In contrast, other MLL family members, such as MLL3, mediate the establishment of mono and di-methylation states. These unique differences between MLL family members highlights the abundant complexity inherent to the establishment of the cell's epigenetic landscape.

The polycomb group (PcG) is responsible for the establishment of a repressive transcriptional state at hundreds of developmentally relevant target genes, but are best known for their role in the heritable silencing at HOX gene cluster (122-124). In mammals, PcG contains two main families of protein complexes, Polycomb Repressive Complex 1 (PRC1) and PRC2 (125, 126). As is true with TcX family members, each PcG family contains a functionally diverse set of subtypes further complicating the complete mechanistic understanding of gene targeting, recruitment, and regulation of transcription (Reviewed in 127). Nevertheless, PcG is generally associated with transcriptional silencing via addition of H3K27me3 at GC rich regions, such as promoters (122, 123, 128). The mammalian PRC2 complex is made up of four core subunits, EZH1/2, EED, SUZ12, and RBBP7/4, but will often include multiple additional subunits as well as transient interactions with other epigenetic modifying proteins (Reviewed in 127). The EED subunit confers substrate recognition through its WD repeat domain. A caging mechanism binds PRC2 to H3K27me3 and stimulates the methyltransferase activity of the catalytic EZH1/2. PRC2 then copies H3K27me3 onto the surrounding, naïve, H3-tails (129-131). This positive feed-forward loop facilitates the maintenance of a repressed state throughout cell divisions. However,

PRC2 recruitment is negatively influenced by histone methylation such as H3K4me3 and H3K36me3, providing evidence for chromatin-context dependent restriction of histone methylation boundaries (132, 133).

Functional characterization of histone methyltransferases has identified important roles in development and differentiation programmes of cells across many model systems. For example, Mll1 null mutations result in embryonic lethality in mouse models, but heterozygous Mll1 knockouts display skeletal transformations in cervical, thoracic and lumbar vertebrae that mimic the phenotypes observed for HOX gene deletions (117, 134). Furthermore, Mll1 mice with a mutation to the Set1 domain show skeletal abnormalities and low levels of H3K4me2/3 at HOX gene promoters (135). The Suvv39h1/2 histone methyltransferases regulates H3K9 methylation and heterochromatin formation. Their disruption results in chromosome instability, spermatogenesis failure, increased tumorigenesis, and long telomeres (136, 137). Deficiency of the PRC2 subunits Ezh2, Eed and Suz12 is embryonic lethal at early stages of mouse development (138-140).

The covalent modification of a histone tail by methylation was initially described as an irreversible event. However, the discovery of Jumonjie C-domain containing dioxygenases (141, 142), and amine oxidases, such as Lysine Specific Histone Demethylase (LSD1, now known as KDM1A) (143) was the first step to connecting lysine methylation as a true, reversible, epigenetic modification. KDM1A catalyzes an FAD-dependent oxidative reaction to remove H3K4me1 and H3K4me2, but not H3K4me3 (143, 144) (Figure 2). The N-terminal protein contains a SWIRM domain that provides structural stability and plays a role in the interaction with histone-tails (145). KDM1As active site binds a 21 amino acid sequence on the H3 N-terminal tail via electrostatic interactions. However, the lack of a DNA binding motif on the

protein surface does not provide KDM1A the ability to target nucleosomal histones assembled into a chromatin fiber (146-148). As a result, KDM1A relies on the binding ability of its many known protein complexes (147, 148). A large tower domain extending from its catalytic center provides a multi-docking site which facilitates a diverse range of protein-interactions that define, in part, KDM1As regulatory function in the cell (149-152). The best described interaction partner of KDM1A is the co-repressor complex, CoREST. CoREST is required for KDM1A demethylase activity of H3K4me1 and 2 and thus they function together to repress transcription at targeted sites (148, 153). However, if KDM1A is associated with the androgen receptor (AR) it can switch demethylation activity towards the repressive methyl marks of H3K9me1 and 2 to facilitate transcription at AR-dependent targets (154). In addition to histone demethylation, KDM1A can target methyl groups on non-histone proteins. For instance, KDM1a targeted demethylation of the DNA methyltransferase DNMT1 improves its stability, in ES cells and has proven to be essential in the maintenance of global DNA methylation (155).

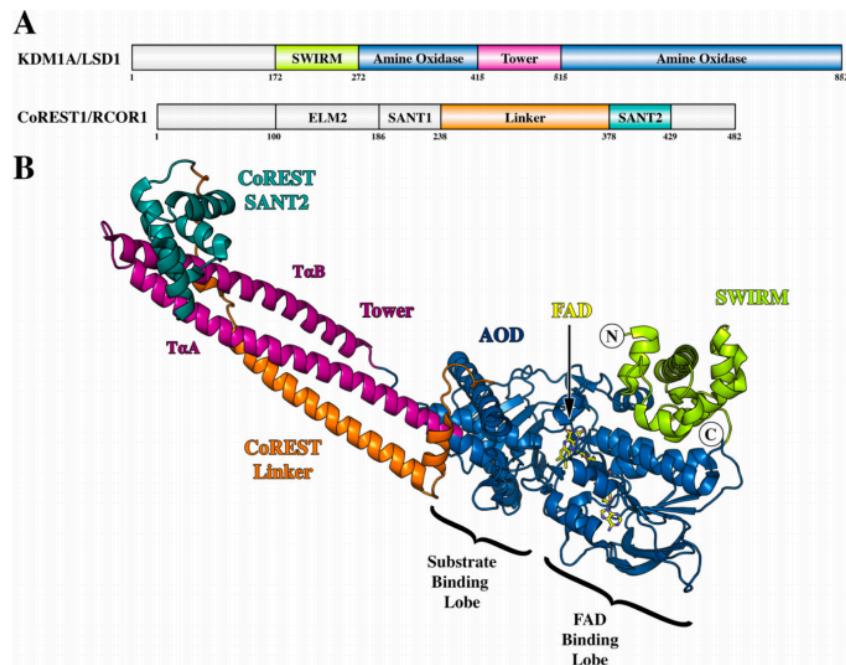


Figure 1.2 KDM1A-CoREST protein complex domain and structure overview (144)

But histone modifying enzymes need not rely on their enzymatic activity alone to assert their function. For example, the histone H3K27me3 demethylases, Jmjd3 and UTX, interact with a master regulator of T-cell differentiation to control gene transcription independently of their demethylase activity (156, 157). Similarly, a major PcG member, PRC1, does not contain enzymatic activity for H3K27me3, but functions to silence specific PcG target genes through multiple mechanisms including the ubiquitination of H2AK119, an association with DNA effector proteins, direct binding of DNA, or the binding H3K27me3 itself (158).

1.4.3 DNA Methylation

Dating back to its initial discovery in 1948, DNA methylation has been established as one of the main pillars of epigenetic regulation and inheritance (159). In eukaryotes, methylation occurs predominantly on the 5' carbon of cytosine (5mC) in CpG dinucleotides. CpG dinucleotides are statistically underrepresented in eukaryotic genomes; likely due to a higher frequency of spontaneous deamination that's associated with 5mC (160, 161). Therefore, over evolutionary time only punctate clusters of hypomethylated CpG dinucleotides remained intact, forming (G+C)-rich domains known as CpG Islands (CGI) (162-164). Genome wide sequencing has revealed that over half of gene promoters contain an unmethylated CGI (165, 166). But the bulk of the genome's DNA methylation occurs at non-CGI CpGs within the gene body or at repetitive genomic elements (165, 167). There are, however, well-known examples of CGI methylation occurring at the promoter region of imprinted genes (discussed later) (168, 169, reviewed in 170).

DNA methylation at promoters is commonly associated with transcriptional inactivation and long-term gene silencing; however, evidence for causality of silencing is not so clear. For

years DNA methylation was thought to directly inhibit gene expression through steric exclusion of transcription factors (TFs) targeting specific DNA motifs. In addition, methylated DNA can interact with methyl-CpG-binding proteins (MBPs) or DNA methyltransferases that recruit co-repressive complexes to promote silencing via remodeling of histone modifications (171-176). But CpG methylation may not initiate the repressive state. For example, methylation can be acquired after the establishment of gene silencing to reinforce transcriptional shutdown (177). Moreover, CpG methylation at intragenic regions may counterintuitively promote transcriptional elongation by protecting against transcription at ectopic promoters and transposons inside the gene's body (166, 178, 179). Consequently, DNA methylation may play a major role in genome integrity through silencing of transposable elements.

Three DNA methyltransferases (DNMTs) are responsible for the global maintenance and establishment of methylation during DNA replication and reprogramming events. During DNA replication, an identical copy of DNA is formed through a semi-conserved mechanism. The replication machinery establishes the replication fork and unwinds the DNA helix into two parental templates to be used in synthesis of complementary daughter strands (Reviewed in 180). CpG methylation on the nascent DNA strands is also semi-conserved, and produces two hemimethylated helices. In order to maintain DNA methylation levels through replication, the protein Uhrf1 recognizes hemimethylated DNA at the replication fork and recruits DNMT1 to replicate the parental methylation state on the naïve daughter strand (181, 182). Consequently, DNMT1 serves a critical role in methylation inheritance between cell divisions. For example, mice and ES cells deficient in Dnmt1 lose 90% of their DNA methylation and die in early in embryogenesis (183).

During embryogenesis and development of the germline, global DNA methylation patterns are de-novo established after waves of epigenetic reprogramming remove most CpG methylation (discussed later). DNMT3a and its co-factor DNMT3L are essential for this process, and proper establishment of global DNA levels are required for embryo survival (184-187). Moreover, de-novo methylation machinery is required for setting of the germline and somatic imprints required for embryonic development. Indeed, altered methylation at these loci leads to disorders of overgrowth, declining behavior, and multiple cancers (188-193).

Although the recruitment of de-novo methylation machinery to specific loci remains an open question, the co-factor DNMT3L may link DNA methylation with the surrounding chromatin landscape. A binding domain in DNMT3L can recognize the unmethylated form of H3K4 and recruit DNMT3a for DNA methylation of these sites (194). Notably, aberrant retention of H3K4me resulting from oocyte KDM1B deficiency leads to a failure in imprinting due to a blocked DNMT3L-DNA interaction (195). In contrast, DNMT3L can act to prevent DNA methylation silencing at bivalent genes by competing with DNMT3a/b for PRC2 interactions at the poised promoter and excluding methyltransferase activity at these sites (196). On the other hand, DNMTs have also been shown to interact directly with histone methyltransferases, such as Suv39h1/2, Heterochromatin Protein 1 (HP-1) and G9a, to target *de novo* methylation at satellite repeats or promoters (174, 197-199). In summary, the establishment and maintenance of DNA methylation, especially at imprinted genes, are intimately involved in epigenetic inheritance, and may be regulated, in part, by already established chromatin modifications.

1.4.4 RNA and its Subclasses

1.4.4.1 Overview

RNA's role as a component of the epigenome has gained traction in recent years through the concerted focus on non-coding transcripts (ncRNA) (200-202). Analogous to the hypothetical Dark Matter of our universe, ncRNAs encompass a massive component of the mammalian genome, but as a whole, they are poorly understood. For instance, genome-wide RNA-sequencing efforts have demonstrated that the abundance of ncRNA far outweighs any protein-coding counterparts, with the majority of the mammalian genome being actively transcribed into non-coding species (67, 203-206). But not all ncRNAs are so poorly described. The role of ribosomal (rRNA) and transfer RNA (tRNA) in translation has been well established, but with the initial discovery of small interfering RNA (siRNA) (207), and microRNAs (miRNA) (208) a post-transcriptional regulatory function for these new class of transcripts has emerged. Since then, a surge of additional ncRNA subtypes has been discovered including short hairpin RNA (shRNA), Piwi interacting RNA (piRNA), tRNA-derived small RNA (tsRNA), also known as tRNA derived RNA fragments (tRf) (209), and long non-coding RNA (lncRNA) (210) among others (211-213). These regulatory classes of ncRNA commonly enact their function through mRNA complementarity and the subsequent inactivation of transcription products. However, some may act by establishing stable epigenetic changes at target loci via recruitment of epigenetic machinery, the modulation of protein activity, or even the simple act of being transcribed (214-219). Therefore, both the structure and sequence of ncRNA contain epigenetic information that is capable of shaping cellular activity. These characteristics may provide an additional means towards epigenetic control of gene regulation and may prove relevant in the context of epigenetic inheritance.

1.4.4.2 The Small ncRNA

This class of 20-30nt single-strand RNA molecule is increasingly diverse in their processing, structure, and biological function, but can be generalized into three main categories: miRNA, siRNA and piwiRNA (201, 213, 220). Both miRNA and siRNA begin as endogenous dsRNA precursors that are first cleaved by endoribonuclease RNAase III enzymes known as the Dicer family. miRNA and piRNA rely on association with a class of effector proteins known as the RNA-induced-silencing-complex (RISC) to activate silencing of mRNA postranscriptionally (221, 222). piRNA, which will be described later in the context of spermatogenesis, are not derived from Dicer activity, and rely on piwi-specific proteins for processing and functionality (223, 224)

ncRNA plays an important role in the regulation of epigenetic modifiers, and in turn, the control of their downstream epigenetic processes. For example, endogenous miRNA is capable of self-directing PRC2 by directly interacting with Suz12 (225). In T-cell and ESC differentiation, a short RNA fragment is detectable at promoters of genes targeted for epigenetic silencing. Through a stem-loop structure, the resulting miRNA directly recruits PRC2 to the target site and promotes the acquisition of H3K27me3 and stable gene repression in cis. Further examples of PRC2 targeting include miR-320-dependent coordination of the RNA effector protein Argonaute-1, Ezh2 and H3K27me3 to the promoter of POLR3D to establish gene silencing in mammalian cells (226). miRNA-promoter regulatory mechanisms are present at key pluripotency genes Nanog, Oct4, and Sox2 in mouse embryos (227), and due to the abundance of candidate miRNA target sites, are likely to be a factor across the mammalian genome (228, 229). RNA directed activity can direct heterochromatin formation at larger regions of the genome, such as in the context of X-inactivation (230). The small ncRNA that is internal to the Xist

transcript (RepA) binds directly to Ezh2 and is required for initiating X-chromosome inactivation, the recruitment of PRC2, and the spread of H3K27me3 along the inactive X (230). And in cancer cell lines, the miRNA class of *miR-29* can indirectly increase gene expression of tumour suppressor genes by targeting *Dnmt1* and *Dnmt3* transcripts to deplete methyltransferase activity and revert aberrant DNA hypermethylation (231, 232).

1.4.4.3 The Long ncRNA

In addition to these small RNA subtypes, the last decade has revealed a long non-coding RNA (lncRNA) subclass, which resembles mRNA in length (>200 bp) and splicing, but does not encode for any protein. Like small ncRNA, the lncRNA are made up of a diverse set of transcripts which arise mainly from protein coding loci but are also generated from intronic or intergenic sequences. lncRNA are capable of associating with epigenetic machinery to establish specific patterns of regulatory states either from local or distal interactions (233-236). A classic example of lncRNA dependent control of a major developmental locus is shown with the HOTAIR transcript. HOTAIR's primary role is the repression of the 40kbp HOXD locus (237, 238). This expansive range is achieved by acting as a modular scaffold from which the histone modifying complexes PRC2 and LSD1/CoREST/REST can bind and be chaperoned to their targets (238-241). This recruitment mechanism facilitates the establishment of heterochromatin through cooperative removal of H3K4 methylation and subsequent H3K27 tri-methylation and has been described in multiple similar cases (240, 241).

Genomic regions of known activating enhancer activity, such as those marked by H3K4me1 and H3K27ac (79, 242) have been shown to produce lncRNA transcripts termed enhancer-RNA (eRNA) (243, 244). The eRNAs regulate gene expression across large, and local, regions of the genome via mechanisms which involve transcription-factor binding (245, 246),

higher-order chromatin looping (244), and regulation of chromatin accessibility (247, 248). The expression of lncRNA is likely coordinated through epigenetic patterns localized to an enhancer region in a similar manner to what is described for protein-coding genes. Indeed, activating histone marks H3K4me3 and H3K36me3 have been associated with the TSS of lncRNA at regions rich in CpGs (249). Moreover, bivalent, H3K4me3 and H3K27me3, eRNAs have been described as poised, and functional in developmental pathways (249-251). In-vivo characterization of targeted enhancer lncRNA in a mouse model revealed functional roles in heart and brain development (250). Ablation of specific lncRNA further resulted in peri- and post-natal lethality (250). One of these genes, *Fendrr*, plays an essential role in embryogenesis through the long-term regulation of PRC2 directed silencing against *Foxf1* and *Pitx2* promoters in lateral mesoderm cells (215, 252).

1.5 Epigenetic Features of the Germline

1.5.1 Background

Relatively little is known about how the aforementioned epigenetic systems can contribute to intergenerational and transgenerational epigenetic inheritance. The phenomenon relies on an as yet undefined ability of parental epigenetic memory carried by the germline to escape a series of genome-wide reprogramming events in the next generations. During these periods of epigenetic remodelling, active and passive DNA demethylation accompanies major changes to chromatin landscape and gene expression in an effort to erase pre-programmed parental marks (Reviewed in 253). As development progresses, epigenetic patterns regulating lineage specification are restored and maintained across cell divisions.

1.5.2 Germline Establishment and Epigenetic Reprogramming

The establishment of the mouse germline begins early in the post-implantation embryo at embryonic day E6.25 (254, 255). Germ-cell fate commitment begins in the prestreak epiblast due to protein signals from bone morphogenic proteins BMP4, BMP8b, SMADs SMAD1/3/5, and WNT3 which are released from the extraembryonic ectoderm (256). These factors induce the expression of the transcriptional repressor, *Blimp1*, in a small number of cells in the proximal posterior epiblast (254, 255). The subsequent expression of specification factors *Prdm14* and *Ap2γ*(257, 258) creates what is widely accepted as the regulatory network required for germ cell lineage specification (259, 260). By E7.2, a founding population of ~40 PGCs are present at the base of the allantois (261). At E8, the PGCs begin their migration from the mesoderm and colonize the developing hindgut by E9 before moving to the genital ridge at E10. Once established in the genital ridge at E12.5, signals from the surrounding somatic cells direct sex differentiation and the formation of an embryonic gonad. The end of PGC development occurs at E13.5. This distinction is marked by either the female PGCs entering meiosis to arrest in prophase I until birth, or the male PGCs remaining briefly proliferative as pro-spermatogonia before entering mitotic quiescence until birth.

In the early stages of germline development, during migration to the genital ridge, PGCs undergo epigenetic erasure of genomic DNA methylation imprints and epigenetic signatures that were inherited from the cells of the prestreak epiblast. This reprogramming facilitates the establishment of a pluripotent cell state required for gametogenesis. The best characterized events of germline epigenetic erasure during this time involve DNA demethylation, the removal of parental imprints, and X-chromosome activation (262-267). In addition, chromatin remodeling events during mouse PGC migration are essential in the suppression of somatic mesoderm development and the subsequent specification of the germline (257, 268-271). These events are

tied to the global elimination of DNA methylation and the expression of the PGC specification factors *Blimp1* and *Prdm14* at E6.5 (257, 272, 273). Acting with BLIMP1 in the nucleus, an arginine methyltransferase, PRMT5, establishes repressive histone modifications on H2A and H4 (272). At E7.75, a substantial decrease in the repressive histone modification H3K9me2 is caused by the downregulation of the *Glp/G9a* histone methyltransferase complex, with nearly all PGCs showing H3K9me2 depletion by E8.75 (269, 274, 275). These events are controlled through the transcriptional repressor PRDM14 (257, 273). At E8.5, PGCs begin a gradual enrichment for H3K27me3 via PRC2-dependent activity that's regulated in part by PRDM14 (257, 270, 273, 275, 276). The germline H3K27me3 differs from that observed in ESC and somatic cells at developmental loci that may be key to maintaining suppression of mesodermal transcription (266, 276, 277). Therefore, H3K27me3 may be critical in further defining germline pluripotency (270).

In parallel, global DNA demethylation, including the removal of parental methylation imprints, is achieved through a combination of enzymatic activity, passive dilution through DNA replication, and base excision repair. The methyloxidase enzymes TET1 and TET2 are expressed between E9.5-E10.5 and actively convert almost all 5mC to 5-hydroxymethylcytosine (5hmC) by E11.5 (265). The resulting accumulation of 5hmC is then both passively and actively cleared from the genome through replication-dependent dilution, and the base excision repair pathway, respectively (265, 278). In addition, the passive dilution of 5mC also occurs in tandem through the suppression of de-novo and maintenance DNA methyltransferase activity by PRDM14 (273). Consequently, global DNA methylation reaches its lowest level of any mammalian cell by E13.5 (266).

It is at this point of germline development that the epigenome is in its most naïve state.

Nonetheless, DNA methylation persists at high levels in repeat elements such as retrotransposons (265-267, 279). Maintenance of methylation, as well as H3K9me3 and H3K27me3 histone methylation, at these repetitive regions may play a role in chromosome stability and protection against retrotransposition. Interestingly, resistance to DNA methylation also occurs at rare instances of functionally relevant genomic locations such as enhancers, CGI, promoters and gene bodies in both mouse and human PGCs (265, 266, 279). The penetrance of these methylation “escapees” support a possible mechanism for DNA methylation in transgenerational epigenetic inheritance. Possible mechanisms of inheritance surrounding these regions will be introduced later.

The newly established epigenetic ground state requires resetting and replacement of the epigenetic information required to support germline development and embryonic development post fertilization. This shift from pluripotency to the unipotency of gametogenesis includes germ-line specific variations of major epigenetic events such as de novo DNA methylation, chromatin modifications, and the sex-specific reprogramming of genomic imprints. De novo DNA methylation occurs asynchronously in male and female germ cells (280-282). In the female, low levels of methylation remain from E13.5 to E16.5, while the pro-spermatogonia of the male have already regained 50% of their global DNA methylation by E16.5 and fully established by birth (281). Genomic imprints are set at this time primarily via DNA methylation at gene promoters to regulate parent-of-origin gene expression in the next generation. The mechanisms underlying genomic imprinting acquisition and maintenance are the best understood examples of epigenetic inheritance of transcriptional control. Nevertheless, it remains undetermined how methylation machinery recognizes and is recruited to the specific regions

targeted for imprinting. Possible mechanisms involve recruitment through chromatin structure and histone modifications, as well as DNA sequence and transcriptional status (283-285).

Imprinted genes are generally found in clusters that can be defined as differentially methylated regions (DMR). These regions are controlled by an associated lncRNA or other similar regulatory element that's responsible for the monoallelic-expression of genes within the imprinting cluster. As such, the origin of these regulatory elements is often referred to as an imprint control region (ICR). A hallmark of genomic imprinting is that they are maintained throughout all somatic lineages of the next generation and are necessary for proper embryonic and extra-embryonic development (286, 287). To date, over 100 genes have been identified to be controlled by genomic imprinting in mammals. Although the majority of imprints are set in the maternal germ-line, some classical examples exist in the sperm. For example, paternal imprinting of the H19 locus illustrates a well conserved model of allele-restricted gene expression which plays a critical role in embryonic growth and development (Reviewed in 188). This locus contains the paternally expressed Igf2, and a maternally expressed lncRNA, H19 is flanked by downstream enhancers (288-291). The reciprocal allelic expression is controlled by the paternal methylation of a shared ICR 2kbp upstream of the H19 promoter and ~80kbp downstream of the distal Igf2 (292, 293). Methylation of the H19-ICR on the paternal allele directly silences H19-expression and promotes enhancer-mediated activation of paternal Igf2. In contrast, the hypomethylated maternal H19-ICR functions as a transcriptional insulator by binding CTCF to form restrictive boundaries that exclude enhancer activity at the maternal Igf2 (294-296). As a result of the insulator boundary, all enhancer activity is directed in cis to the promoter of the maternal H19 lncRNA.

Functional studies of imprinted lncRNA highlight how transcriptional control of a locus is the result of collaborative function between multiple epigenetic systems (233, 297-301). For example, the H19-lncRNA is required for the recruitment of the methyl-CpG binding domain protein 1 (MBD1) to genes within the cluster (300). Only the combination of H19-lncRNA and MBD1 is sufficient for the subsequent enrichment of repressive H3K9me3 at the target genes. In addition to H19, imprinted lncRNA may mediate the recruitment of repressive chromatin modifications to gene targets in trans across the locus to establish long term gene silencing (302-304). Moreover, the Xist lncRNA plays a similar role in the mechanisms of dosage-compensation and maintenance of X-inactivation (305-309). Xist coats the X-chromosome and interacts directly with a repressor complex, Sharp, to initiate X-inactivation through the Smrt co-repressor complex (310, 311). Subsequent silencing machinery, such as polycomb repressive complex 2 (PRC2), is then recruited to maintain the inactive X by the spreading of H3K27me3 heterochromatin domains (230, 311-314).

Recently, DNA-independent modes of genomic imprinting have implicated histone methylation as a novel mechanism of imprint control (315). Advancements in next-generation sequencing have enabled allele-specific assignment of sequencing reads from NGS assays in mouse preimplantation embryos. Using transposase (Tn5) hypersensitivity as a readout of open chromatin in the pronuclear stage 5 (PN5) zygote, paternal chromatin specific hypersensitive sites that are absent on the maternal allele have been identified (315). Strikingly, the difference in chromatin accessibility was not driven by the presence of maternal DMRs, or even by maternal DNA methylation. Instead, maternal H3K27me3 was inherited from the oocyte in a primarily DNA-independent manner. The maternally biased enrichment of H3K27me3 was detected at 76 hypomethylated genes in the ICM of blastocysts and had likely been maintained in

an imprinted-like state from the zygote. Moreover, imprinted-like genes displayed paternally biased expression in the morula that required maternal H3K27me3 and were independent of known DNA-dependent imprinted genes. However, unlike DNA-dependent imprints which are maintained long-term, the monoallelic-expression bias of imprint-like genes was reduced by the trophectoderm stage and lost in the epiblast. Nevertheless, the identification of histone-dependent genomic imprints brings forward a novel mechanism of epigenetic inheritance across generations.

1.5.3 Chromatin Remodeling in Spermatogenesis and Spermiogenesis

Spermatogenesis is well a coordinated, and highly organized process which can be divided into species-specific developmental stages (316-320). Continuous production of the massive number of sperm cells required for the reproductive lifespan of the male depends on the stem cell population called the spermatogonial stem cells (SSCs) and the balance between their ability to self-renewal or give rise to progenitors committed to spermatogenic differentiation. After sex determination in the embryonic germline, PGCs are integrated into the seminiferous chords and briefly proliferate as prospermatozoa before a period of mitotic arrest until shortly after birth. During this time, prospermatozoa undergo morphological and cellular changes and emerge on the outer basement membrane as the undifferentiated A_{single} or A_s spermatogonia (321). In the rodent, only a subset of A_s spermatogonia derived from a stem-cell niche are able to function as SSC (Reviewed in 322). Thus, during neonatal development, a period of expansion is required to establish a pool suitable for maintaining the SSC and non-stem cell progenitor spermatogonia population. SSC may be enriched, but not purified, by using specific cell surface markers such as Thy1, TSPAN8 (323-325). Further enrichment may be achieved using cells

expressing high level of GFP signal when isolated from Id4-eGFP reporter mice, as well as using combinations of reporter gene and cell surface marker signal (325, 326). As the undifferentiated A_{single} spermatogonia proliferate, they undergo incomplete cytokinesis between cell divisions, thus forming a syncytium of interconnected cells. These cellular contacts create populations of undifferentiated spermatogonia pairs (A_{paired}), or multiplets (A_{aligned}). Extrinsic and intrinsic signals such as retinoic acid (RA) and STRA8 (327-329), and the RNA-binding protein, DAZL (330) induce an irreversible differentiation into A1 spermatogonia. This transition occurs asynchronously throughout the seminiferous epithelium due to pulses of RA at stages VIII and IX of the spermatogenic cycle (316, 331, 332). These events form the spermatogenic wave and are responsible for the continuous release of mature spermatozoa.

The now differentiating spermatogonia continues mitotic divisions of A1-A4 spermatogonia, intermediate, and finally B-spermatogonia before forming the tetraploid primary spermatocyte to initiate the meiotic prophase. It is through the reductive divisions of meiosis that the haploid gamete is formed. At the time of prophase I, chromosomes condense homologous pairs form along the synaptonemal complex. In the pachytene spermatocyte, genetic recombination shuffles parental information prior to separation of sister chromatids. As the synaptonemal complex is degraded and the homologous chromosomes begin to separate, the spindle fibers attach and move the pairs along the metaphase plate. Meiosis I completes by the segregation of chromosomes into two diploid daughter cells called the secondary spermatocytes. Meiosis II then completes the chromosome reduction through the segregation of sister chromatids, a cell division, and the production of four haploid spermatids.

Chromatin in the spermatocyte undergoes dynamic changes to its histone modifications and histone variant composition which are indispensable to meiosis and progression of

spermatogenesis (Reviewed in 333). Indeed, dynamic modulation of histone H3 methylation begins from leptotene spermatocytes and continues throughout spermatogenesis (334). In mouse and marmoset, levels of H3K4 methylation are inversely correlated with the protein abundance of the histone demethylase KDM1A, which enacts the majority of its function in the pachytene spermatocyte (334). This is suggestive of specific and temporal regulation of histone methylation throughout germ cell development. Indeed, KDM1A is essential for germ cell differentiation and entry into meiosis (335, 336). Moreover, the histone methyltransferase PRDM9 plays critical roles in meiotic recombination through the recognition of DNA recombination hotspots and trimethylation of H3K4 and H3K36 on adjacent nucleosomes (337-339). The *Suv39h* histone methyltransferases regulate acquisition and maintenance of heterochromatin in embryogenesis and spermatogenesis by trimethylation of H3K9 (340, 341). In the spermatocyte, a double knockout of *Suv39h1* and *Suv39h2* leads to chromosome instability, nonhomologous chromosome associations and complete failure of spermatogenesis (136).

The male germline contains one of the largest and most diverse populations of variant histones when compared to other mammalian cell types (333, 342). Despite the seemingly minor differences in amino acid sequence, histone variants play important roles in chromatin organization and gene expression and are critical for fertility and proper sperm development. For example, mice with heterozygous germline mutation of the H3.3 encoding gene *H3f3b* results in marked reduction of H3.3 and either complete mitotic arrest in round-spermatids, or reduced protamine incorporation, sperm head abnormalities, and infertility (343-345). Furthermore, *H3f3b* null mice showed reduced gene expression in the round spermatid cause by abnormal spreading of H3K9me3 during spermatogenesis (344). The sperm are also host to testis-specific

H2A and H2B histone variants which are associated with heterochromatin and nucleosome remodeling (346-348).

Variant incorporation is regulated by multiple processes but is driven primarily by transcription and nucleosome turnover. In somatic cells, where these mechanisms are best understood, ATP- and DNA-replication-dependent remodelers act to disrupt core histones and replace them with either recycled or newly synthesized, naïve, histones (Reviewed in 349). These processes generally result in rapid nucleosome exchange at active promoters (350-355). However, the presence of localized chromatin factors may influence the rate of turnover at specific regions. For example, PRC2 at heterochromatin regions may suppress nucleosome turnover, reduce variant incorporation and thus maintain epigenetic memory of the chromatin state (352-355). Alternatively, histone variant incorporation may be replication independent and occur through the action of histone-chaperones. HIRA promotes the inclusion of H3.3 at active genes and promoter regions in both germ and somatic cells (278, 356-360). The actions of HIRA are crucial for the inclusion of H3.3 at polycomb repressed loci and facilitate the establishment of bivalent domains found in ESC, spermatocytes and spermatids (95, 96, 361, 362).

Post-meiotic development in the haploid round spermatid initiates the chromatin compaction and cell condensation that's characteristic of the mature sperm. The acquisition of this specialized morphology involves major chromatin remodeling events that reshape the nuclear structure. Nuclear elongation begins after a transcriptional burst in the round spermatid that signifies the end of germ cell transcriptional activity. The nucleosomes of the elongating spermatid begin to be replaced by the small, highly basic, proteins called protamines (363). In mammalian spermiogenesis, the nucleosome to protamine transition is mediated by a protein intermediate known as the transition protein (364, 365). These structural proteins enable the

efficient folding of DNA into a compact toroid structure and improve DNA resistance to damage by nucleases and reactive oxygen species (ROS) (366-368).

The mechanisms to explain how histone replacement occurs remain unknown, but recent studies suggest the incorporation of histone variants destabilizes the nucleosomes and facilitates its eviction. For example, the incorporation of testis-specific H2A/H2B histone variants, or the combination of H3.3 and H2A.Z somatic-type histone variants within a nucleosome can induce nucleosome instability and create regions of open chromatin (346-348, 358, 369). The TH2B variant may have the largest impact on the protamine transition as it replaces nearly all H2B in male germ cells (370). Additional nucleosome instability may be conferred by the hyperacetylation of histone tails. The association of open chromatin and histone acetylation has been well described in transcriptionally active somatic cells; however, in many species, the nucleosomes of the elongating spermatid remain hyperacetylated despite no transcriptional activity (371, 372). Moreover, a testis specific bromodomain containing chromatin remodeler, *Brdt*, induces stage dependent remodeling in the spermatid that's dependent on histone hyperacetylation (373, 374). Together, these processes support the hypothesis that a successful protamine transition first relies on the destabilization and expansion of spermatid nucleosomes prior to chromatin compaction (375-377).

In spite of the aforementioned transition events, nucleosome replacement is not complete (378-381). In the mature sperm of mouse and men, estimates of nucleosome retention range between 1% to 8%, and 10% to 15%, respectively (251, 378). At the beginning of this thesis work, the role of sperm nucleosomes was not well described. Pioneering studies of the sperm epigenome have determined that nucleosomes remain preferentially enriched at certain transcriptional-start sites in mouse and men (251, 380-384). Moreover, retained nucleosomes are

marked by histone modifications that associate with genes involved in development, and housekeeping processes (380, 381). This suggests nucleosome retention is a non-random event. Indeed, sperm nucleosomes are highly correlated with the location of hypomethylated CGIs and transcriptional status in the round spermatid (382). Furthermore, the specific variant composition and associated modifications are also determined by these parameters (385). For example, CGIs that were associated with high levels of transcription in the round spermatid are enriched with the histone variant H3.3 and the active histone modifications H3K4me3, H3K4me1, H3K27ac, and H3K9ac (251, 385). As a result, nucleosome retention may be regulated by nucleosome turnover at these active sites. Conversely, low expression CGIs in the spermatid experience reduced nucleosome turnover and retain nucleosomes with the canonical form of histones H3.1 and H3.2 marked by repressive H3K27me3 (251, 385). Together, these results suggest a model where resistance to nucleosome eviction in spermiogenesis is established through the competitive binding of transcription factors and chaperones at active CGIs. Alternatively, competitive interference between nucleosome eviction machinery and histone-bound repressive factors, such as PRC2, may promote nucleosome retention at transcriptionally silent genes (385).

1.5.4 RNA Content in Mature Sperm

A unique component of spermatocytes is a large ribonucleoprotein granule known as the chromatoid body (CB). Composed of hundreds of RNA binding proteins, piwi-interacting RNA processing machinery, and mRNA and ncRNA molecules, the CB plays a major role in controlling the RNA content of the male germ cell (386-390). Removal of specific CB components impairs spermatogenesis and fertility, and thus implicates the RNA pool in the proper development of sperm (391).

After a transcriptional burst in the round spermatid, the ensuing chromatin condensation reduces spermatid transcription to a halt. However, RNA transcripts were detected in sperm as early as 1989 (392). Although widely accepted as the historical remains of meiotic and spermatid transcription, sperm RNA may have a functional role. The first examination of human sperm RNA came from microarray studies, where 3000-7000 transcripts were discovered in sperm from fertile men (393). This was followed by a serial analysis of gene expression, detecting 4000-5000 sperm mRNA transcripts (394). Microarrays have since been used to show RNA transcripts are abundant in the sperm of other mammals (395, 396), *Drosophila* (397), and plants (398). With the coming of age of RNA-seq, quantitative assessment of the sperm's RNA population has revealed an abundance of sperm borne transcripts. The discovery of over 22,000 human species of coding and non-coding sperm RNA suggests high potential for functional relevance when delivered to the oocyte (399). But sperm transcripts are typically of very poor quality; heavily fragmented, with extensive truncation at the 3' end (399). Moreover, they exist in extremely low levels, nearly 200-fold lower than somatic cells (399-401). Nevertheless, small non-coding RNAs and even small RNA fragments have been shown to have functional roles in models of mouse and men (Reviewed in 402)

Among the most abundant small RNAs in human sperm are the PIWI-interacting RNA (piRNA), which are typically between 24-31nt in length (224). Most piRNA exist in clusters that span transposon regions, and function to repress transposable elements during meiosis. When piRNA transcripts are loaded into PIWI proteins, the complex functions to silence these regions through specific targeting and complementary binding (403-407). The piRNA clusters found in human sperm contain artefacts of dead protein-coding genes, known as pseudogenes, which are transcribed and processed similarly to transposons (224). piRNA derived from these

pseudogenes contributes to a large population sperm-borne piRNA and are predicted to function through silencing of their parent-protein-coding gene via direct targeting or through the associated binding of epigenetic machinery (224). In *Drosophila*, the piRNA-PIWI complex can even act to increase expression of some genes by interaction with nascent transcripts at euchromatin, or function to silence regions by associating directly with DNA at heterochromatin (404). Moreover, sites with piwi-dependent H3K9 methylation in the *Drosophila* ovary can recruit a protein complex in the HP1 family that functions to transcribe piRNA precursors locally. This mechanism may serve as a self-renewing feedback loop similar to those already well described in fungi and plants (405, 408-412). The directed reinforcement of an epigenetic signal may contribute to its perpetuation across gamete development and thus promote the inheritance of epigenetic memories to the offspring of the next generation.

1.6 Epigenetic Mechanisms of Inheritance

1.6.1 Background

The sperm epigenome is well positioned to regulate developmental plasticity of the next generation. However, the common theme in the examples presented above is that there is very little mechanistic description that's capable of explaining how this is achieved. Moreover, in examples of transgenerational epigenetic inheritance, the mechanistic underpinnings are even more complex, as the molecular signal must be able to escape or resist major epigenetic reprogramming events in the PGC and embryo. Knowledge of how paternal epigenetic inheritance occurs across generations is critical to understanding how the father's lifestyle impacts his offspring's health. Limitations to our understanding have historically been due to the

lack of robust functional tests of epigenetic components. Nonetheless, DNA methylation, RNA and histone modifications continue to be implicated in the transmission of paternal effects.

1.6.2 Inheritance through DNA methylation

The first molecular evidence for germ-line transgenerational inheritance was described at endogenous metastable epialleles in the mouse. Epialleles, such as agouti variable yellow (A^{vy}), and axin fused ($Axin^{FU}$) are formed from the random insertion of an intracisternal A particle (IAP) retrotransposon near or inside the gene (413). Transcriptional control of the locus is then hijacked by the promoter elements of the retrotransposon and becomes dependent on their epigenetic state. For example, DNA methylation of the IAP is inversely correlated with transcriptional activity of the agouti gene: hypomethylation of the A^{vy} IAP results in the hyperexpression of agouti and a characteristic yellow coat and obesity phenotype, while hypermethylation results in normal expression and regular variation in coat colour. In yellow, agouti-expressing dams, the inheritance of yellow coat phenotype by offspring occurs at a frequency that defies Mendelian ratios and suggests germ-line inheritance of the IAP epigenetic-state. However, the coat colour of A^{vy} sires has no effect on offspring phenotype. In contrast, the $Axin^{FU}$ phenotype of a kinked tail is transgenerationally inherited in both the maternal and paternal lineages. Interestingly, the DNA methylation state of the $Axin^{FU}$ IAP in the somatic tissue matches the methylation state in the mature sperm, providing an example of paternal epigenetic inheritance that escapes germline reprogramming (413).

1.6.2.1 Models of Nutritional Exposure

The idea that diet and environment can influence the epigenome has been supported in recent studies, but a dearth of evidence exists to describe any mechanism. Altering the levels of the B vitamin folate (B_9) is a powerful tool to study how the epigenome responds to

environmental insults in animal models. Folate functions in maintaining adequate levels of SAM, which acts as the principle methyl donor for DNA methylation and other biological methylation reactions. Therefore, folate plays an essential role in the establishment of the methylome.

Because many of the critical genes associated with embryonic development are subject to parent-of-origin imprinting or have an association with germline methylation, perturbations in DNA methylation during germ-line development may result in long term susceptibility to disease in the next generation.

Paternal folate deficiency in Sprague Dawley rats resulted in lower folate and DNA methylation levels in the liver of offspring in comparison to folate supplemented fathers (414). Interestingly, because the maternal rat was supplemented with folate, the nursing offspring would receive supplemented levels of folate through the nursing period; however, this was not sufficient to produce the recovery of hepatic methylation to the levels of those from the supplemented folate group after 21 days. This implicates that paternal epimutations may have been established in the mature sperm preconceptually and contribute to the changes in the offspring epigenome. It's important to consider that folate over-supplementation may be a confounding factor that could saturate methylation-donor pools, promoting hypermethylation of certain genomic regions in comparison to normal folate levels (2mg/kg). Therefore, it is unclear if changes observed in the offspring were a direct consequence of hypo or hypermethylated changes in the germline.

When male mice were subject to a lifetime folate deficiency, beginning in utero, multiple regions of the sperm epigenome were differentially methylated in comparison to the folate sufficient males as detected by pyrosequencing (415). Specifically, the genes affected were associated with metabolism and developmental processes with implications in chronic disease

such as diabetes. The increased rates of birth defects observed in the folate deficient offspring strengthens the argument for a paternally mediated effect on offspring in response to lifestyle. These results were recently repeated in the case of folate deficiency, while also expanded to show that over-supplementation of folate can also alter the sperm methylome of mice and men (416, 417).

The susceptibility of sperm to environmental exposure is further exemplified in human newborns of obese parents. DNA from cord blood leukocytes was analyzed in a small cohort of babies from obese fathers via bisulphite sequencing and revealed significant hypomethylation at the DMRs of three genes, and a trend for hypermethylation at two DMRs (418). Thus, depending on the DMR studied, pre-conception exposure to nutritional excess may subtly influence the normal programming state of sperm DNA methylation. Moreover, rats fed a high-fat diet beginning at 4 weeks of age sired female progeny with impaired insulin secretion and altered methylation at *Il13ra2* gene (419). These observations share in common a nutritional insult occurring during the pre-pubertal window and onset of spermatogenesis. The resulting subtle changes to DNA methylation patterns are suggestive of a susceptibility to environmental influence during this highly dynamic differentiation period (334).

Nutritional deficiencies have also been shown to influence the methylation of discrete loci associated with transcriptional expression during early embryonic life (420). A mouse model which restricted pregnant dam nutrition by 50% in the 3rd week of pregnancy, which corresponds with the window of male PGC epigenetic programming, resulted in hypomethylation at distinct DMRs in F1 mature sperm despite postnatal exposure to a normal diet. Remarkably, the paternal effect persisted through to the F2, as gene expression was altered at regions which showed methylation differences in the F1. Although the DNA methylation change itself did not persist

into the tissues, the effects on gene expression may have been established early on in development, prior to resetting. While mechanistically it is unclear how nutritional conditions induce germline epigenetic changes capable of influencing subsequent generations, it is becoming evident that the male germ line epigenome is susceptible to environmental exposures (35, 415, 416, 421-424). It must be considered that any changes to the pool of methyl-donors would not only alter the DNA methylation landscape, but also affect histone methylation. As a result, nutritional studies should not exclude the chromatin component of sperm when searching for a mechanism for epigenetic inheritance.

1.6.3 Inheritance Through RNA

1.6.3.1 Potential Role of Sperm Borne Transcripts Post-Fertilization

Intact sperm specific mRNAs have been recovered from the zygote post fertilization (425). Mounting evidence for roles of intact transcripts with roles in post-fertilization have been identified to include signalling pathway members such as the JAK/STAT pathway, WNT signalling and EGR3 (399), perhaps serving as pre-transcribed and readily translatable genes for immediate function in development. It has also been suggested that some RNA transcripts may serve dual functions. That is, a gene coding mRNA is capable of performing non-coding, regulatory roles, depending on its structure or sequence (Reviewed in 426). Moreover, a gene ontology analysis of the most abundant, intact, transcripts showed enrichments for functional roles in post-transcriptional regulation and RNA binding (399). These processes are suggestive of a mechanism by which post-fertilization transcription can be regulated (399). In particular, the intact sperm RNA INTS1 encodes a subunit of the complex required to process spliceosome components for production of small RNA, and is required for blastocyst growth (427). RNA

molecules like these may function to regulate transcription activation in the early zygote and propagate small RNA with roles in development. As an example, the highly abundant sperm borne microRNA transcript, *mir-34c*, was shown to be essential for the first cleavage division of mouse embryos (428), yet appears to be paternally contributed (399).

Functional tests of RNAs contribution to inheritance patterns have been applied through microinjections of RNA mixtures to recapitulate mutant phenotypes. For example, a heterozygote mouse model containing a null insertion mutation in *cKit* presents with a white-tailed phenotype that persists to a variable degree in the genetically wild-type progeny (429). Surprisingly, this phenotype was replicated in wild-type progeny when either total sperm RNA from cKit-mutant heterozygotes, or isolated miRNA which specifically targets the cKit transcript, was injected into wild-type zygotes (429). In addition, zygotic microinjection of synthesized miRNA recapitulated a hypothalamic-pituitary-adrenal (HPA) stress reactivity phenotype from a mouse model of paternal stress (430). Other RNA-microinjection studies have confirmed that sperm RNA introduced to the zygote can induce epigenetic modulation of their transcripts (34, 430-433). These results suggest that non-genetic inheritance has a significant RNA component. It remains unknown how these transient RNA molecules can function on their own to mediate epigenetic inheritance across cell divisions and generations. Nevertheless, a growing body of work suggest RNA may facilitate the stable propagation of paternal effects from the soma, to the germline, to be inherited by the next generation (158).

In agreement with this theory, a mouse model of early-life maternal stress induced a transgenerational phenotype that was inherited through the male germline (34, 434). Unpredictable maternal separation during pre-weaning age was shown to perpetuate a reduction in avoidance and response to fear in the F2 progeny (34). Interestingly, the phenotype was

reproduced when wild-type zygotes were microinjected with total sperm-RNA from stressed males. In this study, the F1 sperm showed altered levels of small ncRNA, with an increased abundance in several miRNA transcripts and a decreased abundance in piRNA. Although the changes to sncRNA were not observed in the F2 sperm, somatic tissues such as hippocampus and serum of both F1 and F2 stressed males were affected. Therefore, it was suggested that the epigenetic information conveyed by ncRNA in the F1 generation is transmitted to other non-genomic systems such as DNA methylation, or chromatin modifications for maintenance of phenotype transmission. However, in non-mammalian systems such as *C. elegans*, starvation induced changes to germline ncRNA pools may remain stable across multiple generations (435). This stability is suggestive of a regulatory mechanism that perpetuates the normally transient, unstable RNA molecule.

While the exact roles of thousands of ncRNA remain unknown it is difficult to describe a mechanism by which RNA alone can mediate transgenerational epigenetic inheritance. However, if we consider the ability of small RNA to self-renew, lncRNA's capacity to modulate transcription factor activity and histone modifications at local or distal regions, and their critical role in developmental processes, one can imagine that sperm delivery of these molecules could transmit a molecular signal that directs gene expression and remodelling events in the next generation.

1.6.4 Inheritance through Histone Modifications

As introduced above, the nucleosomes in sperm are well positioned to regulate the early stages of embryonic development. However, there remains little mechanistic evidence to suggest that sperm nucleosomes have any impact. Studies from mouse and men have shown through

immunofluorescence that post-translationally modified sperm histones are detectable in the early zygote (436, 437). A possible regulatory role for these nucleosomes in the mouse 2-cell embryo was shown by 2-cell gene expression changes after their retention in sperm was disrupted (438). Moreover, in men, incomplete nucleosome replacement to protamines is associated with reduced fertility and poor pregnancy outcomes (439-441). These results may explain the structural importance of nucleosomes, but do not address the epigenetic modifications present on the histones.

Still, it remains poorly understood how histone modifications are capable of mediating the heritable transmission of parental epigenetic memory between generations. Only recently has definitive evidence shown that histone methylation is capable of this phenomenon. In the absence of maternal PRC2, *C. elegans* larva display paternally inherited H3K27me3 that is maintained up to the 16-24 cell stage (442). In the reverse condition, where paternal PRC2 is absent, paternal H3K27me3 is not re-established in the zygote or early embryo. This study was the first to demonstrate the faithful propagation of a paternally inherited germ-line histone modification during the early stages of embryonic development. A recent set of elegant experiments in *Drosophila* embryos identified a causal role for histone modification in the inheritance of a parental epigenetic memory around the HOX gene cluster (443). Once H3K27me3 HOX gene silencing was established, the inherited memory was perpetuated for at least 8 cell divisions by the direct recognition and copying of the parental mark with embryonic PRC2. The inheritance of pre-existing H3K27me3 from the germ-line to the embryo is again supported in a *Drosophila* model of maternal transmission (444). This study knocked down E(z), the PRC2 catalytic subunit responsible for establishment of H3K27me3, creating an embryo deficient in its ability to de-novo establish H3K27me3 (E(z)-KD). However, early E(z)-KD

embryos still showed detectable levels of H3K27me3. Unlike wild-type embryos, which perpetuate the inherited H3K27me3 up to zygotic genome activation (ZGA), E(z) deficient embryos display rapid depletion of H3K27me3 after only 2-3 cell divisions. Interestingly, HOX genes were already marked with H3K27me3 at the earliest embryonic stage, but when they became depleted, the Ez-KD embryos displayed homeotic transformations and none survived embryogenesis. Embryonic survival rates were unable to be restored when Ez was overexpressed from ZGA onward, highlighting the requirement for inherited H3K27me3. The factor driving this phenotype was confirmed to be epigenetically regulated by H3K27me3 itself. Mutations of lysine 27 on H3 K27 to methionine mimics the observed embryonic lethality and homeotic transformations. Loss of H3K27me3 in Ez-KD embryos resulted in abnormal spreading of H3K27ac into a large subclass of poised or repressed enhancers are normally marked by H3K27me3 and H3K4me1. The accumulation of active marks resulted in the transcriptional up-regulation of a subset of genes that are turned on later in development. This suggests that maternally inherited H3K27me3 prevents precocious activation through regulation of both lineage specific genes and enhancers.

1.7 References

1. J. Alfoldi, K. Lindblad-Toh, Comparative genomics as a tool to understand evolution and disease. *Genome research* **23**, 1063-1068 (2013).
2. P. M. Visscher, M. A. Brown, M. I. McCarthy, J. Yang, Five years of GWAS discovery. *Am J Hum Genet* **90**, 7-24 (2012).
3. J. Shendure *et al.*, DNA sequencing at 40: past, present and future. *Nature* **550**, 345-353 (2017).
4. T. A. Manolio *et al.*, Finding the missing heritability of complex diseases. *Nature* **461**, 747-753 (2009).
5. E. E. Eichler *et al.*, Missing heritability and strategies for finding the underlying causes of complex disease. *Nat Rev Genet* **11**, 446-450 (2010).

6. T. A. Manolio, Cohort studies and the genetics of complex disease. *Nat Genet* **41**, 5-6 (2009).
7. R. Barouki, P. D. Gluckman, P. Grandjean, M. Hanson, J. J. Heindel, Developmental origins of non-communicable disease: implications for research and public health. *Environ Health* **11**, 42 (2012).
8. J. G. Wilson, Experimental studies on congenital malformations. *J Chronic Dis* **10**, 111-130 (1959).
9. A. C. Haugen, T. T. Schug, G. Collman, J. J. Heindel, Evolution of DOHaD: the impact of environmental health sciences. *J Dev Orig Health Dis* **6**, 55-64 (2015).
10. D. E. McBride, Septic abortion followed by septic shock. *J Am Osteopath Assoc* **60**, 545-552 (1961).
11. W. Lenz, K. Knapp, Thalidomide embryopathy. *Arch Environ Health* **5**, 100-105 (1962).
12. W. Lenz, A short history of thalidomide embryopathy. *Teratology* **38**, 203-215 (1988).
13. M. T. Miller, K. Stromland, Teratogen update: thalidomide: a review, with a focus on ocular findings and new potential uses. *Teratology* **60**, 306-321 (1999).
14. D. J. Barker, The fetal and infant origins of adult disease. *BMJ* **301**, 1111 (1990).
15. D. J. Barker, C. Osmond, J. Golding, D. Kuh, M. E. Wadsworth, Growth in utero, blood pressure in childhood and adult life, and mortality from cardiovascular disease. *BMJ* **298**, 564-567 (1989).
16. D. J. Barker, C. Osmond, C. M. Law, The intrauterine and early postnatal origins of cardiovascular disease and chronic bronchitis. *J Epidemiol Community Health* **43**, 237-240 (1989).
17. D. J. Barker, P. D. Winter, C. Osmond, B. Margetts, S. J. Simmonds, Weight in infancy and death from ischaemic heart disease. *Lancet* **2**, 577-580 (1989).
18. C. N. Hales, D. J. Barker, Type 2 (non-insulin-dependent) diabetes mellitus: the thrifty phenotype hypothesis. 1992. *Int J Epidemiol* **42**, 1215-1222 (2013).
19. C. Osmond, D. J. Barker, Fetal, infant, and childhood growth are predictors of coronary heart disease, diabetes, and hypertension in adult men and women. *Environ Health Perspect* **108 Suppl 3**, 545-553 (2000).
20. L. C. Schulz, The Dutch Hunger Winter and the developmental origins of health and disease. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 16757-16758 (2010).
21. T. Roseboom, S. de Rooij, R. Painter, The Dutch famine and its long-term consequences for adult health. *Early Hum Dev* **82**, 485-491 (2006).
22. L. H. Lumey, F. W. Van Poppel, The Dutch famine of 1944-45: mortality and morbidity in past and present generations. *Soc Hist Med* **7**, 229-246 (1994).
23. A. C. Ravelli *et al.*, Cardiovascular disease in survivors of the Dutch famine. *Nestle Nutr Workshop Ser Pediatr Program* **55**, 183-191; discussion 191-185 (2005).
24. A. C. Ravelli, J. H. van Der Meulen, C. Osmond, D. J. Barker, O. P. Bleker, Obesity at the age of 50 y in men and women exposed to famine prenatally. *Am J Clin Nutr* **70**, 811-816 (1999).
25. G. P. Ravelli, Z. A. Stein, M. W. Susser, Obesity in young men after famine exposure in utero and early infancy. *N Engl J Med* **295**, 349-353 (1976).

26. Z. Stein, M. Susser, G. Saenger, F. Marolla, Famine and human development: The Dutch hunger winter of 1944-1945. (1975).
27. C. N. Hales, D. J. Barker, The thrifty phenotype hypothesis. *Br Med Bull* **60**, 5-20 (2001).
28. R. A. Waterland, R. L. Jirtle, Early nutrition, epigenetic changes at transposons and imprinted genes, and enhanced susceptibility to adult chronic diseases. *Nutrition* **20**, 63-68 (2004).
29. R. A. Waterland, C. Garza, Potential mechanisms of metabolic imprinting that lead to chronic disease. *Am J Clin Nutr* **69**, 179-197 (1999).
30. R. L. Jirtle, M. K. Skinner, Environmental epigenomics and disease susceptibility. *Nat Rev Genet* **8**, 253-262 (2007).
31. J. J. Heindel, L. N. Vandenberg, Developmental origins of health and disease: a paradigm for understanding disease cause and prevention. *Curr Opin Pediatr* **27**, 248-253 (2015).
32. M. H. Vickers, Developmental programming and transgenerational transmission of obesity. *Ann Nutr Metab* **64 Suppl 1**, 26-34 (2014).
33. M. Pembrey, R. Saffery, L. O. Bygren, E. Network in Epigenetic, E. Network in Epigenetic, Human transgenerational responses to early-life experience: potential impact on development, health and biomedical research. *J Med Genet* **51**, 563-572 (2014).
34. K. Gapp *et al.*, Implication of sperm RNAs in transgenerational inheritance of the effects of early trauma in mice. *Nature neuroscience* **17**, 667-669 (2014).
35. M. D. Anway, A. S. Cupp, M. Uzumcu, M. K. Skinner, Epigenetic transgenerational actions of endocrine disruptors and male fertility. *Science* **308**, 1466-1469 (2005).
36. O. Babenko, I. Kovalchuk, G. A. Metz, Stress-induced perinatal and transgenerational epigenetic programming of brain development and mental health. *Neurosci Biobehav Rev* **48**, 70-91 (2015).
37. J. J. Faith, J. F. Colombel, J. I. Gordon, Identifying strains that contribute to complex diseases through the study of microbial inheritance. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 633-640 (2015).
38. J. R. Seckl, M. J. Meaney, Early life events and later development of ischaemic heart disease. *Lancet* **342**, 1236 (1993).
39. D. D. Francis, M. J. Meaney, Maternal care and the development of stress responses. *Curr Opin Neurobiol* **9**, 128-134 (1999).
40. J. Fabia, T. D. Thuy, Occupation of father at time of birth of children dying of malignant diseases. *Br J Prev Soc Med* **28**, 98-100 (1974).
41. D. A. Savitz, E. A. Whelan, R. C. Kleckner, Effect of parents' occupational exposures on risk of stillbirth, preterm delivery, and small-for-gestational-age infants. *Am J Epidemiol* **129**, 1201-1218 (1989).
42. E. van Wijngaarden, P. A. Stewart, A. F. Olshan, D. A. Savitz, G. R. Bunin, Parental occupational exposure to pesticides and childhood brain cancer. *Am J Epidemiol* **157**, 989-997 (2003).
43. M. Feychtig, N. Plato, G. Nise, A. Ahlbom, Paternal occupational exposures and childhood cancer. *Environ Health Perspect* **109**, 193-196 (2001).
44. J. F. Logman, L. E. de Vries, M. E. Hemels, S. Khattak, T. R. Einarson, Paternal organic solvent exposure and adverse pregnancy outcomes: a meta-analysis. *Am J Ind Med* **47**, 37-44 (2005).

45. M. El-Helaly, K. Abdel-Elah, A. Haussein, H. Shalaby, Paternal occupational exposures and the risk of congenital malformations--a case-control study. *Int J Occup Med Environ Health* **24**, 218-227 (2011).
46. E. National Academies of Sciences, Medicine, M. Institute of, P. Board on the Health of Select, H. Committee to Review the Health Effects in Vietnam Veterans of Exposure to, *Veterans and Agent Orange: Update 2014*. (National Academies Press, Washington (DC), 2016), pp. 1114.
47. G. Kaati, L. O. Bygren, S. Edvinsson, Cardiovascular and diabetes mortality determined by nutrition during parents' and grandparents' slow growth period. *Eur J Hum Genet* **10**, 682-688 (2002).
48. M. E. Pembrey *et al.*, Sex-specific, male-line transgenerational responses in humans. *Eur. J. Hum. Genet.* **14**, 159-166 (2006).
49. L. O. Bygren, G. Kaati, S. Edvinsson, Longevity determined by paternal ancestors' nutrition during their slow growth period. *Acta Biotheor* **49**, 53-59 (2001).
50. J. Golding, M. Pembrey, R. Jones, A. S. Team, ALSPAC--the Avon Longitudinal Study of Parents and Children. I. Study methodology. *Paediatr Perinat Epidemiol* **15**, 74-87 (2001).
51. A. R. Ness, The Avon Longitudinal Study of Parents and Children (ALSPAC)--a resource for the study of the environmental determinants of childhood obesity. *Eur J Endocrinol* **151 Suppl 3**, U141-149 (2004).
52. K. Northstone, J. Golding, G. Davey Smith, L. L. Miller, M. Pembrey, Prepubertal start of father's smoking and increased body fat in his sons: further characterisation of paternal transgenerational responses. *Eur J Hum Genet* **22**, 1382-1386 (2014).
53. G. Kaati, L. O. Bygren, M. Pembrey, M. Sjostrom, Transgenerational response to nutrition, early life circumstances and longevity. *Eur J Hum Genet* **15**, 784-790 (2007).
54. T. H. Chen, Y. H. Chiu, B. J. Boucher, Transgenerational effects of betel-quid chewing on the development of the metabolic syndrome in the Keelung Community-based Integrated Screening Program. *Am J Clin Nutr* **83**, 688-692 (2006).
55. A. M. Yen *et al.*, A population-based study of the association between betel-quid chewing and the metabolic syndrome in men. *Am J Clin Nutr* **83**, 1153-1160 (2006).
56. R. Loomba *et al.*, Parental obesity and offspring serum alanine and aspartate aminotransferase levels: the Framingham heart study. *Gastroenterology* **134**, 953-959 (2008).
57. S. A. Porter *et al.*, Aminotransferase levels are associated with cardiometabolic risk above and beyond visceral fat and insulin resistance: the Framingham Heart Study. *Arterioscler Thromb Vasc Biol* **33**, 139-146 (2013).
58. K. Luger, T. J. Rechsteiner, A. J. Flaus, M. M. Waye, T. J. Richmond, Characterization of nucleosome core particles containing histone proteins made in bacteria. *J Mol Biol* **272**, 301-311 (1997).
59. K. Luger, A. W. Mader, R. K. Richmond, D. F. Sargent, T. J. Richmond, Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**, 251-260 (1997).
60. T. H. Eickbush, E. N. Moudrianakis, The compaction of DNA helices into either continuous supercoils or folded-fiber rods and toroids. *Cell* **13**, 295-306 (1978).

61. G. Li, D. Reinberg, Chromatin higher-order structures and gene regulation. *Curr Opin Genet Dev* **21**, 175-186 (2011).
62. A. P. Wolffe, J. J. Hayes, Chromatin disruption and modification. *Nucleic acids research* **27**, 711-720 (1999).
63. P. A. Grant, A tale of histone modifications. *Genome biology* **2**, REVIEWS0003 (2001).
64. M. Grunstein, Histone acetylation in chromatin structure and transcription. *Nature* **389**, 349-352 (1997).
65. C. Tse, T. Sera, A. P. Wolffe, J. C. Hansen, Disruption of higher-order folding by core histone acetylation dramatically enhances transcription of nucleosomal arrays by RNA polymerase III. *Mol Cell Biol* **18**, 4629-4638 (1998).
66. M. Shogren-Knaak *et al.*, Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science* **311**, 844-847 (2006).
67. E. P. Consortium, An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74 (2012).
68. Y. Shiio, R. N. Eisenman, Histone sumoylation is associated with transcriptional repression. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 13225-13230 (2003).
69. J. C. Tanny, H. Erdjument-Bromage, P. Tempst, C. D. Allis, Ubiquitylation of histone H2B controls RNA polymerase II transcription elongation independently of histone H3 methylation. *Genes & development* **21**, 835-847 (2007).
70. A. Shilatifard, Chromatin modifications by methylation and ubiquitination: Implications in the regulation of gene expression. *Annual Review of Biochemistry* **75**, 243-269 (2006).
71. M. A. Freitas, A. R. Sklenar, M. R. Parthun, Application of mass spectrometry to the identification and quantification of histone post-translational modifications. *Journal of cellular biochemistry* **92**, 691-700 (2004).
72. A. M. Arnaudo, B. A. Garcia, Proteomic characterization of novel histone post-translational modifications. *Epigenetics Chromatin* **6**, 24 (2013).
73. H. Huang, S. Lin, B. A. Garcia, Y. Zhao, Quantitative proteomic analysis of histone modifications. *Chem Rev* **115**, 2376-2418 (2015).
74. S. B. Rothbart, B. D. Strahl, Interpreting the language of histone and DNA modifications. *Biochim Biophys Acta* **1839**, 627-643 (2014).
75. H. Santos-Rosa *et al.*, Active genes are tri-methylated at K4 of histone H3. *Nature* **419**, 407-411 (2002).
76. A. J. Bannister *et al.*, Spatial distribution of di- and tri-methyl lysine 36 of histone H3 at active genes. *J Biol Chem* **280**, 17732-17736 (2005).
77. B. E. Bernstein *et al.*, Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell* **120**, 169-181 (2005).
78. A. Barski *et al.*, High-resolution profiling of histone methylations in the human genome. *Cell* **129**, 823-837 (2007).
79. N. D. Heintzman *et al.*, Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459**, 108-112 (2009).
80. A. Visel *et al.*, ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* **457**, 854-858 (2009).

81. R. Cao *et al.*, Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science* **298**, 1039-1043 (2002).
82. A. H. Peters *et al.*, Histone H3 lysine 9 methylation is an epigenetic imprint of facultative heterochromatin. *Nat Genet* **30**, 77-80 (2002).
83. J. C. Rice *et al.*, Histone methyltransferases direct different degrees of methylation to define distinct chromatin domains. *Molecular cell* **12**, 1591-1598 (2003).
84. A. J. Bannister, T. Kouzarides, Regulation of chromatin by histone modifications. *Cell Res* **21**, 381-395 (2011).
85. S. L. Berger, The complex language of chromatin regulation during transcription. *Nature* **447**, 407-412 (2007).
86. T. Kouzarides, Chromatin modifications and their function. *Cell* **128**, 693-705 (2007).
87. C. R. Vakoc, S. A. Mandat, B. A. Olenchock, G. A. Blobel, Histone H3 lysine 9 methylation and HP1gamma are associated with transcription elongation through mammalian chromatin. *Molecular cell* **19**, 381-391 (2005).
88. J. H. Martens *et al.*, The profile of repeat-associated histone lysine methylation states in the mouse epigenome. *The EMBO journal* **24**, 800-812 (2005).
89. C. R. Vakoc, M. M. Sachdeva, H. Wang, G. A. Blobel, Profile of histone lysine methylation across transcribed mammalian chromatin. *Mol Cell Biol* **26**, 9185-9195 (2006).
90. B. E. Bernstein *et al.*, A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**, 315-326 (2006).
91. T. S. Mikkelsen *et al.*, Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**, 553-560 (2007).
92. P. J. Rugg-Gunn, B. J. Cox, A. Ralston, J. Rossant, 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* **107**, 10783-10790 (2010).
93. O. Alder *et al.*, Ring1B and Suv39h1 delineate distinct chromatin states at bivalent genes during early mouse lineage commitment. *Development* **137**, 2483-2492 (2010).
94. P. Voigt *et al.*, Asymmetrically modified nucleosomes. *Cell* **151**, 181-193 (2012).
95. B. J. Lesch, G. A. Dokshin, R. A. Young, J. R. McCarrey, D. C. Page, A set of genes critical to development is epigenetically poised in mouse germ cells from fetal stages through completion of meiosis. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 16061-16066 (2013).
96. B. J. Lesch, S. J. Silber, J. R. McCarrey, D. C. Page, Parallel evolution of male germline epigenetic poising and somatic development in animals. *Nat Genet* **48**, 888-894 (2016).
97. M. P. Creyghton *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* **107**, 21931-21936 (2010).
98. A. Rada-Iglesias *et al.*, A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **470**, 279-283 (2011).
99. G. E. Zentner, P. J. Tesar, P. C. Scacheri, Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions. *Genome research* **21**, 1273-1283 (2011).
100. S. Heinz, C. E. Romanoski, C. Benner, C. K. Glass, The selection and function of cell type-specific enhancers. *Nat Rev Mol Cell Biol* **16**, 144-154 (2015).

- 101.Y. Huang, J. Fang, M. T. Bedford, Y. Zhang, R. M. Xu, Recognition of histone H3 lysine-4 methylation by the double tudor domain of JMJD2A. *Science* **312**, 748-751 (2006).
- 102.H. Li *et al.*, Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF. *Nature* **442**, 91-95 (2006).
- 103.P. V. Pena *et al.*, Molecular mechanism of histone H3K4me3 recognition by plant homeodomain of ING2. *Nature* **442**, 100-103 (2006).
- 104.B. D. Strahl, C. D. Allis, The language of covalent histone modifications. *Nature* **403**, 41-45 (2000).
- 105.J. E. Brownell *et al.*, Tetrahymena histone acetyltransferase A: a homolog to yeast Gcn5p linking histone acetylation to gene activation. *Cell* **84**, 843-851 (1996).
- 106.J. Taunton, C. A. Hassig, S. L. Schreiber, A mammalian histone deacetylase related to the yeast transcriptional regulator Rpd3p. *Science* **272**, 408-411 (1996).
- 107.R. J. Sims, 3rd, K. Nishioka, D. Reinberg, Histone lysine methylation: a signature for chromatin function. *Trends Genet* **19**, 629-639 (2003).
- 108.B. E. Bernstein *et al.*, The NIH Roadmap Epigenomics Mapping Consortium. *Nat Biotechnol* **28**, 1045-1048 (2010).
- 109.X. B. Shi *et al.*, ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression. *Nature* **442**, 96-99 (2006).
- 110.G. G. Wozniak, B. D. Strahl, Hitting the 'mark': Interpreting lysine methylation in the context of active transcription. *Biochimica Et Biophysica Acta-Gene Regulatory Mechanisms* **1839**, 1353-1361 (2014).
- 111.G. E. Zentner, S. Henikoff, Regulation of nucleosome dynamics by histone modifications. *Nature structural & molecular biology* **20**, 259-266 (2013).
- 112.C. Nislow, E. Ray, L. Pillus, Set1, a yeast member of the trithorax family, functions in transcriptional silencing and diverse cellular processes. *Molecular biology of the cell* **8**, 2421-2436 (1997).
- 113.T. Miller *et al.*, COMPASS: a complex of proteins associated with a trithorax-related SET domain protein. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 12902-12907 (2001).
- 114.N. J. Krogan *et al.*, COMPASS, a histone H3 (Lysine 4) methyltransferase required for telomeric silencing of gene expression. *J Biol Chem* **277**, 10753-10755 (2002).
- 115.A. Shilatifard, The COMPASS Family of Histone H3K4 Methylases: Mechanisms of Regulation in Development and Disease Pathogenesis. *Annual Review of Biochemistry*, Vol 81 **81**, 65-95 (2012).
- 116.P. Ingham, R. Whittle, Trithorax - a New Homoeotic Mutation of Drosophila-Melanogaster Causing Transformations of Abdominal and Thoracic Imaginal Segments .1. Putative Role during Embryogenesis. *Molecular & General Genetics* **179**, 607-614 (1980).
- 117.D. Y. Benjamin, J. L. Hess, S. E. Horning, G. A. Brown, S. J. Korsmeyer, Altered Hox expression and segmental identity in Mll-mutant mice. *Nature* **378**, 505-508 (1995).
- 118.S. Glaser *et al.*, Multiple epigenetic maintenance factors implicated by the loss of Mll2 in mouse development. *Development* **133**, 1423-1432 (2006).
- 119.Y. Dou *et al.*, Regulation of MLL1 H3K4 methyltransferase activity by its core components. *Nature structural & molecular biology* **13**, 713-719 (2006).

- 120.P. M. Dehe *et al.*, Protein interactions within the Set1 complex and their roles in the regulation of histone 3 lysine 4 methylation. *J Biol Chem* **281**, 35404-35412 (2006).
- 121.A. Roguev *et al.*, The *Saccharomyces cerevisiae* Set1 complex includes an Ash2 homologue and methylates histone 3 lysine 4. *The EMBO journal* **20**, 7137-7148 (2001).
- 122.L. A. Boyer *et al.*, Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* **441**, 349-353 (2006).
- 123.J. A. Simon, R. E. Kingston, Mechanisms of polycomb gene silencing: knowns and unknowns. *Nat Rev Mol Cell Biol* **10**, 697-708 (2009).
- 124.D. Beuchle, G. Struhl, J. Muller, Polycomb group proteins and heritable silencing of *Drosophila* Hox genes. *Development* **128**, 993-1004 (2001).
- 125.J. Muller, P. Verrijzer, Biochemical mechanisms of gene regulation by polycomb group protein complexes. *Curr Opin Genet Dev* **19**, 150-158 (2009).
- 126.R. Margueron *et al.*, Ezh1 and Ezh2 Maintain Repressive Chromatin through Different Mechanisms. *Molecular cell* **32**, 503-518 (2008).
- 127.R. Margueron, D. Reinberg, The Polycomb complex PRC2 and its mark in life. *Nature* **469**, 343-349 (2011).
- 128.M. Ku *et al.*, Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. *PLoS genetics* **4**, e1000242 (2008).
- 129.R. Margueron *et al.*, Role of the polycomb protein EED in the propagation of repressive histone marks. *Nature* **461**, 762-767 (2009).
- 130.K. H. Hansen *et al.*, A model for transmission of the H3K27me3 epigenetic mark. *Nature cell biology* **10**, 1291-1300 (2008).
- 131.C. Xu *et al.*, Binding of different histone marks differentially regulates the activity and specificity of polycomb repressive complex 2 (PRC2). *Proceedings of the National Academy of Sciences of the United States of America* **107**, 19266-19271 (2010).
- 132.W. Yuan *et al.*, H3K36 methylation antagonizes PRC2-mediated H3K27 methylation. *J Biol Chem* **286**, 7983-7989 (2011).
- 133.S. F. Pinter *et al.*, Spreading of X chromosome inactivation via a hierarchy of defined Polycomb stations. *Genome research* **22**, 1864-1876 (2012).
- 134.S. L. Hostikka, J. Gong, E. M. Carpenter, Axial and appendicular skeletal transformations, ligament alterations, and motor neuron loss in *Hoxc10* mutants. *Int J Biol Sci* **5**, 397-410 (2009).
- 135.R. Terranova, H. Agherbi, A. Boned, S. Meresse, M. Djabali, Histone and DNA methylation defects at Hox genes in mice expressing a SET domain-truncated form of Mll. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 6629-6634 (2006).
- 136.A. H. Peters *et al.*, Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability. *Cell* **107**, 323-337 (2001).
- 137.M. Garcia-Cao, R. O'Sullivan, A. H. Peters, T. Jenuwein, M. A. Blasco, Epigenetic regulation of telomere length in mammalian cells by the Suv39h1 and Suv39h2 histone methyltransferases. *Nat Genet* **36**, 94-99 (2004).
- 138.D. O'Carroll *et al.*, The Polycomb-group gene Ezh2 is required for early mouse development. *Molecular and Cellular Biology* **21**, 4330-4336 (2001).

- 139.C. Faust, A. Schumacher, B. Holdener, T. Magnuson, The eed mutation disrupts anterior mesoderm production in mice. *Development* **121**, 273-285 (1995).
- 140.D. Pasini, A. P. Bracken, M. R. Jensen, E. L. Denchi, K. Helin, Suz12 is essential for mouse development and for EZH2 histone methyltransferase activity. *Embo Journal* **23**, 4061-4071 (2004).
- 141.Y. Tsukada *et al.*, Histone demethylation by a family of JmjC domain-containing proteins. *Nature* **439**, 811-816 (2006).
- 142.J. R. Whetstine *et al.*, Reversal of histone lysine trimethylation by the JMJD2 family of histone demethylases. *Cell* **125**, 467-481 (2006).
- 143.Y. Shi *et al.*, Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* **119**, 941-953 (2004).
- 144.J. M. Burg, J. J. Gonzalez, K. R. Maksimchuk, D. G. McCafferty, Lysine-Specific Demethylase 1A (KDM1A/LSD1): Product Recognition and Kinetic Analysis of Full-Length Histones. *Biochemistry* **55**, 1652-1662 (2016).
- 145.Y. Chen *et al.*, Crystal structure of human histone lysine-specific demethylase 1 (LSD1). *Proceedings of the National Academy of Sciences of the United States of America* **103**, 13956-13961 (2006).
- 146.R. Baron, N. A. Vellore, LSD1/CoREST reversible opening-closing dynamics: discovery of a nanoscale clamp for chromatin and protein binding. *Biochemistry* **51**, 3151-3153 (2012).
- 147.R. Baron, N. A. Vellore, LSD1/CoREST is an allosteric nanoscale clamp regulated by H3-tail molecular recognition. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 12509-12514 (2012).
- 148.Y. J. Shi *et al.*, Regulation of LSD1 histone demethylase activity by its associated factors. *Molecular cell* **19**, 857-864 (2005).
- 149.Y. Lin *et al.*, The SNAG domain of Snail1 functions as a molecular hook for recruiting lysine-specific demethylase 1. *The EMBO journal* **29**, 1803-1816 (2010).
- 150.F. Forneris, C. Binda, A. Adamo, E. Battaglioli, A. Mattevi, Structural basis of LSD1-CoREST selectivity in histone H3 recognition. *J Biol Chem* **282**, 20070-20074 (2007).
- 151.F. Forneris *et al.*, A highly specific mechanism of histone H3-K4 recognition by histone demethylase LSD1. *J Biol Chem* **281**, 35289-35295 (2006).
- 152.T. Nakamura *et al.*, ALL-1 is a histone methyltransferase that assembles a supercomplex of proteins involved in transcriptional regulation. *Molecular cell* **10**, 1119-1128 (2002).
- 153.M. G. Lee, C. Wynder, N. Cooch, R. Shiekhattar, An essential role for CoREST in nucleosomal histone 3 lysine 4 demethylation. *Nature* **437**, 432-435 (2005).
- 154.E. Metzger *et al.*, LSD1 demethylates repressive histone marks to promote androgen-receptor-dependent transcription. *Nature* **437**, 436-439 (2005).
- 155.J. Wang *et al.*, The lysine demethylase LSD1 (KDM1) is required for maintenance of global DNA methylation. *Nat Genet* **41**, 125-129 (2009).
- 156.S. A. Miller, A. C. Huang, M. M. Miazgowicz, M. M. Brassil, A. S. Weinmann, Coordinated but physically separable interaction with H3K27-demethylase and H3K4-methyltransferase activities are required for T-box protein-mediated activation of developmental gene expression. *Genes & development* **22**, 2980-2993 (2008).

- 157.S. A. Miller, S. E. Mohn, A. S. Weinmann, Jmjd3 and UTX play a demethylase-independent role in chromatin remodeling to regulate T-box family member-dependent gene expression. *Molecular cell* **40**, 594-605 (2010).
- 158.R. Andersson *et al.*, An atlas of active enhancers across human cell types and tissues. *Nature* **507**, 455-461 (2014).
- 159.R. D. Hotchkiss, The quantitative separation of purines, pyrimidines, and nucleosides by paper chromatography. *J Biol Chem* **175**, 315-332 (1948).
- 160.D. N. Cooper, M. Krawczak, Cytosine methylation and the fate of CpG dinucleotides in vertebrate genomes. *Hum Genet* **83**, 181-188 (1989).
- 161.M. M. Suzuki, A. Bird, DNA methylation landscapes: provocative insights from epigenomics. *Nat Rev Genet* **9**, 465-476 (2008).
- 162.A. P. Bird, CpG-rich islands and the function of DNA methylation. *Nature* **321**, 209-213 (1986).
- 163.M. Gardiner-Garden, M. Frommer, CpG islands in vertebrate genomes. *J Mol Biol* **196**, 261-282 (1987).
- 164.S. Tweedie, J. Charlton, V. Clark, A. Bird, Methylation of genomes and genes at the invertebrate-vertebrate boundary. *Molecular and Cellular Biology* **17**, 1469-1475 (1997).
- 165.M. C. Lorincz, D. R. Dickerson, M. Schmitt, M. Groudine, Intragenic DNA methylation alters chromatin structure and elongation efficiency in mammalian cells. *Nature structural & molecular biology* **11**, 1068-1075 (2004).
- 166.A. M. Deaton *et al.*, Cell type-specific DNA methylation at intragenic CpG islands in the immune system. *Genome research* **21**, 1074-1086 (2011).
- 167.S. Saxonov, P. Berg, D. L. Brutlag, A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 1412-1417 (2006).
- 168.A. Bird, DNA methylation patterns and epigenetic memory. *Genes & development* **16**, 6-21 (2002).
- 169.L. Shen *et al.*, Genome-wide profiling of DNA methylation reveals a class of normally methylated CpG island promoters. *PLoS genetics* **3**, 2023-2036 (2007).
- 170.M. S. Bartolomei, A. C. Ferguson-Smith, Mammalian genomic imprinting. *Cold Spring Harb Perspect Biol* **3**, a002592 (2011).
- 171.X. Nan *et al.*, Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature* **393**, 386-389 (1998).
- 172.P. L. Jones *et al.*, Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nat Genet* **19**, 187-191 (1998).
- 173.H. H. Ng, A. Bird, DNA methylation and chromatin modification. *Curr Opin Genet Dev* **9**, 158-163 (1999).
- 174.F. Fuks *et al.*, The methyl-CpG-binding protein MeCP2 links DNA methylation to histone methylation. *J Biol Chem* **278**, 4035-4040 (2003).
- 175.S. A. Sarraf, I. Stancheva, Methyl-CpG binding protein MBD1 couples histone H3 methylation at lysine 9 by SETDB1 to DNA replication and chromatin assembly. *Molecular cell* **15**, 595-605 (2004).
- 176.A. Hellman, A. Chess, Gene body-specific methylation on the active X chromosome. *Science* **315**, 1141-1143 (2007).

- 177.A. M. Deaton, A. Bird, CpG islands and the regulation of transcription. *Genes & development* **25**, 1010-1022 (2011).
- 178.H. Wu *et al.*, Dnmt3a-dependent nonpromoter DNA methylation facilitates transcription of neurogenic genes. *Science* **329**, 444-448 (2010).
- 179.A. K. Maunakea, I. Chepelev, K. Cui, K. Zhao, Intragenic DNA methylation modulates alternative splicing by recruiting MeCP2 to promote exon recognition. *Cell Res* **23**, 1256-1269 (2013).
- 180.A. Groth, W. Rocha, A. Verreault, G. Almouzni, Chromatin challenges during DNA replication and repair. *Cell* **128**, 721-733 (2007).
- 181.M. Bostick *et al.*, UHRF1 plays a role in maintaining DNA methylation in mammalian cells. *Science* **317**, 1760-1764 (2007).
- 182.X. Liu *et al.*, UHRF1 targets DNMT1 for DNA methylation through cooperative binding of hemi-methylated DNA and methylated H3K9. *Nature communications* **4**, 1563 (2013).
- 183.E. Li, T. H. Bestor, R. Jaenisch, Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69**, 915-926 (1992).
- 184.M. Kaneda *et al.*, Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. *Nature* **429**, 900-903 (2004).
- 185.Y. Kato *et al.*, Role of the Dnmt3 family in de novo methylation of imprinted and repetitive sequences during male germ cell development in the mouse. *Hum Mol Genet* **16**, 2272-2280 (2007).
- 186.D. Bourc'his, G. L. Xu, C. S. Lin, B. Bollman, T. H. Bestor, Dnmt3L and the establishment of maternal genomic imprints. *Science* **294**, 2536-2539 (2001).
- 187.D. Bourc'his, T. H. Bestor, Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature* **431**, 96-99 (2004).
- 188.W. Reik, W. Dean, J. Walter, Epigenetic reprogramming in mammalian development. *Science* **293**, 1089-1093 (2001).
- 189.A. C. Lossie *et al.*, Distinct phenotypes distinguish the molecular classes of Angelman syndrome. *J Med Genet* **38**, 834-845 (2001).
- 190.S. B. Cassidy, D. J. Driscoll, Prader-Willi syndrome. *European Journal of Human Genetics* **17**, 3-13 (2009).
- 191.J. B. Beckwith, Extreme cytomegaly of the adrenal fetal cortex, omphalocele, hyperplasia of kidneys and pancreas, and Leydig-cell hyperplasia: Another syndrome. *Western Society for Pediatric Research* **11**, (1963).
- 192.H. Wiedemann, Complex malformatif familial avec hernie ombilicale et macroglossie. *Un "syndrome nouveau".*
- 193.H. K. Silver, W. Kiyasu, J. George, W. C. Deamer, Syndrome of congenital hemihypertrophy, shortness of stature, and elevated urinary gonadotropins. *Pediatrics* **12**, 368-376 (1953).
- 194.D. Jia, R. Z. Jurkowska, X. Zhang, A. Jeltsch, X. Cheng, Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. *Nature* **449**, 248-251 (2007).
- 195.D. N. Ciccone *et al.*, KDM1B is a histone H3K4 demethylase required to establish maternal genomic imprints. *Nature* **461**, 415-U115 (2009).
- 196.F. Neri *et al.*, Dnmt3L antagonizes DNA methylation at bivalent promoters and favors DNA methylation at gene bodies in ESCs. *Cell* **155**, 121-134 (2013).

- 197.F. Fuks, P. J. Hurd, R. Deplus, T. Kouzarides, The DNA methyltransferases associate with HP1 and the SUV39H1 histone methyltransferase. *Nucleic acids research* **31**, 2305-2312 (2003).
- 198.S. Epsztajn-Litman *et al.*, De novo DNA methylation promoted by G9a prevents reprogramming of embryonically silenced genes. *Nature structural & molecular biology* **15**, 1176-1183 (2008).
- 199.B. Lehnertz *et al.*, The methyltransferase G9a regulates HoxA9-dependent transcription in AML. *Genes & development* **28**, 317-327 (2014).
- 200.M. Salditt-Georgieff, J. E. Darnell, Jr., Further evidence that the majority of primary nuclear RNA transcripts in mammalian cells do not contribute to mRNA. *Mol. Cell. Biol.* **2**, 701-707 (1982).
- 201.R. W. Carthew, E. J. Sontheimer, Origins and Mechanisms of miRNAs and siRNAs. *Cell* **136**, 642-655 (2009).
- 202.J. L. Rinn, H. Y. Chang, Genome regulation by long noncoding RNAs. *Annu Rev Biochem* **81**, 145-166 (2012).
- 203.P. Carninci *et al.*, The transcriptional landscape of the mammalian genome. *Science* **309**, 1559-1563 (2005).
- 204.P. Kapranov *et al.*, RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science* **316**, 1484-1488 (2007).
- 205.S. Djebali *et al.*, Landscape of transcription in human cells. *Nature* **489**, 101-108 (2012).
- 206.M. K. Iyer *et al.*, The landscape of long noncoding RNAs in the human transcriptome. *Nat Genet* **47**, 199-208 (2015).
- 207.A. Fire *et al.*, Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391**, 806-811 (1998).
- 208.R. C. Lee, R. L. Feinbaum, V. Ambros, The *C. elegans* heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* **75**, 843-854 (1993).
- 209.H. Peng *et al.*, A novel class of tRNA-derived small RNAs extremely enriched in mature mouse sperm. *Cell Res* **22**, 1609-1612 (2012).
- 210.T. Derrien *et al.*, The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression. *Genome research* **22**, 1775-1789 (2012).
- 211.D. P. Bartel, MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**, 281-297 (2004).
- 212.R. J. Taft *et al.*, Small RNAs derived from snoRNAs. *RNA* **15**, 1233-1240 (2009).
- 213.V. N. Kim, J. Han, M. C. Siomi, Biogenesis of small RNAs in animals. *Nat Rev Mol Cell Biol* **10**, 126-139 (2009).
- 214.W. R. Jeck, N. E. Sharpless, Detecting and characterizing circular RNAs. *Nat Biotechnol* **32**, 453-461 (2014).
- 215.P. Grote, B. G. Herrmann, The long non-coding RNA Fendrr links epigenetic control mechanisms to gene regulatory networks in mammalian embryogenesis. *RNA biology* **10**, 1579-1585 (2013).
- 216.C. E. Senner *et al.*, Disruption of a conserved region of Xist exon 1 impairs Xist RNA localisation and X-linked gene silencing during random and imprinted X chromosome inactivation. *Development* **138**, 1541-1550 (2011).

- 217.J. Feng *et al.*, The Evf-2 noncoding RNA is transcribed from the Dlx-5/6 ultraconserved region and functions as a Dlx-2 transcriptional coactivator. *Genes & development* **20**, 1470-1484 (2006).
- 218.M. Guttman, J. L. Rinn, Modular regulatory principles of large non-coding RNAs. *Nature* **482**, 339-346 (2012).
- 219.J. A. Martens, P. Y. Wu, F. Winston, Regulation of an intergenic transcript controls adjacent gene transcription in *Saccharomyces cerevisiae*. *Genes & development* **19**, 2695-2704 (2005).
- 220.H. Ishizu, H. Siomi, M. C. Siomi, Biology of PIWI-interacting RNAs: new insights into biogenesis and function inside and outside of germlines. *Genes & development* **26**, 2361-2373 (2012).
- 221.G. Meister, T. Tuschl, Mechanisms of gene silencing by double-stranded RNA. *Nature* **431**, 343-349 (2004).
- 222.S. M. Hammond, E. Bernstein, D. Beach, G. J. Hannon, An RNA-directed nuclease mediates post-transcriptional gene silencing in *Drosophila* cells. *Nature* **404**, 293-296 (2000).
- 223.M. A. Carmell *et al.*, MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline. *Developmental cell* **12**, 503-514 (2007).
- 224.L. Pantano *et al.*, The small RNA content of human sperm reveals pseudogene-derived piRNAs complementary to protein-coding genes. *RNA* **21**, 1085-1095 (2015).
- 225.A. Kanhere *et al.*, Short RNAs are transcribed from repressed polycomb target genes and interact with polycomb repressive complex-2. *Molecular cell* **38**, 675-688 (2010).
- 226.D. H. Kim, P. Saetrom, O. Snove, Jr., J. J. Rossi, MicroRNA-directed transcriptional gene silencing in mammalian cells. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 16230-16235 (2008).
- 227.Y. Tay, J. Zhang, A. M. Thomson, B. Lim, I. Rigoutsos, MicroRNAs to Nanog, Oct4 and Sox2 coding regions modulate embryonic stem cell differentiation. *Nature* **455**, 1124-1128 (2008).
- 228.B. P. Lewis, C. B. Burge, D. P. Bartel, Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120**, 15-20 (2005).
- 229.H. Dweep, N. Gretz, miRWalk2.0: a comprehensive atlas of microRNA-target interactions. *Nat Methods* **12**, 697 (2015).
- 230.J. Zhao, B. K. Sun, J. A. Erwin, J. J. Song, J. T. Lee, Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* **322**, 750-756 (2008).
- 231.M. Fabbri *et al.*, MicroRNA-29 family reverts aberrant methylation in lung cancer by targeting DNA methyltransferases 3A and 3B. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 15805-15810 (2007).
- 232.R. Garzon *et al.*, MicroRNA-29b induces global DNA hypomethylation and tumor suppressor gene reexpression in acute myeloid leukemia by targeting directly DNMT3A and 3B and indirectly DNMT1. *Blood* **113**, 6411-6418 (2009).
- 233.R. R. Pandey *et al.*, Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. *Molecular cell* **32**, 232-246 (2008).
- 234.A. Wagschal *et al.*, G9a histone methyltransferase contributes to imprinting in the mouse placenta. *Mol Cell Biol* **28**, 1104-1113 (2008).

- 235.A. Wutz, Gene silencing in X-chromosome inactivation: advances in understanding facultative heterochromatin formation. *Nat Rev Genet* **12**, 542-553 (2011).
- 236.F. C. Beckedorff *et al.*, The intronic long noncoding RNA ANRASSF1 recruits PRC2 to the RASSF1A promoter, reducing the expression of RASSF1A and increasing cell proliferation. *PLoS genetics* **9**, e1003705 (2013).
- 237.J. L. Rinn *et al.*, Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* **129**, 1311-1323 (2007).
- 238.R. A. Gupta *et al.*, Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* **464**, 1071-1076 (2010).
- 239.M. C. Tsai *et al.*, Long noncoding RNA as modular scaffold of histone modification complexes. *Science* **329**, 689-693 (2010).
- 240.S. Y. Ng, G. K. Bogu, B. S. Soh, L. W. Stanton, The Long Noncoding RNA RMST Interacts with SOX2 to Regulate Neurogenesis. *Molecular cell* **51**, 349-359 (2013).
- 241.M. Guttman *et al.*, lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature* **477**, 295-300 (2011).
- 242.N. D. Heintzman *et al.*, Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet* **39**, 311-318 (2007).
- 243.T. K. Kim *et al.*, Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465**, 182-187 (2010).
- 244.D. Wang *et al.*, Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. *Nature* **474**, 390-394 (2011).
- 245.K. W. Vance *et al.*, The long non-coding RNA Paupar regulates the expression of both local and distal genes. *The EMBO journal* **33**, 296-311 (2014).
- 246.C. A. Melo *et al.*, eRNAs are required for p53-dependent enhancer activity and gene transcription. *Molecular cell* **49**, 524-535 (2013).
- 247.L. Pnueli, S. Rudnizky, Y. Yosefzon, P. Melamed, RNA transcribed from a distal enhancer is required for activating the chromatin at the promoter of the gonadotropin alpha-subunit gene. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 4369-4374 (2015).
- 248.K. Mousavi *et al.*, eRNAs promote transcription by establishing chromatin accessibility at defined genomic loci. *Molecular cell* **51**, 606-617 (2013).
- 249.S. Sati, S. Ghosh, V. Jain, V. Scaria, S. Sengupta, Genome-wide analysis reveals distinct patterns of epigenetic features in long non-coding RNA loci. *Nucleic acids research* **40**, 10018-10031 (2012).
- 250.M. Sauvageau *et al.*, Multiple knockout mouse models reveal lincRNAs are required for life and brain development. *Elife* **2**, e01749 (2013).
- 251.Y. H. Jung *et al.*, Chromatin States in Mouse Sperm Correlate with Embryonic and Adult Regulatory Landscapes. *Cell Rep* **18**, 1366-1382 (2017).
- 252.P. Grote *et al.*, The tissue-specific lncRNA Fendrr is an essential regulator of heart and body wall development in the mouse. *Developmental cell* **24**, 206-214 (2013).
- 253.E. Heard, R. A. Martienssen, Transgenerational epigenetic inheritance: myths and mechanisms. *Cell* **157**, 95-109 (2014).
- 254.Y. Ohinata *et al.*, Blimp1 is a critical determinant of the germ cell lineage in mice. *Nature* **436**, 207-213 (2005).

- 255.M. Saitou, B. Payer, D. O'Carroll, Y. Ohinata, M. A. Surani, Blimp1 and the emergence of the germ line during development in the mouse. *Cell Cycle* **4**, 1736-1740 (2005).
- 256.Y. Ohinata *et al.*, A signaling principle for the specification of the germ cell lineage in mice. *Cell* **137**, 571-584 (2009).
- 257.M. Yamaji *et al.*, Critical function of Prdm14 for the establishment of the germ cell lineage in mice. *Nat Genet* **40**, 1016-1022 (2008).
- 258.S. Weber *et al.*, Critical Function of AP-2gamma/TCFAP2C in Mouse Embryonic Germ Cell Maintenance. *Biology of Reproduction* **82**, 214-223 (2010).
- 259.E. Magnusdottir *et al.*, A tripartite transcription factor network regulates primordial germ cell specification in mice. *Nature cell biology* **15**, 905-915 (2013).
- 260.W. W. Tang, T. Kobayashi, N. Irie, S. Dietmann, M. A. Surani, Specification and epigenetic programming of the human germ line. *Nat Rev Genet* **17**, 585-600 (2016).
- 261.K. A. Lawson, W. J. Hage, Clonal analysis of the origin of primordial germ cells in the mouse. *Ciba Found Symp* **182**, 68-84; discussion 84-91 (1994).
- 262.M. Monk, M. Boubelik, S. Lehnert, Temporal and regional changes in DNA methylation in the embryonic, extraembryonic and germ cell lineages during mouse embryo development. *Development* **99**, 371-382 (1987).
- 263.P. Hajkova *et al.*, Epigenetic reprogramming in mouse primordial germ cells. *Mech Dev* **117**, 15-23 (2002).
- 264.J. Lee *et al.*, Erasing genomic imprinting memory in mouse clone embryos produced from day 11.5 primordial germ cells. *Development* **129**, 1807-1817 (2002).
- 265.J. A. Hackett *et al.*, Germline DNA demethylation dynamics and imprint erasure through 5-hydroxymethylcytosine. *Science* **339**, 448-452 (2013).
- 266.S. Seisenberger *et al.*, The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells. *Molecular cell* **48**, 849-862 (2012).
- 267.S. Guibert, T. Forne, M. Weber, Global profiling of DNA methylation erasure in mouse primordial germ cells. *Genome research* **22**, 633-641 (2012).
- 268.M. Saitou, S. C. Barton, M. A. Surani, A molecular programme for the specification of germ cell fate in mice. *Nature* **418**, 293-300 (2002).
- 269.Y. Yabuta, K. Kurimoto, Y. Ohinata, Y. Seki, M. Saitou, Gene expression dynamics during germline specification in mice identified by quantitative single-cell gene expression profiling. *Biol Reprod* **75**, 705-716 (2006).
- 270.Y. Seki *et al.*, Extensive and orderly reprogramming of genome-wide chromatin modifications associated with specification and early development of germ cells in mice. *Developmental biology* **278**, 440-458 (2005).
- 271.P. Hajkova, Epigenetic reprogramming in the germline: towards the ground state of the epigenome. *Philos Trans R Soc Lond B Biol Sci* **366**, 2266-2273 (2011).
- 272.K. Ancelin *et al.*, Blimp1 associates with Prmt5 and directs histone arginine methylation in mouse germ cells. *Nature cell biology* **8**, 623-630 (2006).
- 273.M. Yamaji *et al.*, PRDM14 Ensures Naive Pluripotency through Dual Regulation of Signaling and Epigenetic Pathways in Mouse Embryonic Stem Cells. *Cell Stem Cell* **12**, 368-382 (2013).

- 274.M. Tachibana *et al.*, Histone methyltransferases G9a and GLP form heteromeric complexes and are both crucial for methylation of euchromatin at H3-K9. *Genes & development* **19**, 815-826 (2005).
- 275.Y. Seki *et al.*, Cellular dynamics associated with the genome-wide epigenetic reprogramming in migrating primordial germ cells in mice. *Development* **134**, 2627-2638 (2007).
- 276.Y. S. Chan *et al.*, A PRC2-dependent repressive role of PRDM14 in human embryonic stem cells and induced pluripotent stem cell reprogramming. *Stem Cells* **31**, 682-692 (2013).
- 277.K. Kurimoto, M. Saitou, in *Cold Spring Harbor symposia on quantitative biology*. (Cold Spring Harbor Laboratory Press, 2015), vol. 80, pp. 147-154.
- 278.P. Hajkova *et al.*, Genome-wide reprogramming in the mouse germ line entails the base excision repair pathway. *Science* **329**, 78-82 (2010).
- 279.W. W. Tang *et al.*, A Unique Gene Regulatory Network Resets the Human Germline Epigenome for Development. *Cell* **161**, 1453-1467 (2015).
- 280.T. L. Davis, G. J. Yang, J. R. McCarrey, M. S. Bartolomei, The H19 methylation imprint is erased and re-established differentially on the parental alleles during male germ cell development. *Human molecular genetics* **9**, 2885-2894 (2000).
- 281.S. K. Kota, R. Feil, Epigenetic transitions in germ cell development and meiosis. *Developmental cell* **19**, 675-686 (2010).
- 282.N. Kubo *et al.*, DNA methylation and gene expression dynamics during spermatogonial stem cell differentiation in the early postnatal mouse testis. *BMC Genomics* **16**, 624 (2015).
- 283.S. A. Smallwood *et al.*, Dynamic CpG island methylation landscape in oocytes and preimplantation embryos. *Nat Genet* **43**, 811-814 (2011).
- 284.A. C. Ferguson-Smith, Genomic imprinting: the emergence of an epigenetic paradigm. *Nat Rev Genet* **12**, 565-575 (2011).
- 285.M. Chotalia *et al.*, Transcription is required for establishment of germline methylation marks at imprinted genes. *Genes & development* **23**, 105-117 (2009).
- 286.J. McGrath, D. Solter, Completion of Mouse Embryogenesis Requires Both the Maternal and Paternal Genomes. *Cell* **37**, 179-183 (1984).
- 287.M. A. Surani, S. C. Barton, M. L. Norris, Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis. *Nature* **308**, 548-550 (1984).
- 288.M. S. Bartolomei, S. Zemel, S. M. Tilghman, Parental imprinting of the mouse H19 gene. *Nature* **351**, 153-155 (1991).
- 289.T. M. DeChiara, E. J. Robertson, A. Efstratiadis, Parental imprinting of the mouse insulin-like growth factor II gene. *Cell* **64**, 849-859 (1991).
- 290.A. C. Ferguson-Smith, B. Cattanach, S. Barton, C. Beechey, M. Surani, Embryological and molecular investigations of parental imprinting on mouse chromosome 7. *Nature* **351**, 667-670 (1991).
- 291.P. A. Leighton, R. S. Ingram, J. Eggenschwiler, A. Efstratiadis, S. M. Tilghman, Disruption of imprinting caused by deletion of the H19 gene region in mice. *Nature* **375**, 34-39 (1995).
- 292.J. L. Thorvaldsen, K. L. Duran, M. S. Bartolomei, Deletion of the H19 differentially methylated domain results in loss of imprinted expression of H19 and Igf2. *Genes & development* **12**, 3693-3702 (1998).

- 293.M. Srivastava *et al.*, H19 and Igf2 monoallelic expression is regulated in two distinct ways by a shared cis acting regulatory region upstream of H19. *Genes & development* **14**, 1186-1195 (2000).
- 294.D. P. Barlow, M. S. Bartolomei, Genomic imprinting in mammals. *Cold Spring Harb Perspect Biol* **6**, (2014).
- 295.S. Kurukuti *et al.*, CTCF binding at the H19 imprinting control region mediates maternally inherited higher-order chromatin conformation to restrict enhancer access to Igf2. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 10684-10689 (2006).
- 296.J. Q. Ling *et al.*, CTCF mediates interchromosomal colocalization between Igf2/H19 and Wsb1/Nf1. *Science* **312**, 269-272 (2006).
- 297.F. Sleutels, D. P. Barlow, The origins of genomic imprinting in mammals. *Adv Genet* **46**, 119-163 (2002).
- 298.D. Mancini-Dinardo, S. J. Steele, J. M. Levorse, R. S. Ingram, S. M. Tilghman, Elongation of the Kcnq1ot1 transcript is required for genomic imprinting of neighboring genes. *Genes & development* **20**, 1268-1282 (2006).
- 299.C. I. M. Seidl, S. H. Stricker, D. P. Barlow, The imprinted Air ncRNA is an atypical RNAPII transcript that evades splicing and escapes nuclear export. *Embo Journal* **25**, 3565-3575 (2006).
- 300.P. Monnier *et al.*, H19 lncRNA controls gene expression of the Imprinted Gene Network by recruiting MBD1. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 20693-20698 (2013).
- 301.L. Williamson *et al.*, UV Irradiation Induces a Non-coding RNA that Functionally Opposes the Protein Encoded by the Same Gene. *Cell* **168**, 843-+ (2017).
- 302.T. Nagano *et al.*, The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. *Science* **322**, 1717-1720 (2008).
- 303.R. Terranova *et al.*, Polycomb group proteins Ezh2 and Rnf2 direct genomic contraction and imprinted repression in early mouse embryos. *Developmental cell* **15**, 668-679 (2008).
- 304.A. Gabory *et al.*, H19 acts as a trans regulator of the imprinted gene network controlling growth in mice. *Development* **136**, 3413-3421 (2009).
- 305.G. Borsani *et al.*, Characterization of a murine gene expressed from the inactive X chromosome. *Nature* **351**, 325-329 (1991).
- 306.N. Brockdorff *et al.*, The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. *Cell* **71**, 515-526 (1992).
- 307.C. J. Brown *et al.*, The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. *Cell* **71**, 527-542 (1992).
- 308.G. D. Penny, G. F. Kay, S. A. Sheardown, S. Rastan, N. Brockdorff, Requirement for Xist in X chromosome inactivation. *Nature* **379**, 131-137 (1996).
- 309.Y. Marahrens, B. Panning, J. Dausman, W. Strauss, R. Jaenisch, Xist-deficient mice are defective in dosage compensation but not spermatogenesis. *Genes Dev.* **11**, 156-166 (1997).

- 310.Y. Shi *et al.*, Sharp, an inducible cofactor that integrates nuclear receptor repression and activation. *Genes & development* **15**, 1140-1151 (2001).
- 311.C. A. McHugh *et al.*, The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature* **521**, 232-236 (2015).
- 312.E. M. Riising *et al.*, Gene silencing triggers polycomb repressive complex 2 recruitment to CpG islands genome wide. *Molecular cell* **55**, 347-360 (2014).
- 313.K. Plath *et al.*, Role of histone H3 lysine 27 methylation in X inactivation. *Science* **300**, 131-135 (2003).
- 314.Y. Hasegawa *et al.*, The matrix protein hnRNP U is required for chromosomal localization of Xist RNA. *Developmental cell* **19**, 469-476 (2010).
- 315.A. Inoue, L. Jiang, F. Lu, T. Suzuki, Y. Zhang, Maternal H3K27me3 controls DNA methylation-independent imprinting. *Nature* **547**, 419-424 (2017).
- 316.C. P. Leblond, Y. Clermont, Definition of the stages of the cycle of the seminiferous epithelium in the rat. *Ann N Y Acad Sci* **55**, 548-573 (1952).
- 317.C. Leblond, Y. Clermont, Spermiogenesis of rat, mouse, hamster and guinea pig as revealed by the “periodic acid-fuchsin sulfurous acid” technique. *Developmental Dynamics* **90**, 167-215 (1952).
- 318.E. F. Oakberg, A Description of Spermiogenesis in the Mouse and Its Use in Analysis of the Cycle of the Seminiferous Epithelium and Germ Cell Renewal. *American Journal of Anatomy* **99**, 391-& (1956).
- 319.E. F. Oakberg, Duration of spermatogenesis in the mouse and timing of stages of the cycle of the seminiferous epithelium. *Am J Anat* **99**, 507-516 (1956).
- 320.Y. Clermont, Kinetics of spermatogenesis in mammals: seminiferous epithelium cycle and spermatogonial renewal. *Physiol Rev* **52**, 198-236 (1972).
- 321.D. J. McLean, P. J. Friel, D. S. Johnston, M. D. Griswold, Characterization of spermatogonial stem cell maturation and differentiation in neonatal mice. *Biol Reprod* **69**, 2085-2091 (2003).
- 322.J. M. Oatley, R. L. Brinster, The germline stem cell niche unit in mammalian testes. *Physiol Rev* **92**, 577-595 (2012).
- 323.B. P. Hermann *et al.*, Transcriptional and translational heterogeneity among neonatal mouse spermatogonia. *Biol Reprod* **92**, 54 (2015).
- 324.H. Kubota, M. R. Avarbock, R. L. Brinster, Culture conditions and single growth factors affect fate determination of mouse spermatogonial stem cells. *Biology of reproduction* **71**, 722-731 (2004).
- 325.F. Chan *et al.*, Functional and molecular features of the Id4+ germline stem cell population in mouse testes. *Genes & development* **28**, 1351-1362 (2014).
- 326.K. Mutoji *et al.*, TSPAN8 expression distinguishes spermatogonial stem cells in the prepubertal mouse testis. *Biology of reproduction* **95**, 117, 111-114 (2016).
- 327.M. Oulad-Abdelghani *et al.*, Characterization of a premeiotic germ cell-specific cytoplasmic protein encoded by Stra8, a novel retinoic acid-responsive gene. *J Cell Biol* **135**, 469-477 (1996).
- 328.A. E. Baltus *et al.*, In germ cells of mouse embryonic ovaries, the decision to enter meiosis precedes premeiotic DNA replication. *Nat Genet* **38**, 1430-1434 (2006).

- 329.E. L. Anderson *et al.*, Stra8 and its inducer, retinoic acid, regulate meiotic initiation in both spermatogenesis and oogenesis in mice. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 14976-14980 (2008).
- 330.Y. F. Lin, M. E. Gill, J. Koubova, D. C. Page, Germ Cell-Intrinsic and -Extrinsic Factors Govern Meiotic Initiation in Mouse Embryos. *Science* **322**, 1685-1687 (2008).
- 331.C. A. Hogarth, M. D. Griswold, The key role of vitamin A in spermatogenesis. *J Clin Invest* **120**, 956-962 (2010).
- 332.M. D. Griswold, Spermatogenesis: The Commitment to Meiosis. *Physiol Rev* **96**, 1-17 (2016).
- 333.S. Kimmins, P. Sassone-Corsi, Chromatin remodelling and epigenetic features of germ cells. *Nature* **434**, 583-589 (2005).
- 334.M. Godmann *et al.*, Dynamic regulation of histone H3 methylation at lysine 4 in mammalian spermatogenesis. *Biol Reprod* **77**, 754-764 (2007).
- 335.R. Lambrot, C. Lafleur, S. Kimmins, The histone demethylase KDM1A is essential for the maintenance and differentiation of spermatogonial stem cells and progenitors. *Faseb Journal* **29**, 4402-4416 (2015).
- 336.D. A. Myrick *et al.*, KDM1A/LSD1 regulates the differentiation and maintenance of spermatogonia in mice. *PLoS One* **12**, e0177473 (2017).
- 337.E. D. Parvanov, P. M. Petkov, K. Paigen, Prdm9 controls activation of mammalian recombination hotspots. *Science* **327**, 835 (2010).
- 338.F. Baudat *et al.*, PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice. *Science* **327**, 836-840 (2010).
- 339.N. R. Powers *et al.*, The Meiotic Recombination Activator PRDM9 Trimethylates Both H3K36 and H3K4 at Recombination Hotspots In Vivo. *PLoS genetics* **12**, e1006146 (2016).
- 340.M. Lachner, D. O'Carroll, S. Rea, K. Mechtler, T. Jenuwein, Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* **410**, 116-120 (2001).
- 341.S. Rea *et al.*, Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* **406**, 593-599 (2000).
- 342.K. Sarma, D. Reinberg, Histone variants meet their match. *Nat Rev Mol Cell Biol* **6**, 139-149 (2005).
- 343.M. C. W. Tang *et al.*, Contribution of the two genes encoding histone variant h3.3 to viability and fertility in mice. *PLoS Genet.* **11**, e1004964 (2015).
- 344.B. T. Yuen, K. M. Bush, B. L. Barrilleaux, R. Cotterman, P. S. Knoepfler, Histone H3.3 regulates dynamic chromatin states during spermatogenesis. *Development* **141**, 3483-3494 (2014).
- 345.M. C. Tang *et al.*, Contribution of the two genes encoding histone variant h3.3 to viability and fertility in mice. *PLoS genetics* **11**, e1004964 (2015).
- 346.J. Govin *et al.*, Pericentric heterochromatin reprogramming by new histone variants during mouse spermiogenesis. *J Cell Biol* **176**, 283-294 (2007).
- 347.B. Li *et al.*, Preferential occupancy of histone variant H2AZ at inactive promoters influences local histone modifications and chromatin remodeling. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 18385-18390 (2005).
- 348.M. Boulard *et al.*, The NH₂ tail of the novel histone variant H2BFWT exhibits properties distinct from conventional H2B with respect to the assembly of mitotic chromosomes. *Mol Cell Biol* **26**, 1518-1526 (2006).

- 349.C. Alabert, A. Groth, Chromatin replication and epigenome maintenance. *Nat Rev Mol Cell Biol* **13**, 153-167 (2012).
- 350.P. B. Mason, K. Struhl, The FACT complex travels with elongating RNA polymerase II and is important for the fidelity of transcriptional initiation in vivo. *Molecular and Cellular Biology* **23**, 8323-8333 (2003).
- 351.A. Saunders *et al.*, Tracking FACT and the RNA polymerase II elongation complex through chromatin in vivo. *Science* **301**, 1094-1096 (2003).
- 352.M. F. Dion *et al.*, Dynamics of replication-independent histone turnover in budding yeast. *Science* **315**, 1405-1408 (2007).
- 353.A. Jamai, R. M. Imoberdorf, M. Strubin, Continuous histone H2B and transcription-dependent histone H3 exchange in yeast cells outside of replication. *Molecular cell* **25**, 345-355 (2007).
- 354.A. Rufiange, P. E. Jacques, W. Bhat, F. Robert, A. Nourani, Genome-wide replication-independent histone H3 exchange occurs predominantly at promoters and implicates H3 K56 acetylation and Asf1. *Molecular cell* **27**, 393-405 (2007).
- 355.R. B. Deal, J. G. Henikoff, S. Henikoff, Genome-wide kinetics of nucleosome turnover determined by metabolic labeling of histones. *Science* **328**, 1161-1164 (2010).
- 356.D. Ray-Gallet *et al.*, HIRA is critical for a nucleosome assembly pathway independent of DNA synthesis. *Molecular cell* **9**, 1091-1100 (2002).
- 357.H. Tagami, D. Ray-Gallet, G. Almouzni, Y. Nakatani, Histone H3.1 and H3.3 complexes mediate nucleosome assembly pathways dependent or independent of DNA synthesis. *Cell* **116**, 51-61 (2004).
- 358.C. Jin *et al.*, H3.3/H2A.Z double variant-containing nucleosomes mark 'nucleosome-free regions' of active promoters and other regulatory regions. *Nat Genet* **41**, 941-945 (2009).
- 359.T. Tamura *et al.*, Inducible deposition of the histone variant H3.3 in interferon-stimulated genes. *J Biol Chem* **284**, 12217-12225 (2009).
- 360.P. Hajkova *et al.*, Chromatin dynamics during epigenetic reprogramming in the mouse germ line. *Nature* **452**, 877-881 (2008).
- 361.A. D. Goldberg *et al.*, Distinct factors control histone variant H3.3 localization at specific genomic regions. *Cell* **140**, 678-691 (2010).
- 362.L. A. Banaszynski *et al.*, Hira-dependent histone H3.3 deposition facilitates PRC2 recruitment at developmental loci in ES cells. *Cell* **155**, 107-120 (2013).
- 363.R. Balhorn, A model for the structure of chromatin in mammalian sperm. *J Cell Biol* **93**, 298-305 (1982).
- 364.W. Kistler, C. Noyes, R. Hsu, R. Heinrikson, The amino acid sequence of a testis-specific basic protein that is associated with spermatogenesis. *Journal of Biological Chemistry* **250**, 1847-1853 (1975).
- 365.P. C. Yelick *et al.*, Mouse transition protein 1 is translationally regulated during the postmeiotic stages of spermatogenesis. *Molecular reproduction and development* **1**, 193-200 (1989).
- 366.J. Castillo, L. Simon, S. de Mateo, S. Lewis, R. Oliva, Protamine/DNA ratios and DNA damage in native and density gradient centrifuged sperm from infertile patients. *J Androl* **32**, 324-332 (2011).

- 367.S. Francis, S. Yelumalai, C. Jones, K. Coward, Aberrant protamine content in sperm and consequential implications for infertility treatment. *Hum Fertil (Camb)* **17**, 80-89 (2014).
- 368.S. M. Wykes, S. A. Krawetz, The structural organization of sperm chromatin. *J Biol Chem* **278**, 29471-29477 (2003).
- 369.S. H. Syed *et al.*, The incorporation of the novel histone variant H2AL2 confers unusual structural and functional properties of the nucleosome. *Nucleic acids research* **37**, 4684-4695 (2009).
- 370.E. Montellier *et al.*, Chromatin-to-nucleoprotamine transition is controlled by the histone H2B variant TH2B. *Genes & development* **27**, 1680-1692 (2013).
- 371.M. L. Meistrich, P. K. Trostleweige, R. L. Lin, Y. M. Bhatnagar, C. D. Allis, Highly Acetylated H4 Is Associated with Histone Displacement in Rat Spermatids. *Molecular Reproduction and Development* **31**, 170-181 (1992).
- 372.M. Hazzouri *et al.*, Regulated hyperacetylation of core histones during mouse spermatogenesis: involvement of histone deacetylases. *Eur J Cell Biol* **79**, 950-960 (2000).
- 373.J. Gaucher *et al.*, Bromodomain-dependent stage-specific male genome programming by Brdt. *The EMBO journal* **31**, 3809-3820 (2012).
- 374.E. Y. Shang, H. D. Nickerson, D. C. Wen, X. Y. Wang, D. J. Wolgemuth, The first bromodomain of Brdt, a testis-specific member of the BET sub-family of double-bromodomain-containing proteins, is essential for male germ cell differentiation. *Development* **134**, 3507-3515 (2007).
- 375.A. M. Khalil, F. Z. Boyar, D. J. Driscoll, Dynamic histone modifications mark sex chromosome inactivation and reactivation during mammalian spermatogenesis. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 16583-16587 (2004).
- 376.K. Hayashi, Y. Matsui, Meisetz, a novel histone tri-methyltransferase, regulates meiosis-specific epigenesis. *Cell Cycle* **5**, 615-620 (2006).
- 377.S. Rousseaux *et al.*, Establishment of male-specific epigenetic information. *Gene* **345**, 139-153 (2005).
- 378.R. Balhorn, B. L. Gledhill, A. J. Wyrobek, Mouse sperm chromatin proteins: quantitative isolation and partial characterization. *Biochemistry* **16**, 4074-4080 (1977).
- 379.M. Gardiner-Garden, M. Ballesteros, M. Gordon, P. P. Tam, Histone- and protamine-DNA association: conservation of different patterns within the beta-globin domain in human sperm. *Mol Cell Biol* **18**, 3350-3356 (1998).
- 380.S. S. Hammoud *et al.*, Distinctive chromatin in human sperm packages genes for embryo development. *Nature* **460**, 473-478 (2009).
- 381.U. Brykczynska *et al.*, Repressive and active histone methylation mark distinct promoters in human and mouse spermatozoa. *Nature structural & molecular biology* **17**, 679-687 (2010).
- 382.S. Erkek *et al.*, Molecular determinants of nucleosome retention at CpG-rich sequences in mouse spermatozoa. *Nature structural & molecular biology* **20**, 868-875 (2013).
- 383.K. Siklenka *et al.*, Disruption of histone methylation in developing sperm impairs offspring health transgenerationally. *Science* **350**, aab2006 (2015).
- 384.G. D. Johnson, M. Jodar, R. Pique-Regi, S. A. Krawetz, Nuclease Footprints in Sperm Project Past and Future Chromatin Regulatory Events. *Sci Rep* **6**, 25864 (2016).

- 385.S. Erkek *et al.*, Molecular determinants of nucleosome retention at CpG-rich sequences in mouse spermatozoa. *Nature structural & molecular biology* **20**, 1236 (2013).
- 386.O. Meikar, M. Da Ros, H. Liljenback, J. Toppari, N. Kotaja, Accumulation of piRNAs in the chromatoid bodies purified by a novel isolation protocol. *Exp Cell Res* **316**, 1567-1575 (2010).
- 387.N. Kotaja *et al.*, The chromatoid body of male germ cells: similarity with processing bodies and presence of Dicer and microRNA pathway components. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 2647-2652 (2006).
- 388.M. Nguyen Chi *et al.*, Temporally regulated traffic of HuR and its associated ARE-containing mRNAs from the chromatoid body to polysomes during mouse spermatogenesis. *PLoS One* **4**, e4900 (2009).
- 389.A. L. Arkov, A. Ramos, Building RNA-protein granules: insight from the germline. *Trends Cell Biol* **20**, 482-490 (2010).
- 390.S. Kuramochi-Miyagawa *et al.*, MVH in piRNA processing and gene silencing of retrotransposons. *Genes & development* **24**, 887-892 (2010).
- 391.N. Kotaja, P. Sassone-Corsi, The chromatoid body: a germ-cell-specific RNA-processing centre. *Nat Rev Mol Cell Biol* **8**, 85-90 (2007).
- 392.C. A. Pessot *et al.*, Presence of RNA in the sperm nucleus. *Biochem Biophys Res Commun* **158**, 272-278 (1989).
- 393.G. C. Ostermeier, D. J. Dix, D. Miller, P. Khatri, S. A. Krawetz, Spermatozoal RNA profiles of normal fertile men. *Lancet* **360**, 772-777 (2002).
- 394.Y. Zhao *et al.*, Characterization and quantification of mRNA transcripts in ejaculated spermatozoa of fertile men by serial analysis of gene expression. *Hum Reprod* **21**, 1583-1590 (2006).
- 395.E. P. Consortium *et al.*, Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**, 799-816 (2007).
- 396.C. C. Yang *et al.*, Seasonal effect on sperm messenger RNA profile of domestic swine (*Sus Scrofa*). *Anim Reprod Sci* **119**, 76-84 (2010).
- 397.B. E. Fischer *et al.*, Conserved properties of *Drosophila* and human spermatozoal mRNA repertoires. *Proc Biol Sci* **279**, 2636-2644 (2012).
- 398.F. Borges *et al.*, Comparative transcriptomics of *Arabidopsis* sperm cells. *Plant Physiol* **148**, 1168-1181 (2008).
- 399.E. Sendler *et al.*, Stability, delivery and functions of human sperm RNAs at fertilization. *Nucleic acids research* **41**, 4104-4117 (2013).
- 400.G. D. Johnson *et al.*, The sperm nucleus: chromatin, RNA, and the nuclear matrix. *Reproduction* **141**, 21-36 (2011).
- 401.R. Goodrich, G. Johnson, S. A. Krawetz, The preparation of human spermatozoal RNA for clinical analysis. *Archives of andrology* **53**, 161-167 (2007).
- 402.P. Kumar, C. Kuscu, A. Dutta, Biogenesis and Function of Transfer RNA-Related Fragments (tRFs). *Trends Biochem Sci* **41**, 679-689 (2016).
- 403.G. Sienski, D. Donertas, J. Brennecke, Transcriptional silencing of transposons by Piwi and maelstrom and its impact on chromatin state and gene expression. *Cell* **151**, 964-980 (2012).

- 404.X. A. Huang *et al.*, A major epigenetic programming mechanism guided by piRNAs. *Developmental cell* **24**, 502-516 (2013).
- 405.A. Le Thomas *et al.*, Transgenerationally inherited piRNAs trigger piRNA biogenesis by changing the chromatin of piRNA clusters and inducing precursor processing. *Genes & development* **28**, 1667-1680 (2014).
- 406.J. Brennecke *et al.*, Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. *Cell* **128**, 1089-1103 (2007).
- 407.M. C. Siomi, K. Sato, D. Pezic, A. A. Aravin, PIWI-interacting small RNAs: the vanguard of genome defence. *Nat Rev Mol Cell Biol* **12**, 246-258 (2011).
- 408.F. Mohn, G. Sienski, D. Handler, J. Brennecke, The rhino-deadlock-cutoff complex licenses noncanonical transcription of dual-strand piRNA clusters in Drosophila. *Cell* **157**, 1364-1379 (2014).
- 409.M. R. Motamedi *et al.*, Two RNAi complexes, RITS and RDRC, physically interact and localize to noncoding centromeric RNAs. *Cell* **119**, 789-802 (2004).
- 410.K. Noma *et al.*, RITS acts in cis to promote RNA interference-mediated transcriptional and post-transcriptional silencing. *Nat Genet* **36**, 1174-1180 (2004).
- 411.S. U. Colmenares, S. M. Boker, M. Buhler, M. Dlakic, D. Moazed, Coupling of double-stranded RNA synthesis and siRNA generation in fission yeast RNAi. *Molecular cell* **27**, 449-461 (2007).
- 412.D. Holoch, D. Moazed, RNA-mediated epigenetic regulation of gene expression. *Nat Rev Genet* **16**, 71-84 (2015).
- 413.V. K. Rakyan *et al.*, Transgenerational inheritance of epigenetic states at the murine Axin(Fu) allele occurs after maternal and paternal transmission. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 2538-2543 (2003).
- 414.K. K. Mejos, H. W. Kim, E. M. Lim, N. Chang, Effects of parental folate deficiency on the folate content, global DNA methylation, and expressions of FRalpha, IGF-2 and IGF-1R in the postnatal rat liver. *Nutr Res Pract* **7**, 281-286 (2013).
- 415.R. Lambrot *et al.*, Low paternal dietary folate alters the mouse sperm epigenome and is associated with negative pregnancy outcomes. *Nature communications* **4**, 2889 (2013).
- 416.L. Ly *et al.*, Intergenerational impact of paternal lifetime exposures to both folic acid deficiency and supplementation on reproductive outcomes and imprinted gene methylation. *Molecular Human Reproduction* **23**, 461-477 (2017).
- 417.M. Aarabi *et al.*, High-dose folic acid supplementation alters the human sperm methylome and is influenced by the MTHFR C677T polymorphism. *Human molecular genetics* **24**, 6301-6313 (2015).
- 418.A. Soubry *et al.*, Newborns of obese parents have altered DNA methylation patterns at imprinted genes. *Int J Obes (Lond)* **39**, 650-657 (2015).
- 419.S. F. Ng *et al.*, Chronic high-fat diet in fathers programs beta-cell dysfunction in female rat offspring. *Nature* **467**, 963-966 (2010).
- 420.E. J. Radford *et al.*, In utero effects. In utero undernourishment perturbs the adult sperm methylome and intergenerational metabolism. *Science* **345**, 1255903 (2014).
- 421.M. Manikkam, C. Guerrero-Bosagna, R. Tracey, M. M. Haque, M. K. Skinner, Transgenerational actions of environmental compounds on reproductive disease and

- identification of epigenetic biomarkers of ancestral exposures. *PLoS One* **7**, e31901 (2012).
- 422.M. K. Skinner, C. G.-B. M. Haque, E. Nilsson, R. Bhandari, J. R. McCarrey, Environmentally induced transgenerational epigenetic reprogramming of primordial germ cells and the subsequent germ line. *PLoS one* **8**, e66318 (2013).
- 423.S. F. Ng *et al.*, Chronic high-fat diet in fathers programs beta-cell dysfunction in female rat offspring. *Nature* **467**, 963-966 (2010).
- 424.B. R. Carone *et al.*, Paternally induced transgenerational environmental reprogramming of metabolic gene expression in mammals. *Cell* **143**, 1084-1096 (2010).
- 425.G. C. Ostermeier, D. Miller, J. D. Huntriss, M. P. Diamond, S. A. Krawetz, Reproductive biology: delivering spermatozoan RNA to the oocyte. *Nature* **429**, 154 (2004).
- 426.D. Ulveling, C. Francastel, F. Hube, Identification of potentially new bifunctional RNA based on genome-wide data-mining of alternative splicing events. *Biochimie* **93**, 2024-2027 (2011).
- 427.T. Hata, M. Nakayama, Targeted disruption of the murine large nuclear KIAA1440/Ints1 protein causes growth arrest in early blastocyst stage embryos and eventual apoptotic cell death. *Biochim Biophys Acta* **1773**, 1039-1051 (2007).
- 428.W. M. Liu *et al.*, Sperm-borne microRNA-34c is required for the first cleavage division in mouse. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 490-494 (2012).
- 429.M. Rassoulzadegan *et al.*, RNA-mediated non-mendelian inheritance of an epigenetic change in the mouse. *Nature* **441**, 469-474 (2006).
- 430.A. B. Rodgers, C. P. Morgan, N. A. Leu, T. L. Bale, Transgenerational epigenetic programming via sperm microRNA recapitulates effects of paternal stress. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 13699-13704 (2015).
- 431.K. D. Wagner *et al.*, RNA induction and inheritance of epigenetic cardiac hypertrophy in the mouse. *Developmental cell* **14**, 962-969 (2008).
- 432.V. Grandjean *et al.*, The miR-124-Sox9 paramutation: RNA-mediated epigenetic control of embryonic and adult growth. *Development* **136**, 3647-3655 (2009).
- 433.Q. Chen *et al.*, Sperm tsRNAs contribute to intergenerational inheritance of an acquired metabolic disorder. *Science* **351**, 397-400 (2016).
- 434.K. Gapp *et al.*, Early life stress in fathers improves behavioural flexibility in their offspring. *Nature communications* **5**, 5466 (2014).
- 435.O. Rechavi *et al.*, Starvation-induced transgenerational inheritance of small RNAs in *C. elegans*. *Cell* **158**, 277-287 (2014).
- 436.G. W. van der Heijden *et al.*, Transmission of modified nucleosomes from the mouse male germline to the zygote and subsequent remodeling of paternal chromatin. *Developmental biology* **298**, 458-469 (2006).
- 437.G. W. van der Heijden *et al.*, Sperm-derived histones contribute to zygotic chromatin in humans. *BMC Dev Biol* **8**, 34 (2008).
- 438.M. Ihara *et al.*, Paternal poly (ADP-ribose) metabolism modulates retention of inheritable sperm histones and early embryonic gene expression. *PLoS genetics* **10**, e1004317 (2014).

- 439.R. Balhorn, S. Reed, N. Tanphaichitr, Aberrant protamine 1/protamine 2 ratios in sperm of infertile human males. *Experientia* **44**, 52-55 (1988).
- 440.V. W. Aoki, B. R. Emery, L. H. Liu, D. T. Carrell, Protamine levels vary between individual sperm cells of infertile human males and correlate with viability and DNA integrity. *Journal of Andrology* **27**, 890-898 (2006).
- 441.D. Miller, M. Brinkworth, D. Iles, Paternal DNA packaging in spermatozoa: more than the sum of its parts? DNA, histones, protamines and epigenetics. *Reproduction* **139**, 287-301 (2010).
- 442.L. J. Gaydos, W. Wang, S. Strome, Gene repression. H3K27me and PRC2 transmit a memory of repression across generations and during development. *Science* **345**, 1515-1518 (2014).
- 443.R. T. Coleman, G. Struhl, Causal role for inheritance of H3K27me3 in maintaining the OFF state of a Drosophila HOX gene. *Science* **356**, (2017).
- 444.F. Zenk *et al.*, Germ line-inherited H3K27me3 restricts enhancer function during maternal-to-zygotic transition. *Science* **357**, 212-216 (2017).

Chapter 2: Disruption of histone methylation in developing sperm impairs offspring health transgenerationally

Authors: Keith Siklenka^{1*}, Serap Erkek^{2,3,10*}, Maren Godmann^{4,11}, Romain Lambrot⁴, Serge McGraw⁵, Christine Lafleur⁴, Tamara Cohen⁴, Jianguo Xia^{4,9}, Matthew Suderman⁷, Michael Hallett⁸, Jacquette Trasler^{5,6}, Antoine H.F.M. Peters^{2,3*} and Sarah Kimmins^{1,4*}.

*these authors contributed equally to this work

Affiliations:

1 Department of Pharmacology and Therapeutics¹, Faculty of Medicine, McGill University, Montreal, Canada

2 Friedrich Miescher Institute for Biomedical Research (FMI), CH-4058 Basel, Switzerland

3 Faculty of Sciences, University of Basel, Basel, Switzerland

4 Department of Animal Science, Faculty of Agricultural and Environmental Sciences, McGill University, Montreal, Canada

5 Department of Pediatrics, Faculty of Medicine, McGill University, Montreal, Canada

6 Departments of Human Genetics and Pharmacology & Therapeutics; Research Institute of the McGill University Health Centre at the Montreal Children's Hospital, Montreal, Canada

7 MRC Integrative Epidemiology Unity, School of Social and Community Medicine, University of Bristol, Bristol, United Kingdom

8 McGill Centre for Bioinformatics, School of Computer Science, Faculty of Science, McGill University, Montreal

9 Institute of Parasitology, Faculty of Agricultural and Environmental Sciences, McGill University, Montreal, Canada

10 Current address: EMBL, Meyerhofstrasse 1, 69117 Heidelberg, Germany and DKFZ, Im Neuenheimer Feld 280, 69120 Heidelberg, Germany

11 Current address: Friedrich Schiller University Jena, Center for Molecular Biomedicine

(CMB), Institute of Biochemistry and Biophysics, Department of Cell Biology, Jena, Germany

2.1 Abstract

A father's lifetime experiences can be transmitted to his offspring to affect health and development. The mechanisms underlying paternal epigenetic transmission are unclear. Unlike somatic cells, there are few nucleosomes in sperm and their function in epigenetic inheritance is unknown. We generated transgenic mice in which over-expression of the histone H3 lysine 4 (H3K4) demethylase LSD1/KDM1A during spermatogenesis reduced H3K4 dimethylation in sperm. KDM1A over-expression in one generation severely impaired development and survivability of offspring. These defects persisted transgenerationally in the absence of KDM1A germ line expression and were associated with altered RNA profiles in sperm and offspring. We show that epigenetic inheritance of aberrant development can be initiated by histone demethylase activity in developing sperm, without changes to DNA methylation at CpG-rich regions.

2.2 Introduction

Birth defects occur in three percent of babies and can be caused by genetic factors or environmental exposures with 50% being idiopathic (1, 2). The focus on prevention has largely been on the mother, despite the father contributing half the genetic information and possibly some epigenetic information to the embryo. Studies in humans and animals suggest that epigenetic mechanisms may serve in the transmission of environmentally-induced phenotypic traits from parents to offspring (3-8). Such traits have been associated with altered gene expression and tissue function in first, second and/or third offspring generations (3-8). Depending on the number of generations and parental origin, this phenomenon is termed intergenerational or transgenerational epigenetic inheritance (Fig. 1). Maternal and paternal transmission of such effects has been related to alterations of DNA methylation in germ cells (7,

9-11). DNA methylation occurs at the 5-position of cytosine residues and associates with heritable gene silencing when promoter sequences containing multiple CpG dinucleotides (CpG islands; CGIs) are methylated. Likewise, DNA methylation suppresses transcription of endogenous retrotransposable elements (12). In sperm, DNA methylation at gene promoters is uncommon, whereas repetitive elements are frequently methylated (13-15). Despite being the major focus of studies in epigenetic inheritance, the significance of DNA methylation in paternal epigenetic inheritance is unresolved. Most studies reported only minor changes in DNA methylation in sperm at CpG enriched regions that have been associated with transmission of environmentally induced traits (5, 7, 8, 10, 11). Paternal transgenerational epigenetic inheritance may therefore involve changes in histone states (16-18), RNA (19), and/or other sperm components.

During the final stages of sperm formation, chromatin undergoes extensive remodeling in which most histones are replaced by sperm-specific protamine proteins enabling extensive compaction of DNA in sperm nuclei (20). Nonetheless, a small percentage of retained histones, at about 1% in mouse and 15% in men are retained in mature sperm (16, 18). Several recent studies characterized the genome-wide distribution of nucleosomes in mature sperm of mouse and men, with partially discordant results (17, 18, 21), as reviewed by (22). Using a protocol that enables efficient opening of the extremely compact mouse sperm nucleus, followed by micrococcal nuclease digestion and deep sequencing, we observed an approximately 10-fold over-representation of sequences localizing to promoters enriched in CpG di-nucleotides in nucleosome associated DNA (23). Histone retention in mouse sperm is predictable, occurring mainly at regions of high CpG density and low DNA methylation, e.g. at promoters of housekeeping and development regulating genes (23). Retention at non-methylated CGIs is

conserved between mouse and man (16-18, 24). In sperm, promoter regions of housekeeping genes contain the histone variant H3.3 and are marked by histone H3 lysine 4 di- and trimethylation. In contrast, developmental regulatory genes in sperm contain both H3.3 and canonical H3.1/H3.2 histones and are marked by H3K27me3 at their promoters (16, 17). The function of sperm histones and their modifications in embryonic development, offspring health and epigenetic inheritance is largely unknown.

In *C. elegans*, loss of the H3K4me2 demethylase *spr-5*, is associated with a transgenerational phenotype appearing as progressive sterility due to stable accumulation of H3K4me2 in primordial germ cells (25). The mouse homolog, lysine demethylase 1A (Kdm1a), controls gene expression and development by catalyzing mainly the removal of mono- and dimethylation of H3K4 (26). To study the role of retained histones in sperm for embryo development and transgenerational epigenetic inheritance, we targeted H3K4 methylation, an epigenetic mark associated among others with developmental genes in sperm (23). We aimed at generating a mouse model producing spermatozoa with reduced H3K4me2 within the CGIs of genes implicated in development. Towards this goal, we over-expressed the human KDM1A histone lysine 4 demethylase during spermatogenesis in mice and studied the development and fitness of offspring.

Offspring sired by heterozygous transgenic fathers (TG) and non-transgenic descendants (nonTG) had increased rates of birth defects, neonatal mortality, and altered gene expression. We characterized histone and DNA methylation states in the sperm of TG and nonTG sires. Over-expression of *KDM1A* was associated with specific losses of H3K4me2 at over 2300 genes, including many developmental regulatory genes. Unlike in other examples of paternal transgenerational inheritance (11, 27, 28), we observed no changes in sperm DNA methylation at

CpG-enriched regions. Instead, we measured robust and analogous changes in sperm RNA content of TG and nonTG descendants as well as in their offspring at the two cell stage. These changes in expression and phenotypic abnormalities observed in offspring correlate with altered histone methylation levels at genes in sperm. This study demonstrates that KDM1A activity and related reduced histone methylation during sperm development has catastrophic consequences for offspring and implicates histone methylation and sperm RNA as potential mediators of transgenerational inheritance.

2.3 Results

The *KDM1A* transgene is specifically over-expressed in the mouse male germline

We generated two lines of transgenic mice that express human *KDM1A* histone lysine 4 demethylase under the control of the gonad specific, truncated form of the human polypeptide chain elongation factor 1 alpha (EF-1 α) promoter (29) (Fig. 2A; fig. S1A, B). This promoter is active in testicular germ cells but not somatic tissue (Fig. 2B, fig. S1C, D). Transgenic *KDM1A* mRNA localized to spermatogonia, spermatocytes, and weakly in round spermatids (Fig. 2C). A similar distribution was observed for the transgene marker GFP (fig. S1C). Histopathological evaluation of testes revealed normal spermatogenesis at all ages examined, from the first wave of spermatogenesis and throughout adulthood, in offspring of both lines. Likewise, testis and epididymal weights, as well as sperm counts, were largely unaffected (fig. S2). To assess DNA integrity we monitored phosphorylation of gamma H2A.X and TUNEL as markers of DNA damage. The cellular distribution of H2A.X phosphorylation in spermatogonia, spermatocytes, and spermatids was invariant between transgenic and control animals (fig. S3A). Likewise, we did not observe an increased number of TUNEL positive cells in testes of transgenic males (fig.

S3C, D). Finally, expression of the retrotransposon LINE-1 did not differ between TG, nonTG, or control testes (fig. S3B). Together, these data indicate that *KDM1A* expression does not induce DNA damage nor genome instability(30).

Transgenerational paternal effects in TG and nonTG sired descendants

Breedings from TG sires and nonTG sires were conducted to determine possible inter- and transgenerational paternal effects of *KDM1A* over-expression on fitness of offspring. Pedigrees and generations are shown in Figure 2. Male transgenic offspring were bred to C57BL/6 females generating the experimental heterozygous TG and nonTG brothers. Males from TG² to TG⁴ and nonTG³ to nonTG⁵ generations (representing F₀-F₃ in our transgenerational studies) were bred to C57BL/6 females and we analyzed the offspring for intergenerational and transgenerational effects comprising F₁-F₄ pups (Fig 2D,E). First, we analyzed the pups of TG and nonTG sires over multiple generations for development and survival (Fig. 2F to I). We determined the litter size at birth and weighed, sexed, and examined the pups at 36h and 48h after birth, and on postnatal days (PND) 6 and 21 (table S1). Neonatal survival of offspring sired by TG²⁻⁴ and by nonTG³ males was reduced compared to that of C57BL/6 controls ($p<0.05$) (Fig. 2F, G). We observed cumulative effects of transgene expression on survivability; with each successive generation of exposure of developing sperm to the transgene (TG²⁻⁴), survivability of transgenic offspring comprising generations F₁-F₃ decreased in both lines ($p<0.05$) (Fig. 2F, G). We also observed pups that were abnormal and runted (<75% of average body weight of the litter), with apparent abnormalities in limbs, the skeleton (table S1), and in the skin (flaky as pups and then grey and white flecked as adults) (Fig. 2H, I). Analysis of offspring sired by

individual males show that the described phenomenon is not driven by transmission of abnormalities by a few males to many offspring but by many males transmitting epigenetic alterations to a variable number of offspring (fig. S4). Moreover, frequencies of pup survival and abnormalities were not related to whether the offspring inherited the transgene from a transgenic father or not ($p<0.05$). The presence of possible alterations in the epigenome of haploid spermatozoa that do not carry the actual transgene is likely due to the fact that in TG heterozygous sires, KDM1A is active in germ cells at various stages of spermatogenesis, such as spermatogonia and spermatocytes which are diploid cells. Moreover, even following the two meiotic divisions, haploid spermatids share mRNAs via cellular bridges resulting from incomplete mitotic and meiotic divisions. Consequently, although the transgene is transcribed in only 50% of spermatids, all haploid spermatids within the cellular syncytium will be affected(31, 32). Since transgene expression is restricted to the developing germ line (Fig. 2B, C, fig. S1C, D), and nonTG sires do not express the transgene in their developing sperm, these data argue that the aberrant phenotypes observed in nonTG pups (F_1-F_3) are related to transgenerational epigenetic phenomena.

It is probable that C57BL/6 mothers culled pups perceived as abnormal. As a consequence we may have underestimated the phenotypic abnormalities assessed in the pups (table S1). To address this and to further characterize the developmental defects across generations, we performed detailed phenotypic analysis at embryonic day (E) 18.5. Each TG²⁻⁴ (F_0-F_2) and nonTG³⁻⁵ (F_1-F_3) sire was mated to at least two females and pregnancies assessed for litter size, pre- and post- implantation loss (combined to give total-pregnancy loss), and abnormal fetuses (Fig. 3-4). In TG sired offspring, abnormalities were varied and affected multiple systems. Examples of developmental errors included skeletal anomalies observed as

malformed digits, spinal defects, abnormal craniofacial structures, and failures in the development of limbs and body segments (Fig. 3-4). Likewise, in pregnancies sired by nonTG³ (F₁) and nonTG⁴ (F₂), we observed F₂ and F₃ offspring with visible abnormalities and skeletal defects that were significantly greater than in control breedings ($p<0.05$) (Fig. 3, table S2). Abnormalities occurring in F₃ offspring (sired by a nonTG⁴ father) represent transgenerational effects (Fig. 3-4). Pregnancy loss increased in several TG and nonTG generations (Fig. 3). No gross visible abnormalities were observed in E18.5 F₄ offspring sired by nonTG⁵ males (Fig. 3). To further identify the specific nature of the skeletal defects and to confirm observations in nonTG offspring, a subset were analyzed specifically for skeletal abnormalities. These analyses revealed abnormalities such as errors in ossification, spinal defects, and missing or abnormal craniofacial structures in both TG and nonTG offspring (Fig. 4, table S2). The dilution of the gross abnormalities in F₄ offspring sired by nonTG⁵ was confirmed by the skeletal analysis (table S2). Thus by three generations following transgene exposure there was a normalization of offspring phenotypes in terms of birth defects at the levels detectable by our analysis (gross morphological and skeletal). Nonetheless there is likely an underestimate of offspring abnormalities as in-depth pathology analysis was not performed and abnormalities were limited to those detected by observation of pups or skeletal analysis.

Transgenic *KDM1A* expression alters H3K4 dimethylation in sperm

To relate the developmental defects in offspring to possible changes in chromatin states in parental spermatozoa, we profiled the KDM1A target H3K4me2 in a genome wide manner in sperm from TG³ males (n=11), their non-transgenic littermates (nonTG³, n=9), and age matched C57BL/6 controls (n=11) using the previously mentioned method specifically developed for

analyzing chromatin states in highly condensed sperm (33). We prepared nucleosomes by micrococcal nuclease digestion and performed chromatin immunoprecipitation with H3K4me2-specific antibodies followed by next generation sequencing (ChIP-seq) (fig. S6). In read-count normalized libraries, we observed a reduction in H3K4me2 enrichments at CpG-rich sequences within transcriptional start site (TSS) regions of different genes in TG³ sperm compared to controls (Fig. 5A). To identify TSS regions with altered H3K4me2 profiles in a genome wide manner, we calculated the enrichment of H3K4me2 for genomic regions encompassing 250 base pairs (bps) upstream and downstream (+/-) of TSS for the different sperm samples. Since we previously showed that nucleosomes are retained in mouse spermatozoa at CpG-rich sequences, e.g. close to TSS (23), the selected windows thus correspond to regions with the highest nucleosome occupancy in sperm and are most likely to transmit any histone encoded information to the offspring. Reduced H3K4me2 levels occurred at 28.7% of the 8171 TSS regions marked by H3K4me2 in TG³ samples, whereas 3.4% TSS regions displayed modest elevated levels of H3K4me2 (Fig. 5B,C) The predominant down-regulation of H3K4me2 levels is consistent with the reported H3K4me2 demethylase activity of KDM1A.

Direct sequencing of nucleosome-associated DNA did not reveal differences in nucleosome occupancies between TSS regions with differential H3K4me2 levels in TG³ sperm nor between TG³, nonTG³ littermates, and control samples (fig. S7). These results exclude impairment of nucleosome retention as a putative cause underlying the reduction of H3K4me2 at certain loci in TG³ sperm. To assess the reasoning underlying reduced H3K4me2 occupancy levels at certain TSSs, we categorized the regions according to CpG density and various chromatin attributes such as the occupancy of the histone variant H3.3, high and intermediate H3K4me3 and H3K27me3 levels, as measured in sperm of control animals (23) and which

associate with specific gene functions (Fig. 5B,C) (23). We observe that regions with reduced H3K4me2 levels are CpG-rich (Fig. 5D), suggesting CpG-density related targeting of human KDM1A. Indeed, endogenous murine Kdm1a is particularly localized at CpG-dense TSS regions in mouse spermatocytes that are characterized by H3K4 methylation in sperm (23, 34) (Fig. 5A-C). Moreover, regions with reduced H3K4me2 levels in TG³ sperm have higher occupancy levels of murine Kdm1a compared to regions with unchanged H3K4me2 levels (Fig. 5E), pointing to specificity in human KDM1A activity at target sites of endogenous murine Kdm1a. Together, these data suggest that transgenic human KDM1A is targeted to selected regions that are extensively bound by endogenous murine Kdm1a during spermatogenesis.

TSS regions with reduced H3K4me2 levels are strongly enriched for H3.3, as measured in control sperm (Fig. 5F) (23). Since high H3.3 levels at strong CGIs in sperm relate to high nucleosome turnover at CGIs in spermatids (23), the locally reduced H3K4me2 levels at strong CGI promoters in TG³ sperm may reflect the inability of H3K4 histone methyltransferases to counteract local TG-conferred KDM1A activity and thereby fail to re-establish the H3K4me2 state during nucleosome turnover in spermatids of TG³ males.

Gene ontology analysis suggests that transgenic KDM1A regulates specific classes of genes (tables S3, and S4). Genes of the H3K4me3 high and intermediate chromatin classes with reduced H3K4me2 levels in TG³ males serve functions in various cellular metabolic protein processes. For example reduced H3K4me2 in TG³ sperm was observed for *Pdk1* (3-phosphoinositide dependent protein kinase 1) and *E2F6* (Fig 5A). Conditional ablation of Pdk1 in the pancreas leads to diabetes (35) and knockout mice have impaired embryogenesis, growth and nervous system development (36, 37). Loss of *E2F6* results in mice displaying skeletal defects (38). H3K27me3 marked genes with H3K4me2 reduced levels serve various functions

during embryonic development. For instance, *Gsc* (Goosecoid homeobox) has reduced H3K4me2 in TG³ sperm at a region overlapping H3K27me3, H3K4me3 and Kdm1a binding (Fig. 5B,C). A targeted mutation of *Gsc* gave rise to mice with craniofacial abnormalities and skeletal defects(39). These phenotypes reported in gene ablation models are highly reminiscent of those observed in TG¹⁻⁴ and nonTG³⁻⁴ offspring yet the relationship to altered histone methylation in developing sperm remains unclear. Our gene ontology analysis of embryonically expressed genes indicates that genes with loss of H3K4me2 are involved in morphogenesis, patterning, vasculature development, cartilage development and catabolic and metabolic processes (table S4). In contrast, H3K4me3 high and intermediate chromatin genes with unchanged H3K4me2 levels in TG³ males execute functions in spermatogenesis and various nuclear processes, respectively. These results emphasize the targeting of hKDM1A to specific subsets of genes that may preferentially alter offspring development. Moreover, the gene subsets are directly in line with our phenotypes where defects were not observed on spermatogenesis in transgenic sires, but were observed in the TG-sired offspring.

Changes to histone H3K4me2 are independent of DNA methylation state

In contrast to TG³ spermatozoa, the ChIP-seq analysis of sperm of nonTG³ males (littermates of TG³ animals) did not reveal any differences in H3K4me2 occupancy in comparison to sperm of TG³ or control males (Fig. 5B, C). These data suggest that the phenotypic aberrations observed in offspring of nonTG³ sires are not directly related to reduced H3K4me2 levels, as observed in mature sperm of TG³ sires. Previously, KDM1A was reported to mediate the removal of H3K9 mono- and dimethylation in some cases (40, 41). In various somatic cells, H3K9me2 is an abundant heterochromatic modification that is localized in large

domains throughout the genome (42). H3K9me2 generally does not localize to and is mutually exclusive with regions marked by H3K4me2 and/or H3K27me3 which are prevalent at CGI regions containing nucleosomes in sperm (42). Such anti-correlation minimizes a possible contribution of H3K9me2 to paternal transmission of the embryonic impairment traits observed in TG and nonTG offspring.

Altered DNA methylation has been associated with phenotypes induced by paternal environmental exposures and suggested as a mechanism underlying epigenetic inheritance(43). To address whether reduced H3K4me2 may induce heritable changes in DNA methylation levels, we first used a targeted approach to assess DNA methylation in sperm from TG³ (F₁) and nonTG³ (F₁) littermates. Target selection was based on the identification of regions as being differentially methylated in TG³ sperm for H3K4me2 in comparison to C57BL/6 controls (3 overlapping windows of 250 bp each), and enrichment for CpGs. Selected targets were analyzed for altered DNA methylation by quantitative Sequenom MassARRAY, a technique based on bisulfite conversion followed by mass spectrometry analysis with a resolution at the CpG level. In all 24 targeted genes, we failed to observe any significant differences in DNA methylation in TG³ and nonTG³ sperm in comparison to controls (fig. S8).

Next we used reduced representation bisulfite sequencing (RRBS) for a genome wide comparison of DNA methylation levels predominantly at CpG islands in control, TG³ and nonTG³ sperm. Analysis based on methylation percentages of single CpG sites found across the genome, showed a high degree of similarity between samples of all three genotypes ($R=0.98-0.99$) with no apparent group clustering (Fig. S9A). When testing individual CpG sites for significant association with genotype, we found only few more than expected by chance (fig. S9B, C). In summary, using targeted Sequenom analysis and RRBS, we did not observe an

overrepresentation of changes in CpG methylation levels in spermatozoa of TG³ and nonTG³ transgenic mice versus control samples. These findings indicate that DNA methylation at CpG islands is not implicated in the molecular processes leading to the transgenerational phenotypes observed.

Overexpression of KDM1A is associated with altered sperm RNA content.

RNA analysis of sperm from TG³ and nonTG³ revealed a common non-genetic molecular change. Sperm is a rich source of diverse RNAs that may function as potential mediators of paternally transmitted effects (44). We compared the RNA content of sperm from TG³ and nonTG³ mice to controls using the Affymetrix GeneChip ST2.0 Array, which enables detection of 28,000 coding transcripts and over 7000 non-coding RNAs including around 2000 long intervening/intergenic noncoding RNAs (lincRNAs). TG³ and nonTG³ spermatozoa displayed a comparable molecular change in RNA content relative to the control with 564 RNAs commonly misregulated, among 650 RNAs altered in TG³ sperm and 619 in nonTG³ sperm (\log_2 fold change > 1.0 ; FDR < 0.05). Among the shared misregulated RNAs in the nonTG and TG sperm, 67 were non-coding RNAs and 471 were protein-coding transcripts (table S5). As would be predicted with reduction of a gene activating histone modification, 99% of differential RNAs were reduced in TG³ and nonTG³ sperm compared to the control (643/650 and 613/619 respectively). Over 60% of RNA associated promoters were marked by H3K4me3 in wild-type sperm potentially reflecting transcriptional activity at preceding stages of male germ cell development(23) (Fig. 6B). Only ~3% (15/564) were marked by H3K27me3 in wild-type sperm. Finally, 41 commonly deregulated transcripts were associated with genes with reduced H3K4me2 levels at their TSS in TG³ sperm (Fig. 6B). These intriguing findings offer the

tantalizing possibility that transcripts and regulatory RNAs are transmitted via TG³ and nonTG³ sperm to the embryo and may be implicated in transmission of the phenotype to the offspring.

TG and nonTG sperm alter gene expression in early embryos

To understand if there was a link between genes bearing altered histone methylation in the TG³ sperm to embryo gene expression and offspring phenotypes, we performed array expression analysis from TG⁸, nonTG⁸ and control sired two-cell embryos. We chose to examine two-cell embryos as they are in terms of developmental timing close to sperm in which we observed H3K4me2 changes. We hypothesized that potential changes in paternal chromatin states would be the least diluted in two-cell embryos compared to later stages of development. We identified 874 RNAs in embryos sired by TG⁸ and 123 in nonTG⁸ embryos (F2 generation) that were differentially expressed (table S6) (FDR < 0.3). Most of deregulated genes in TG⁸ sired embryos were up-regulated (80%; 703/874), whereas about half of the genes were up-regulated in nonTG⁸ sired embryos (46%; 56/123). Notably, 71 genes were commonly mis-expressed in two cell embryos sired by TG⁸ and nonTG⁸ littermates (Fig. 6D,E and table S6) (FDR < 0.3). Moreover, 110 genes deregulated in two-cell embryos had reduced H3K4me2 at TSS in TG³ sperm. A subset of these genes are also marked by H3K27me3 and H3K4me3 (Fig. 6C-E). Some of the differentially expressed genes in the embryos that overlap with reduced H3K4me2 in TG³ sperm, or regions in sperm enriched in H3K27me, or H3K4me3, can be linked to the developmental abnormalities observed in TG and nonTG sired pups (Fig. 6C). These data establish that the triggering event leading to altered H3K4me2 in developing sperm has consequences on gene expression in embryos.

2.4 Discussion

Here we show that an increased expression of the chromatin modifier KDM1A during spermatogenesis induces major developmental defects in offspring, which were transmitted paternally for three generations, even in the absence of transgene expression in the germ line of non-transgenic offspring. Our data suggest that the critical initiating event in our model is the alteration in histone methylation in developing male germ cells, leading to offspring abnormalities. We show that changes to histone methylation in the hKDM1A model occurred in the absence of changes to DNA methylation at CGIs in sperm. We also observed altered RNA profiles in sperm of TG and nonTG males (F1 generation) and their two-cell embryonic offspring (F2 generation) (Fig. 6). Our data emphasize that the mechanisms of transgenerational epigenetic inheritance are complex and likely involve several molecular factors like the establishment of chromatin states in spermatogenesis, DNA methylation, and sperm borne RNA.

We observed a large diversity in aberrant phenotypes including abnormalities in bone, skin, reduced survivability, and retarded growth in offspring sired by many, rather than a few fathers. These observations together with the high penetrance of phenotypes that was resolved by the fourth generation of non-transgenic offspring points to an epigenetic mechanism driving impaired development. It is unlikely that hKDM1A expression during spermatogenesis behaves comparable to a chemical mutagen or induces a mutator phenotype in germ cells or offspring given the very high frequency of abnormalities observed (45-47). Consistent with that notion, our assessments using multiple approaches revealed no DNA damage or chromatin instability. Further supporting the stability of sperm chromatin is the normal sperm head morphology. Men and mouse mutants with altered histone content in sperm have misshapen sperm heads lacking compaction (48, 49). Unlike prior studies of transgenerational inheritance, however, we

confirmed by several methods including genome wide DNA methylation analysis by RRBS and targeted analysis with epityper that DNA methylation at principally CGIs was not altered in TG or nonTG sperm.

Transgenic KDM1A expression induces a major reduction in H3K4me2 occupancy locally at many CGI containing promoters in sperm of TG sires. Reductions in H3K4me2 in sperm are observed at sites of high Kdm1a occupancy in spermatocytes supporting the notion that transgenic KDM1A functions preferentially at sites of endogenous Kdm1a proteins. Since nonTG³ animals sire offspring with developmental defects as do TG³ males, reduction in H3K4me2 as observed in TG³ sperm likely does not directly mediate epigenetic inheritance of developmental defects across generations. Instead our data point to the idea that disruptions in H3K4me2 have a cascading effect altering transcription at genes in spermatogenesis and the RNA content of sperm. Correspondingly, H3K4me2 changes in sperm were correlated with changes in gene expression from genes bearing H3K4me3, and H3K27me3, that are normally already present in sperm at CGIs of subsets of genes known to control embryonic development(23). Fitting with the aforementioned, genes with altered expression in two cell embryos sired by either TG or nonTG were associated with genes in sperm enriched in histone methylation (H3K4me2, H3K4me3, H3K27me3). This suggests that altered histone methylation in sperm can influence early embryo gene expression. Similarly, mutation of the *Smarca5*^{MommeD4} allele of the Snf2h chromatin remodeler during spermatogenesis has been associated with changes in expression of the epigenetic sensitive *agouti viable yellow* (*A^{vy}*) gene. As in our study, wild-type offspring of heterozygous *Smarca5*^{MommeD4} sires had variable phenotypes(50) also assigning an important role of proper chromatin regulation during spermatogenesis to offspring fitness and gene expression.

Our data also offer the possibility of sperm RNA content as a contributing factor to the offspring phenotypes. There was remarkable overlap in RNA content of TG and nonTG sperm, with ~85% similarity in RNAs that were different from control sperm RNAs. An example of paternal effects via RNA was demonstrated in *C. elegans*, where induced viral expression leads to paternal transmission of small viRNA across multiple generations. Like our model, the wild-type offspring worm descendants also inherited the viRNA molecules and the phenotype of viral protection. This transgenerational effect was extinguished by the third generation beyond exposure to the transgene thus showing a gradual loss of the phenotype, as we observed (51). However the likelihood that a diffusible agent such as RNA could drive the transgenerational phenotype we observe is low. The reason is that germ cell fate in mammals is specified during gastrulation, occurring many cell divisions after fertilization. In the absence of a means to amplify a diffusible RNA signal, it will be lost over the many cell divisions until germ line specification. We therefore hypothesize a model for mammals in which altered chromatin in developing sperm is associated with abnormal sperm born RNAs. These RNAs could then signal to chromatin to regulate gene expression during development (Fig. 7). Regulatory RNAs made up a significant portion of the differential RNA content in sperm from TG and nonTG males in comparison to controls. Functions of regulatory RNA include control of pluripotency, differentiation, and guidance in the setting of the epigenome (52-54). Sperm is known to transmit long noncoding RNAs to the embryo, which have been postulated to affect the post-fertilization genome (19). If lncRNAs are not correctly transcribed in developing germ cells, altered lncRNA levels could then be transported to the embryo where gene expression could be modified. LncRNAs interact with chromatin regulatory machinery and serve in guiding chromatin changes.

Moreover lincRNA knockouts (*Fendrr*, *Peril* and *Mdgt*) display peri- and post-natal lethality prior to PND20(55), a common observation in our transgenic descendants.

2.5 Conclusion

Our data show for the first time, that abnormalities in histone methylation during spermatogenesis without affecting DNA methylation at CGIs are associated with altered embryo gene expression and development. Our findings of transgenerational epigenetic inheritance highlight the potential for genetic lesions in chromatin modifiers and environmentally-induced alterations in histone methylation during spermatogenesis as underlying causes of birth defects and disease that may be traceable to the father. Individual responses to environment and inheritance of sensitivities and/or predisposition to disease may be shaped by factors such as gene copy number of chromatin modifiers.

2.6 Materials and Methods

Generation of KDM1 Transgenic Mice

For generation of KDM1 transgenic mice (TG), full length human KDM1 cDNA (kindly provided by Dr. Yang Shi, Harvard University); accession number NM_015013, CCDS30627) was subcloned into a modified pIRES-EGFP vector (Clontech). We replaced the cytomegalovirus (CMV) promoter of the expression vector by an isoform of the human elongation factor-1 alpha promoter (hEF-1 α) driving germ cell-specific expression of the KDM1 transgene(29). We microinjected the transgenic construct into C57BL/6 zygotes and obtained two female transgenic founders. We crossed founders with C57BL/6 males to create the two transgenic lines. We subsequently transmitted the transgene by heterozygous males bred to

C57BL/6 females. The two transgenic lines showed similar developmental phenotypes in offspring and served as a control for transgene integration site effects. We extracted genomic DNA from tail biopsies and performed Southern blot analysis to determine the transgene copy number. All C57BL/6 control mice were obtained from Charles River Laboratories International (Wilmington, MA, USA). Mice were watered and fed standard mouse chow *ad libitum* and housed under conditions of controlled light (12-hour light, 12-hour dark cycle) at 21°C and 50% humidity. All animal procedures were approved by the Animal Care and Use Committee of McGill University, Montreal, Canada.

***In Situ* hybridization**

In situ hybridizations were performed as previously described(56), using DIG-labeled RNA probes (Roche).

Genotyping by PCR on Genomic DNA

Genotyping was performed on tail tip DNA as previously described(57). Oligonucleotide primers used for genotyping by PCR were: hsEF1alpha-fw1 (TTC TCA AGC CTC AGA CAG TGG), Flag-re1 (TCG TCA TCG TCC TTG TAG TCC); hsLSD1-ex19-fw1 (TAC GAT CCG TAA CTA CCC AGC), pIRES-EGFP-re2 (TCT TAG CGC AGA AGT CAT GCC).

RNA Isolation and RT-PCR

Total RNA was extracted using Purezol (BioRad, Hercules, CA, USA) according to the manufacturer's instructions and treated with DnaseI (RNeasy Mini Kit, Qiagen). Isolated RNA was reverse transcribed with random primers using the High Capacity cDNA Reverse

Transcription kit (Applied Biosystems, Foster City, CA, USA). cDNAs were subjected to PCR using the following oligonucleotide primers: *beta-Act1-fw* (5' ACC TTC AAC ACC CCM GCC ATG TAC G 3' (M=A/C)), *beta-Act2-re* (5' CTR ATC CAC ATC TGC TGG AAG GTG G 3' (R=A/G)); *hKDM1A-ex2-fw1* (5' TGG ATG AAA GCT TGG CCA ACC 3'), *hsLSD1-ex3-re1* (5' AGA AGT CAT CCG GTC ATG AGG 3'); *hsKDM1A-ex11-fw1* (5' GCC ACC CAG AGA TAT TAC 3'). For embryo gene expression analysis for the transgene, major organs from three transgenic embryos at E18.5 were homogenized via mortar and pestle followed by extrusion through a 25G needle. RNA was extracted using Qiagen RNeasy RNA purification kit (Product # 1050349; Lot# 145015691). 1 ug of RNA from each sample was converted to cDNA using High Capacity cDNA Reverse Transcriptase Kit (Applied Biosystems; Product # 4368814; Lot# 1108153). To determine expression of *hsKDM1A*, *hsKDM1A-ex2-fw1* and *hsLSD1-ex3-re1* primers were used as above. Primers for *Gapdh* were used as loading control (Forward: 5' ACT TTG GCA TTG AAG GGC TG 3'; Reverse: 5' TGG AAG AGT GGG AGT TGC TGT TG 3').

PCR design

Pre-denaturation of one cycle at 94°C for 4 minutes followed by PCR amplification repeated for 40 cycles of denaturation at 94°C for 40 seconds, annealing at 60°C for 25 seconds and elongation at 72°C for 60 seconds. Final elongation performed at 72°C for 10 minutes. PCR reactions were visualized on Qiaxcel advanced system using the QiAxcel DNA Screening kit (Product# 74104; Lot# 145019088)

Sperm Isolation, Sperm Count, Sperm Morphology

For sperm counts spermatozoa were recovered from paired cauda epididymides of sexually mature mice as follows: Cauda epididymides were placed into PBS, cut and gently agitated at 37°C to allow sperm swim out. After 5 min of incubation a 10 µl aliquot of the sperm solution was removed and subjected to hemacytometric counts following standard procedures. For isolation of sperm for ChIP-seq a swim up procedure was used to ensure pure cell population(33).

Analysis of Pups

We bred C57BL/6 (n=18), Line 1 TG²⁻³ (n=10, 14) and nonTG³⁻⁵ (n=8, 7, 4), and Line 2 TG²⁻⁴ (n= 4, 6, 5) males to C57BL/6 females. Shortly after copulation, as determined by the presence of a vaginal plug, males were removed from females. Litter size was determined at birth and pups were sexed, weighed and examined at 36 h and 48h after birth, and on postnatal days (PND) 6 and 21. TG and nonTG pups with and without external malformations, and C57BL/6 controls were subjected to skeletal staining. Abnormal pups were identified by gross morphological appearances such as the presence of a skin abnormality, underdeveloped limb or runting. Frequency of morphologically abnormal pups was calculated based on the total number of offspring per generation.

Analysis of E18.5 Fetuses and assessment of pregnancy loss

We bred and C57BL/6 (n=21), Line 1 TG²⁻³ (n=8, n=12) and nonTG³⁻⁵ (n= 10, n=16, n=14), and Line 2 TG³⁻⁴ (n=2, n=9) and nonTG⁴ (n=9) males to C57BL/6 females. Control matings were age matched so that statistical comparisons were always between matings in which

the sires were the same age. Copulation was determined by the presence of a vaginal plug. The day of the plug was determined as embryonic day 0.5 (E0.5). Fetuses were collected from the uteri of the female mice at E18.5 after mating. Each fetus was carefully examined, and sex, weight and crown-rump length were determined. Moreover, placental gross morphology, weight and diameter were examined. Fetuses with external malformations and the corresponding controls were subjected to skeletal staining or fixed in Bouin's fixative and subsequently subjected to histopathological analysis. We determined the number of ovulations by counting the number of corpora lutea (CL) in E18.5 pregnant females from the above matings (C57BL/6, n=32; L1 TG²⁻³, n=15, 24; L1 nonTG³⁻⁵, n= 18, 30, 20; L2 TG³⁻⁴, n=4, 16; L2 nonTG⁴, n=17; n = females per group). Ovaries were placed into PBS in separate drops and CL were counted blindly with respect to the number of embryos actually recovered using a dissecting microscope with top lightening. The sum of pre- and post-implantation losses were used as a measure of total pregnancy loss. This was done by comparing the number of CL produced with the number of total embryos per group: Total Pregnancy Loss per Group [%] = (#corpora lutea - total embryos) per group / #corpora lutea per group x 100.

Skeletal Staining and Histopathology

Alcian blue and Alizarin red staining of cleared skeletal preparations was performed according to Hogan et al(58). Skin and viscera were removed and carcasses were fixed overnight in 95% EtOH. Afterwards, carcasses were stained overnight in Alcian blue 8GS (80 ml 95% EtOH / 20ml acetic acid / 15mg Alcian blue 8GS), fixed in 95% EtOH for 2-5 h, and transferred to 2% KOH for 24 h. Then, skin and muscles were taken off. Next, carcasses were stained

overnight in 1% KOH / 0.015% Alizarin red S. Next, skeletons were cleared in 1% KOH / 20% Glycerol for approximately 48 h. Skeletons were stored in 20% Glycerol for analysis.

ChIP-Sequencing

ChIP-sequencing experiments were performed as described before(23, 33) using an antibody against H3K4me2 (Millipore 07-030). ChIP-seq libraries were prepared using Illumina ChIP-seq DNA Sample Prep Kit (Cat# IP-102-1001) and sequenced on Illumina GA II. Processing and alignment of the ChIP-seq data were performed as previously(23) and was based on mouse mm9 assembly (July 2007 Build 37 assembly by NCBI and Mouse Genome Sequencing Consortium).

Classification of H3K4me2 signal around transcriptional start sites (TSS) in KDM1A transgenic mouse

Enrichment values for regions surrounding ± 250 bp TSS were calculated in a similar way as described(23). Enrichments refer to the ratio of H3K4me2 signal in sperm of KDM1A TG³ males over H3K4me2 signal in sperm of C57BL/6 wild type (WT) mice. We called genes H3K4me2 positive that had more than or equal to 5 reads (log2 scale) in the regions analyzed in both replicates of WT H3K4me2 (Supplementary Fig. 6a). We used the averages of two H3K4me2 data sets as the WT H3K4me2 measurements. To identify genes with altered H3K4me2 levels in KDM1A TG³ sperm, we determined the ratio of H3K4me2 occupancy around TSS (± 250 bp) in TG³ over WT sperm (Supplementary Fig. 6b). Since the H3K4me2 ratios were not normally distributed for down-regulated TSS regions, we performed k-means clustering analysis coupled to heatmap visualization to set the cut-off for genes with down-

regulated H3K4me2 levels (see Fig. 4b, c; Supplementary Fig. 6). To increase sensitivity, k-means clustering was performed taking the occupancies of nucleosomes, H3K4me3, H3K27me3, H3.3 and H31/H3.2 histones in WT sperm into account. For clusters with over 1000 genes, we chose randomly 1000 genes for heatmap visualization purposes. For setting the right cut-off, we arbitrarily identified the upper 10% of TSS regions having elevated H3K4me2 levels in TG³ over WT sperm. Chromatin snapshots (Fig 4a) were generated as described by Erkek et al.(23). Shortly, chromatin images represent the read counts per base averaged over a moving 300 bps interval. Averaged reads counts were normalized for total read counts across samples.

Sequenom MassARRAY methylation analysis

1µg of DNA from TG (n=5) and nonTG (n=5) and control C57/BL6 (n=5) sperm was bisulfite treated with EZ DNA Methylation Gold Kit (Zymoresearch, D5007). We designed primers for the amplification of different amplicons of selected targets using the Sequenom EpiDesigner application. Targeted regions encompassed a 250bp window showing large reduction of H3K4me2 (ratio of TG versus WT control < -2.5), high CpG density and H3.3 occupancy(23). Sequenom MassARRAY methylation analysis was then performed using the MassARRAY Compact System (Sequenom, Inc. San Diego, CA). This system is based on mass spectrometry (MS) analysis for qualitative and quantitative detection of DNA methylation using homogeneous MassCLEAVE (hMC) base-specific cleavage and matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) MS. Spectra were elaborated by the Epityper software v1.2.0 (Sequenom) which provides methylation values of each CpG unit expressed as percentage. Such values result from the calculation of the ratio mass signals between the methylated and non-methylated DNA.

RRBS analysis

RRBS libraries were generated according to previously published protocols using the gel free technique(59). Briefly, 500ng of DNA from each of the 15 samples (TG³ n=5, nonTG³=5 and control n=5) was used in the RRBS experiments. Multiplexed samples were used in paired end sequencing (HiSeq sequencer, Illumina). BSMAP version 2.6 was used to trim reads (phred quality>30 and Illumina adapters), align reads to mm10 and obtain CpG methylation calls(60).

RNA analysis from sperm

~7x10⁶ sperm was isolated by swim-out from each TG³, or nonTG³, or C57BL/6 males. To increase sperm RNA yield for analysis, sperm were pooled from four or five males to give a total of ~20x10⁶ sperm/replicate (C57BL/6, 1 replicate, n= 5; TG, 3 replicates, n=4 per replicate; nonTG, 3 replicates, n=4 per replicate). RNA was extracted from pooled sperm. Somatic cell contamination was avoided by washing with somatic cell lysis buffer (SCLB; 0.1% SDS, 0.5% Triton-X)(61). After two washes with PBS spermatozoa were homogenized by vortexing in buffer RLT (Qiagen RNeasy) supplemented with 50mM -mercaptoethanol and 100mg 0.2mm steel beads. RNA was extracted through Qiazol (Qiagen) followed by chloroform. The aqueous layer containing RNA was processed with RNeasy RNA extraction kit (Qiagen).

RNA analysis from two cell embryos.

Two cell embryos were collected from multiple C57BL/6 pregnancies sired by either a C57BL/6 (n=8), a TG (n=4) or nonTG (n=5) male. Embryos collected per pregnancy ranged from 1-23 and therefore pooling of embryos from multiple sires within each genotype was necessary. On average, 26 embryos were pooled for total RNA extraction. Plug positive females

were sacrificed and oviducts flushed with PBS + 0.01% BSA to collect embryos. Cells were stored at -80C until embryos from 2 – 6 breedings were collected. For each genotype analyzed the following technical replicates were used: C57BL/6 (3), TG (3) and nonTG (2). Experimental details are summarized in Supplementary Table 7.

Microarray data analysis

RNA from sperm and embryo was converted to ds-cDNA using the SensationPlus WT kit and hybridized to GeneChip Mouse Exon 2.0 ST arrays (Affymetrix). CEL files for sperm and embryo data were separately read into R (version 3.1.0), where the Bioconductor package *oligo* was used to calculate transcript-cluster level expression values with the RMA (Robust Multi-array Average) algorithm. Differential expression analysis between the experimental groups was conducted using the limma package(62). The expected false discovery rate (FDR) was estimated using the Benjamini and Hochberg method, and statistical significance for differential expression of sperm RNA was set to FDR<5% coupled with a minimal difference of 1 on the log2 scale ($|FC|>2$). Statistical significance for differential expression of embryo RNA was set to FDR < 30% due to an n=2 of replicates in nonTG sired embryos. Genes reaching statistical significance were submitted to DAVID(63) and Mouse Genome Informatics (MGI) (<http://www.informatics.jax.org>) for functional analysis.

Functional analysis of ChIP-seq data

Functional analysis was performed using the R package topGO(64). Genes depleted in H3K4me2, which were also enriched in high levels of H3K4me3 or in H3K27me3, were selected

(See supplementary table 3). Enrichment analysis was performed using Fisher's exact test and significance was called at $p < 0.01$.

Statistical analyses

The level of significance for all statistical tests used was set at $P < 0.05$ and all tests were two-tailed. The assessment of abnormalities of TG and NonTG sired pups and E18.5 embryos was done on a per-generation basis. Litter size was depicted as mean \pm s.e.m and was analyzed using Student's T-test. The survivability of pups was estimated applying a Kaplan-Meier analysis followed by a log-rank test using GraphPad Prism 5 software (GraphPad Software Inc., La Jolla, California, USA). Total abnormalities and frequency of live pups, total pregnancy loss and total abnormal E18.5 embryos were tested for a significant difference from control offspring using Fisher's Exact Test, uncorrected for multiple comparisons. DNA methylation data was analyzed using unpaired Student's t-test, and all data were analyzed with the aid of Systat 13 and the R statistical computing environment.

2.6 References

1. I. Lobo, K. Zhaurova, Birth defects: causes and statistics. *Nature Education* **1**, (2008).
2. T. J. Mathews, M. F. MacDorman, Infant mortality statistics from the 2005 period linked birth/infant death data set. *National vital statistics reports : from the Centers for Disease Control and Prevention, National Center for Health Statistics, National Vital Statistics System* **57**, 1-32 (2008).
3. M. D. Anway, A. S. Cupp, M. Uzumcu, M. K. Skinner, Epigenetic transgenerational actions of endocrine disruptors and male fertility. *Science* **308**, 1466-1469 (2005).
4. M. E. Pembrey *et al.*, Sex-specific, male-line transgenerational responses in humans. *Eur J Hum Genet* **14**, 159-166 (2006).
5. B. R. Carone *et al.*, Paternally induced transgenerational environmental reprogramming of metabolic gene expression in mammals. *Cell* **143**, 1084-1096 (2010).

6. S. F. Ng *et al.*, Chronic high-fat diet in fathers programs beta-cell dysfunction in female rat offspring. *Nature* **467**, 963-966 (2010).
7. R. Lambrot *et al.*, Low paternal dietary folate alters the mouse sperm epigenome and is associated with negative pregnancy outcomes. *Nature communications* **4**, 2889 (2013).
8. B. G. Dias, K. J. Ressler, Parental olfactory experience influences behavior and neural structure in subsequent generations. *Nature neuroscience* **17**, 89-96 (2014).
9. M. K. Skinner, C. Guerrero-Bosagna, Role of CpG deserts in the epigenetic transgenerational inheritance of differential DNA methylation regions. *BMC Genomics* **15**, 692 (2014).
10. M. Zeybel *et al.*, Multigenerational epigenetic adaptation of the hepatic wound-healing response. *Nature medicine* **18**, 1369-1377 (2012).
11. E. J. Radford *et al.*, In utero effects. In utero undernourishment perturbs the adult sperm methylome and intergenerational metabolism. *Science* **345**, 1255903 (2014).
12. T. H. Bestor, D. Bourc'his, Transposon silencing and imprint establishment in mammalian germ cells. *Cold Spring Harb Symp Quant Biol* **69**, 381-387 (2004).
13. S. A. Smallwood *et al.*, Dynamic CpG island methylation landscape in oocytes and preimplantation embryos. *Nat Genet* **43**, 811-814 (2011).
14. H. Kobayashi *et al.*, Contribution of intragenic DNA methylation in mouse gametic DNA methylomes to establish oocyte-specific heritable marks. *PLoS genetics* **8**, e1002440 (2012).
15. Z. D. Smith *et al.*, A unique regulatory phase of DNA methylation in the early mammalian embryo. *Nature* **484**, 339-344 (2012).
16. U. Brykczynska *et al.*, Repressive and active histone methylation mark distinct promoters in human and mouse spermatozoa. *Nature structural & molecular biology* **17**, 679-687 (2010).
17. S. Erkek *et al.*, Molecular determinants of nucleosome retention at CpG-rich sequences in mouse spermatozoa. *Nature structural & molecular biology* **20**, 868-875 (2013).
18. S. S. Hammoud *et al.*, Distinctive chromatin in human sperm packages genes for embryo development. *Nature* **460**, 473-478 (2009).
19. E. Sendler *et al.*, Stability, delivery and functions of human sperm RNAs at fertilization. *Nucleic acids research* **41**, 4104-4117 (2013).
20. R. Balhorn, The protamine family of sperm nuclear proteins. *Genome biology* **8**, 227 (2007).
21. B. R. Carone *et al.*, High-resolution mapping of chromatin packaging in mouse embryonic stem cells and sperm. *Developmental cell* **30**, 11-22 (2014).
22. M. Saitou, K. Kurimoto, Paternal nucleosomes: are they retained in developmental promoters or gene deserts?
23. S. Erkek *et al.*, Molecular determinants of nucleosome retention at CpG-rich sequences in mouse spermatozoa. *Nature structural & molecular biology* **20**, 1236 (2013).
24. T. Vavouri, B. Lehner, Chromatin organization in sperm may be the major functional consequence of base composition variation in the human genome. *PLoS genetics* **7**, e1002036 (2011).
25. D. J. Katz, T. M. Edwards, V. Reinke, W. G. Kelly, A *C. elegans* LSD1 demethylase contributes to germline immortality by reprogramming epigenetic memory. *Cell* **137**, 308-320 (2009).

26. Y. Shi *et al.*, Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* **119**, 941-953 (2004).
27. C. Guerrero-Bosagna *et al.*, Epigenetic transgenerational inheritance of vinclozolin induced mouse adult onset disease and associated sperm epigenome biomarkers. *Reproductive toxicology* **34**, 694-707 (2012).
28. Y. Wei *et al.*, Paternally induced transgenerational inheritance of susceptibility to diabetes in mammals. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 1873-1878 (2014).
29. T. Furuchi, K. Masuko, Y. Nishimune, M. Obinata, Y. Matsui, Inhibition of testicular germ cell apoptosis and differentiation in mice misexpressing Bcl-2 in spermatogonia. *Development* **122**, 1703-1709 (1996).
30. J. Blanco-Rodriguez, gammaH2AX marks the main events of the spermatogenic process. *Microscopy research and technique* **72**, 823-832 (2009).
31. C. Cho *et al.*, Haploinsufficiency of protamine-1 or -2 causes infertility in mice. *Nat Genet* **28**, 82-86 (2001).
32. R. E. Braun, R. R. Behringer, J. J. Peschon, R. L. Brinster, R. D. Palmiter, Genetically haploid spermatids are phenotypically diploid. *Nature* **337**, 373-376 (1989).
33. M. Hisano *et al.*, Genome-wide chromatin analysis in mature mouse and human spermatozoa. *Nature protocols* **8**, 2449-2470 (2013).
34. J. Zhang *et al.*, SFMBT1 functions with LSD1 to regulate expression of canonical histone genes and chromatin-related factors. *Genes & development* **27**, 749-766 (2013).
35. N. Hashimoto *et al.*, Ablation of PDK1 in pancreatic beta cells induces diabetes as a result of loss of beta cell mass. *Nat Genet* **38**, 589-593 (2006).
36. B. J. Collins, M. Deak, V. Murray-Tait, K. G. Storey, D. R. Alessi, In vivo role of the phosphate groove of PDK1 defined by knockin mutation. *Journal of cell science* **118**, 5023-5034 (2005).
37. M. A. Lawlor *et al.*, Essential role of PDK1 in regulating cell size and development in mice. *The EMBO journal* **21**, 3728-3738 (2002).
38. M. Courel, L. Friesenhahn, J. A. Lees, E2f6 and Bmi1 cooperate in axial skeletal development. *Developmental dynamics : an official publication of the American Association of Anatomists* **237**, 1232-1242 (2008).
39. M. Blum *et al.*, Gastrulation in the mouse: the role of the homeobox gene goosecoid. *Cell* **69**, 1097-1106 (1992).
40. E. Metzger *et al.*, LSD1 demethylates repressive histone marks to promote androgen-receptor-dependent transcription. *Nature* **437**, 436-439 (2005).
41. M. Wissmann *et al.*, Cooperative demethylation by JMJD2C and LSD1 promotes androgen receptor-dependent gene expression. *Nature cell biology* **9**, 347-353 (2007).
42. F. Lienert *et al.*, Genomic prevalence of heterochromatic H3K9me2 and transcription do not discriminate pluripotent from terminally differentiated cells. *PLoS genetics* **7**, e1002090 (2011).
43. N. A. Youngson, E. Whitelaw, Transgenerational epigenetic effects. *Annual review of genomics and human genetics* **9**, 233-257 (2008).
44. C. Lalancette, D. Miller, Y. Li, S. A. Krawetz, Paternal contributions: new functional insights for spermatozoal RNA. *Journal of cellular biochemistry* **104**, 1570-1579 (2008).

45. L. B. Russell, P. R. Hunsicker, W. L. Russell, Comparison of the genetic effects of equimolar doses of ENU and MNU: while the chemicals differ dramatically in their mutagenicity in stem-cell spermatogonia, both elicit very high mutation rates in differentiating spermatogonia. *Mutation research* **616**, 181-195 (2007).
46. K. L. Burr *et al.*, The effects of MSH2 deficiency on spontaneous and radiation-induced mutation rates in the mouse germline. *Mutation research* **617**, 147-151 (2007).
47. R. C. Barber *et al.*, The effects of in utero irradiation on mutation induction and transgenerational instability in mice. *Mutation research* **664**, 6-12 (2009).
48. M. Ihara *et al.*, Paternal poly (ADP-ribose) metabolism modulates retention of inheritable sperm histones and early embryonic gene expression. *PLoS genetics* **10**, e1004317 (2014).
49. D. T. Carrell, Elucidating the genetics of male infertility: understanding transcriptional and translational regulatory networks involved in spermatogenesis. *International journal of andrology* **31**, 455-456 (2008).
50. S. Chong *et al.*, Modifiers of epigenetic reprogramming show paternal effects in the mouse. *Nat Genet* **39**, 614-622 (2007).
51. O. Rechavi, G. Minevich, O. Hobert, Transgenerational inheritance of an acquired small RNA-based antiviral response in *C. elegans*. *Cell* **147**, 1248-1256 (2011).
52. L. R. Sabin, M. J. Delas, G. J. Hannon, Dogma derailed: the many influences of RNA on the genome. *Molecular cell* **49**, 783-794 (2013).
53. I. Ulitsky, D. P. Bartel, lincRNAs: genomics, evolution, and mechanisms. *Cell* **154**, 26-46 (2013).
54. F. Borges, R. A. Martienssen, Establishing epigenetic variation during genome reprogramming. *RNA biology* **10**, 490-494 (2013).
55. M. Sauvageau *et al.*, Multiple knockout mouse models reveal lincRNAs are required for life and brain development. *Elife* **2**, e01749 (2013).
56. V. Delmas, F. van der Hoorn, B. Mellstrom, B. Jegou, P. Sassone-Corsi, Induction of CREM activator proteins in spermatids: down-stream targets and implications for haploid germ cell differentiation. *Molecular endocrinology* **7**, 1502-1514 (1993).
57. M. Godmann *et al.*, Kruppel-like factor 4 is involved in functional differentiation of testicular Sertoli cells. *Developmental biology* **315**, 552-566 (2008).
58. B. Hogan, R. Beddington, F. Costantini, E. Lacy, *Manipulating the mouse embryo: a laboratory manual*. (Cold Spring Harbor Laboratory Press, New York, ed. 2nd, 1994).
59. S. McGraw *et al.*, Transient DNMT1 suppression reveals hidden heritable marks in the genome. *Nucleic acids research* **43**, 1485-1497 (2015).
60. Y. Xi, W. Li, BSMP: whole genome bisulfite sequence MAPping program. *BMC bioinformatics* **10**, 232 (2009).
61. R. Goodrich, G. Johnson, S. A. Krawetz, The preparation of human spermatozoal RNA for clinical analysis. *Archives of andrology* **53**, 161-167 (2007).
62. M. E. Ritchie *et al.*, limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic acids research* **43**, e47 (2015).
63. W. Huang da, B. T. Sherman, R. A. Lempicki, Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols* **4**, 44-57 (2009).

64. A. Alexa, J. Rahnenfahrer, topGO: enrichment analysis for gene ontology. *R package version*, (2010).
65. E. Heard, R. A. Martienssen, Transgenerational epigenetic inheritance: myths and mechanisms. *Cell* **157**, 95-109 (2014).
66. M. A. Carmell *et al.*, MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline. *Developmental cell* **12**, 503-514 (2007).

Acknowledgements

We thank Michael Stadler, Hélène Royo and Dimos Gaidatzis (FMI) for insightful advice on computational data analysis. We thank Brian K. Hall of Dalhousie University, Jason Tanny and Daniel Bernard of McGill University for their advice, Xavier Giner for technical expertise and the Laboratoire de transgénèse centre de recherché, Centre Hospitalier de l’Université de Montréal (CHUM), Canada for micro-injection and generation of the transgenic *hKDM1A* founders. We thank Tomi Pastinen, Guillaume Bourque, Maxime Caron and platform personnel of the McGill Epigenomics Mapping and Data Coordinating Centers in the McGill University, as well as Génomique Québec Innovation Centre for setting up the RRBS sequencing data analysis pipeline, and Francois Lefebvre for support with microarray analysis.

Financial support: This research was funded by the Canadian Institute of Health Research (SK and JT), Genome Quebec (SK and JT) and the Réseau de Recherche sur la Reproduction Québécoise, FQRNT (SK, JT), Boehringer Ingelheim Fond (SE), Swiss National Science Foundation (31003A_125386) (AP) and the Novartis Research Foundation (AP).

The data reported in this paper are available at Gene Expression Omnibus (GEO). Microarray data including both the CEL files and the transcript-level expression values are under accession number GSE66052 and ChIP-seq and nucleosome data under accession number GSE55471.

Author Contributions

K.S., S.E., M.G., R.L, S.M, J.T., A.H.F.M. P., and S.K conceived and designed the experiments.

K.S., S.E., M.G, R.L, C.L., T.C., S.M., J.T., M.S., A.H.F.M.P. and S.K performed the experiments and analyzed the data. M.H provided advice on the data analysis and the manuscript. K.S., J.T., A.H.F.M.P., and S.K. wrote the manuscript.

The authors declare that there is no conflict of interest that would prejudice the impartiality of this scientific work.

2.7 Figure Legends

Figure 2.1 Transgenerational and intergenerational definitions in maternal (a) and paternal (b) epigenetic inheritance.

In the pregnant female mouse (F_0), exposure to environmental factors (toxicants, nutrients, stress) can alter the soma and the germline of the F_1 . These are considered intergenerational effects. When the offspring F_1 are bred any phenotypic effects in F_2 are considered intergenerational as the F_2 originates from the F_1 germ cell that was exposed *in utero*. In maternal epigenetic inheritance, the F_3 are the first generation that have no direct connection to the exposure and are considered to have a transgenerational phenotype should abnormalities be detected. In males exposure of the F_0 includes his sperm that fertilizes the oocyte to produce the F_1 generation. Thus the F_1 are considered to be intergenerational. Breeding the F_1 to generate the F_2 results in the first unexposed generation. In contrast to females, in the male it is the F_2 and subsequent generations that can be subject to transgenerational phenotypes (65) (65).

Figure 2.2: Transgenic KDM1A expression in developing male germ cells impairs the development and survival of offspring.

(a) Human (hs) *KDM1A* is expressed under the control of the human elongation factor 1 alpha promoter (EF1 α) which has a truncated regulatory region that drives transcription in germ cells. Kozak (K) sequence, internal ribosomal entry site (IRES), green fluorescent protein (GFP), bovine growth hormone (BGH) polyadenylation signal. **(b)** Reverse transcriptase-PCR analysis shows that *KDM1A* transgene expression is restricted to testis. **(c)** *In situ* hybridization reveals *KDM1A* mRNA (blue) in spermatogonia (SG), spermatocytes (SPC) and round spermatids (RS). Elongated spermatids (ES) were negative. **(d, e)** Pedigrees **(d)** In line 1 the female heterozygous transgenic founder (TG^0) was bred to a C57BL/6 male to generate TG^1 male pups and their non-transgenic brothers (non TG^1). We subsequently bred heterozygous transgenic males (TG^{1-4}) to C57BL/6 females. We assayed the effects of paternal transgenic exposure on nonTG descendants over multiple generations (non TG^{3-5}). **(e)** In line 2 a female heterozygous transgenic founder (TG^0) was bred to a wild-type C57BL/6 male to generate TG^1 offspring. The first breeding of the founder resulted only in viable female transgenic offspring. Therefore, from the TG^2 generation onwards the transgene was passed through the father. For both lines we bred TG^{2-4} and non TG^{3-5} males to wild-type C57BL/6 females. Following E18.5 embryonic analysis from line 1 TG^3 sires and their non TG^3 brothers, we sacrificed the males and used their sperm for epigenomic analysis. **(f-g)** Reduced survival of offspring sired by TG^{2-3} and non TG^3 sires in line 1 and **(g)** TG^{2-4} and line 2 **(h)** in comparison to C57BL/6 sired offspring. Survivor curves with different letters are significantly different (log rank statistical test). n indicates the number of offspring. **(h-i)** postnatal day 6 TG sired pups with abnormal skin are shown in comparison to a normal pup sired by a control mating (asterix).

Figure 3: E18.5 fetuses sired by transgenic (TG) and non-transgenic descendants (nonTG) show a range of abnormalities.

(a) Phenotypic frequencies observed in embryonic day E18.5. Total Pregnancy Loss per Group [%] = (#corpora lutea - total embryos) per group / #corpora lutea per group x 100, Abnormal fetus per group [%] = (sum of abnormal fetuses per group / # embryos per group) x 100. Major abnormalities included: craniofacial, skeletal, edema, hemorrhagic gut, extra digits, missing eyes. Minor abnormalities included: pale, color disfigurement, growth retarded. **(b)** Normal E18.5 fetus sired by a C57BL/6 control father. **(c)** runt with limb abnormality (arrow) and hemorrhagic foci, **(d)** extra digit left hind paw, **(e)** blunted snout, underdeveloped forearms and digits (arrow), multiple hemorrhagic foci (wide arrow), **(f)** craniofacial abnormalities, edema, **(g)** hemorrhagic gut, underdeveloped limbs with failure in ossification **(h)-(k)** Severely malformed fetuses, **(j)** umbilical hernia. **(l)** Line 1 total pregnancy loss per group compared with total abnormalities / group. (Control n=298; TG² n=134, TG³ n=197, nonTG³ n=142, nonTG⁴ n=249, nonTG⁵ n=186) **(m)** Line 2 Total pregnancy loss / group compared with total abnormalities / group. (Control n=298; TG³ n=30, TG⁴ n=144, nonTG⁴ n=139) Total Pregnancy Loss per Group. [%] = (#corpora lutea - total embryos) per group / #corpora lutea per group x 100, Abnormal fetus per group [%] = (sum of abnormal fetuses per group / # embryos per group) x 100. Statistical test for Litter size: Student's T-test. Statistical test for total pregnancy loss and abnormalities per group: Fisher's exact test *p<0.05; **p<0.01; ***p<0.001.

Figure 4: Common skeletal abnormalities in pups sired by transgenic (TG) and nonTG³⁻⁴ descendants.

(red=bone, blue=cartilage) (a) control sired C56BL/6 pup at E18.5 and the corresponding skeletal stain. (b) nonTG4 sired fetus at E18.5 and the corresponding skeletal stain. The nonTG4 sired fetus had an abnormal craniofacial structure with reduced parietal ossification (arrow) and an extended cartilage deposit in the snout. Front limb digits were also underdeveloped (arrow). (c) Control C57BL/6 sired fetus (d) TG3 sired fetus with hypo-ossified parietal, frontal and supraoccipital bones and abnormal eye socket, (e) lack of temporal bones (arrow), hypo-ossified skull, abnormal snout, and (f) nonTG4 sired fetus missing mandible (agnathia). (Refer to Figure 2d for origin or generation and nomenclature and Supplementary Fig. 5 for a reference for normal skeletal anatomy and staining).

Figure 5: H3K4me2 occupancy is severely reduced at nucleosome containing TSS regions of CGI genes in TG³ sperm only.

(a) Snapshots of H3K4me2 occupancy at different loci in sperm of TG³ and WT sires (this study). Occupancies of H3K4me3, H3K27me3 and H3.3 in sperm and of Kdm1a in spermatocytes of WT animals were taken from (23) and (34). *Pdpk1* and *E2F6* represent loci marked by high and intermediate H3K4me3 occupancy, respectively, while *Gsc* represent H3K27me3 marked genes as observed in wild-type sperm. All three genes show reduced H3K4me2 levels at TSS sequences strongly enriched for H3.3-containing nucleosomes in WT sperm. We interpret the high histone modification occupancies at flanking sequences with low H3.3 nucleosome occupancy as the result of high ChIP efficiencies even at sites with low histone retention levels in the majority of sperm. (b) (c) Heatmaps showing reduced (b) or unchanged (c) H3K4me2 occupancy at TSS in TG³ sperm for three groups of genes classified according to the H3K4me3 and H3K27me3 states in WT sperm. The plots show CpG density, nucleosome, H3.3, H3.1/H3.2, H3K4me3 and H3K27me3 coverages in WT mice(23), and H3K4me2 coverages around TSS ($\pm 3\text{kb}$) in sperm of C57BL/6 WT, KDM1A TG³ and nonTG³ littermates and Kdm1a coverage in WT spermatocyte(34). (d), (e), and (f) show boxplots for the distributions of CpG percentage, for Kdm1a occupancy around TSS ($\pm 500 \text{ bp}$) and for H3.3 enrichment around TSS ($\pm 250 \text{ bp}$) for different groups of genes. Down and unchanged refer to genes with reduced or unchanged levels, respectively, of H3K4me2 in sperm of KDM1A TG³ males. Number of genes per group are as follows (from left to right): 507, 1121, 1437, 3840, 397, and 593.

Figure 6: Differential gene expression in two cell embryos as related to sperm chromatin content.

a) RNA content in TG³ and nonTG³ sperm is depicted by heatmap of selected genes as related to identified regions of H3K4me2 loss in TG³ sperm, or regions enriched for H3K4me3 or H3K27me3 in C57BL/6 sperm. **(b)** Venn-diagrams comparing number of RNA transcripts with significantly different levels of abundance in the sperm of TG3 and nonTG3 males with respect to control. **c)** The phenotypes observed in TG and nonTG offspring resemble those of mouse mutants for differentially expressed genes. **d)** Heatmap depicting selected RNAs that differed from control sired embryos to those sired by TG8, or nonTG8 (F2 generation embryos), as related to identified regions of H3K4me2 loss in TG3 sperm, or sperm enriched regions for H3K4me3 or H3K27me3. **e)** Venn-diagrams depicting differential gene expression in two cell embryos sired by TG8, nonTG8, or controls as related to identified regions of H3K4me2 loss in TG3 sperm, or sperm enriched regions for H3K4me3 or H3K27me3.

Figure 7. Summary of Model

Disruption of histone methylation in developing sperm by KDM1A transgene over-expression from one generation severely impaired embryo development and survivability of offspring.

These defects were transgenerational and occurred in nonTG descendants in the absence of KDM1A germ line expression suggesting regions in the nonTG germline escape epigenetic reprogramming. Developmental defects in offspring were associated with altered RNA content in sperm and gene expression in embryos.

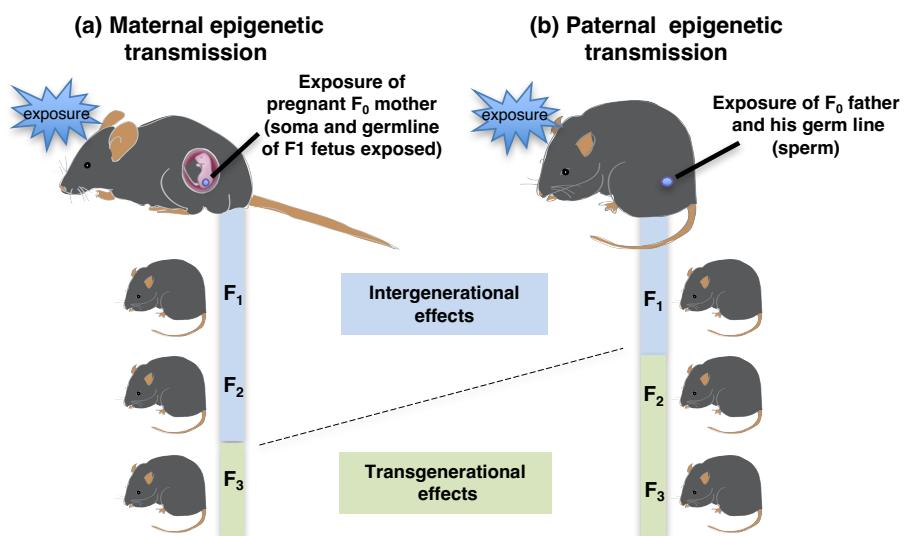


Figure 1

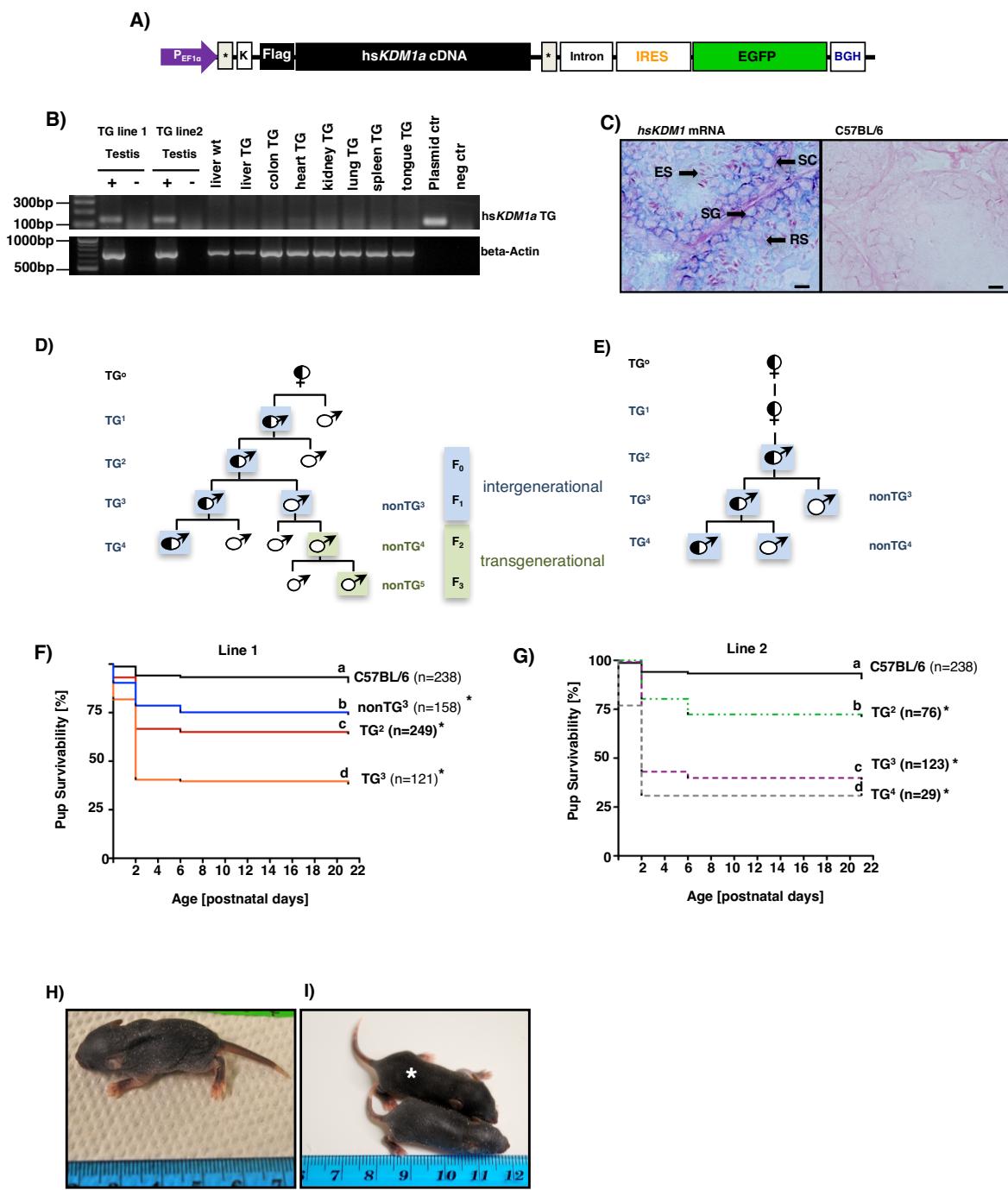


Figure 2

A)

Sire / filial generation (n sires/n litters)	Line 1 TG ² (F ₀) (8 / 15)	Line 1 TG ³ (F ₁) (12/24)	Line 1 nonTG ³ (F ₁) (10/18)	Line 1 nonTG ⁴ (F ₂) (16/30)	Line 1 nonTG ⁵ (F ₃) (14/20)	Line 2 TG ³ (F ₁) (2/4)	Line 2 TG ⁴ (F ₂) (9/16)	Line 2 nonTG ³ (F ₁) (9/17)	Control C57BL/6 (21/32)
Fetuses (n)	134	197	142	249	186	30	144	139	298
Litter size ± stdev	8.93 ± 1.49	8.20 ± 1.66	7.88 ± 1.94	8.3 ± 2.12	9.3 ± 1.89	7.5 ± 4.65	9.0 ± 1.37	8.18 ± 2.10	9.31 ± 1.42
Total preg. loss/group [%]	15.2	19.6**	20.2**	20.7***	12.9	25.0*	22.2***	22.8***	10.8
Abnormal fetuses [%]	25.3***	7.61**	11.3***	6.42**	3.23	10.0*	13.9***	10.1***	1.68
[Abnormal/total]	[34/134]	[15/197]	[16/142]	[16/249]	[6/186]	[3/30]	[20/144]	[14/139]	[5/298]
[Major abnormal/total]	[30/134 (22.4%)]	[12/197 (6.09%)]	[14/142 (9.86%)]	[12/249 (4.82%)]	[5/186 (2.69%)]	[3/30 (10.0%)]	[17/144 (11.8%)]	[12/139 (8.63%)]	[5/298 (1.68%)]
[Minor abnormal/total]	[4/134 (2.99%)]	[3/197 (1.52%)]	[2/142 (1.41%)]	[4/249 (1.61%)]	[1/186 (0.54%)]	[0/30 (0.00%)]	[3/144 (2.08%)]	[2/139 (1.44%)]	[0/298 (0.00%)]

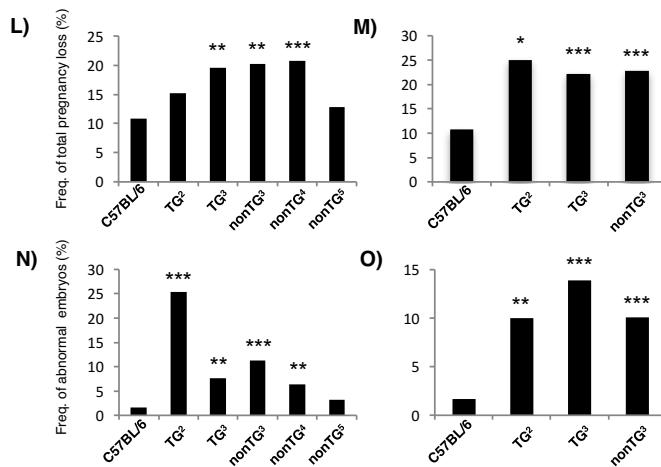


Figure 3

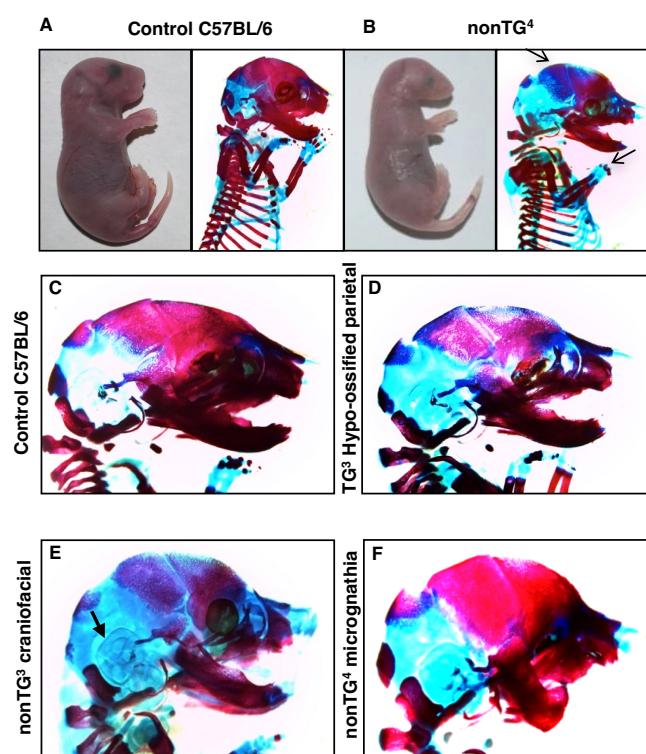


Figure 4

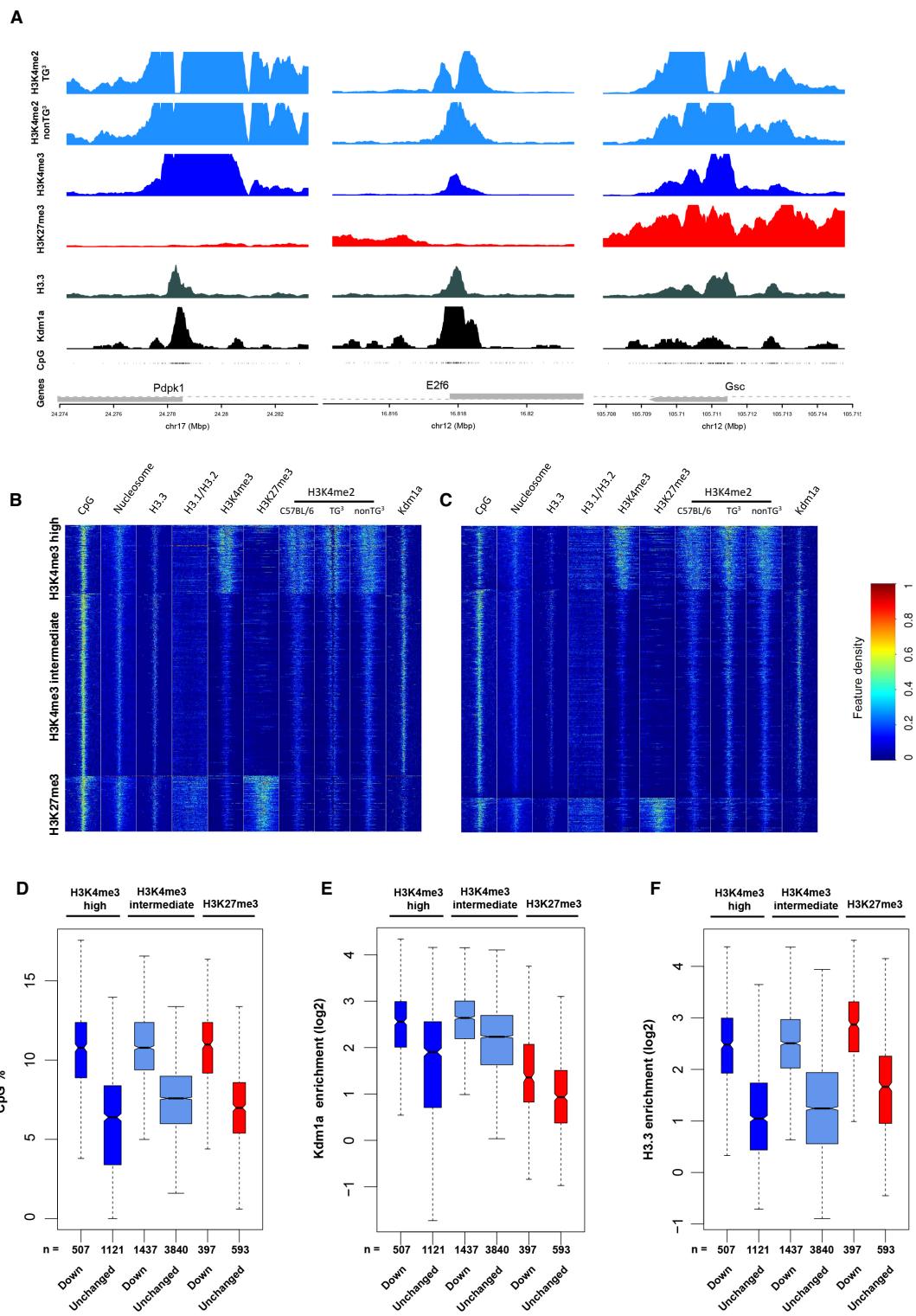


Figure 5

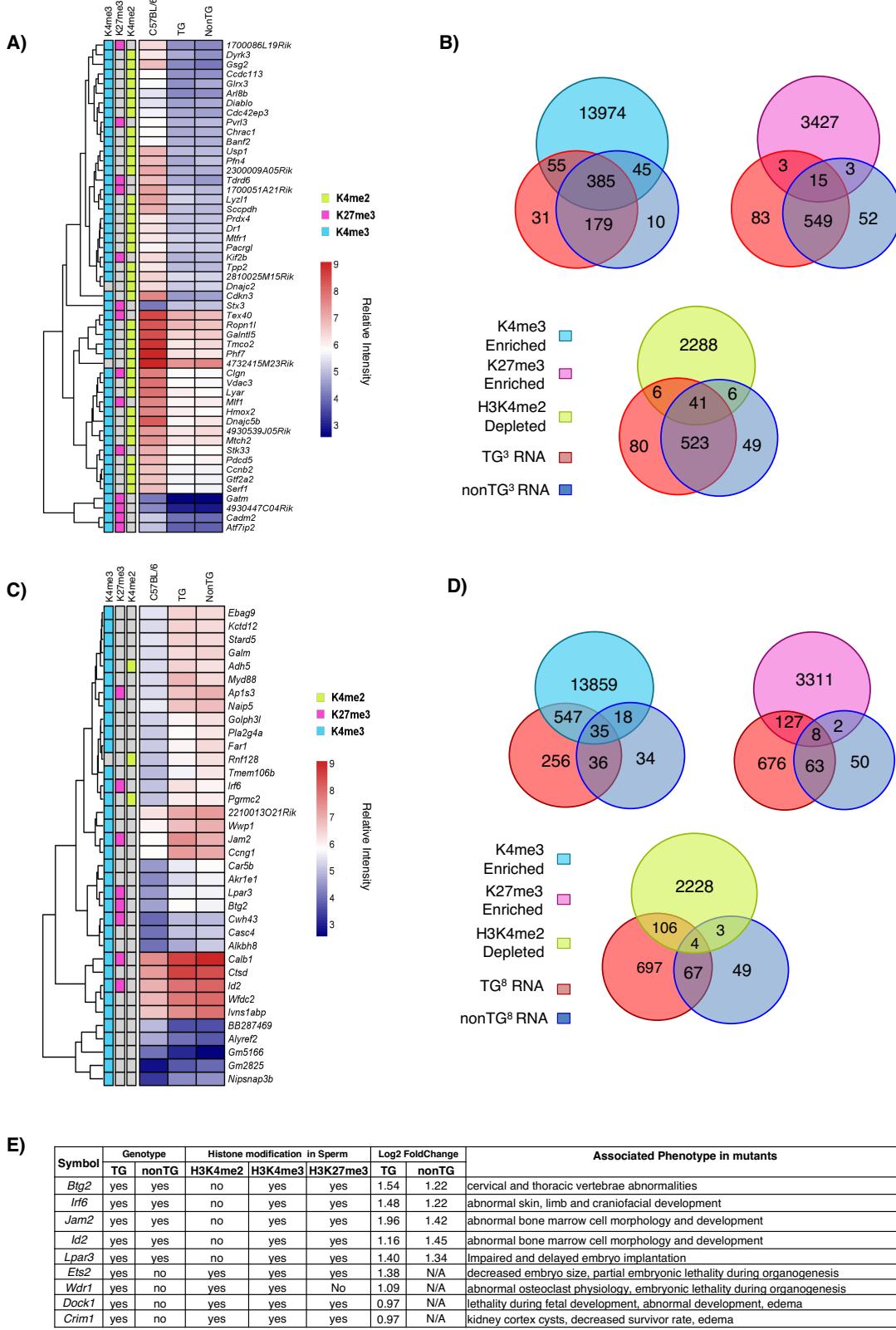


Figure 6

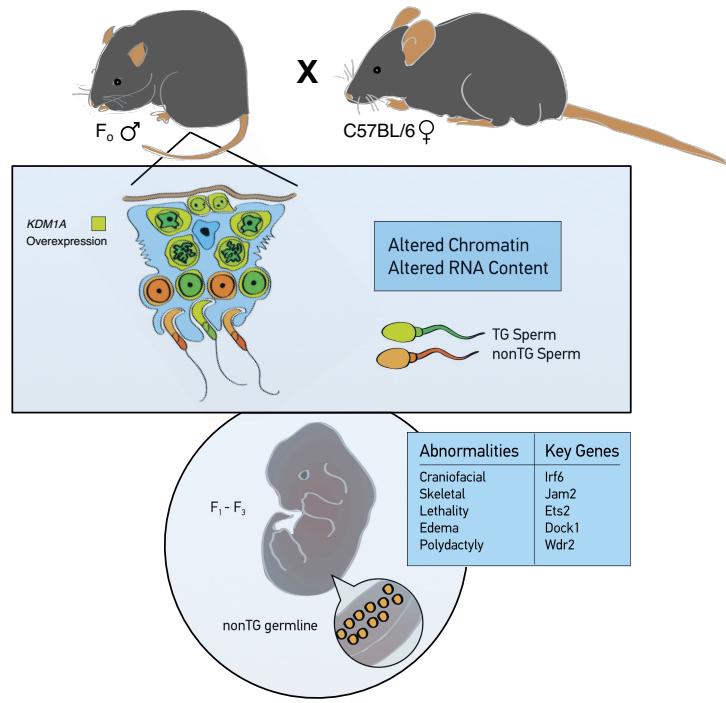


Figure 7

2.8 Appendix I

Supplementary Figure 2.1. Transgenic copy number was determined by Southern blot and expression in embryonic tissues was not detected. (a) Transgenic construct and southern-blot probe location (b) Copy number was assessed by Southern blot followed by densitometry using spiked samples for transgenic (TG) lines 1 and 2. (c) Localization of the transgene marker GFP in a cryostat cross-section from adult testis. (d-e) Total RNA from transgenic embryos at E18.5 was assessed for transgene expression from testis, liver, kidney, gut, lungs, heart, and brain. PCR amplification of cDNA with primers against hsKDM1 did not show transgene expression in these tissues. Positive controls include cDNA from adult transgenic testis and purified plasmid. Gel image was acquired using Qiagen's Qiaxcel Advanced.

Supplementary Figure 2. (a) Sperm counts for line1 TG F2 (TG²; n=10), nonTG (nonTG²; n=5), and for line2 TG F₂ (TG²; n=4), nonTG (nonTG²; n=3), and control (n=7). (b) Body weight (BW), (c) Testes weight to body weight (TW:BW), (d) Epididymal weight to body weight ratio (EW:BW), for TG and nonTG animals of Line1 (F₁ = TG¹ n=8; F₂ = TG² n=26, nonTG² n=9; F₃ = TG³ n=11, nonTG³ n=9), and Line2 (F₁ = TG¹ n=3, nonTG¹ n=3; F₂ = TG² n=6, nonTG² n=6; F₃ = TG³ n=3; F₄ = nonTG⁴ n=4). (e) Fetus bodyweight and (h) placenta weight at embryonic day 18.5 (E18.5) for embryos sired by control C57/Bl6 (n=10), TG (n=12) and nonTG (n=11) fathers. Differences between TG and nonTG groups were compared to control with student's T-test corrected for multiple comparisons by Bonferroni corrections. Values depicted by mean ± standard deviation.

Supplementary Figure 2.3: (a) Panels i, ii: Histopathological analysis of spermatogenesis was performed on testis cross-sections from generations TG²⁻³ and nonTG²⁻³ in comparison to controls. Panels iii, iv: DNA damage did not differ in TG and nonTG germ cells as assessed by normal immunohistochemical staining for phosphorylated gamma H2AX on testis cross-sections from TG²⁻³ (n=4) versus C57BL/6 (n=3). (b) RNA from whole testis of C57BL/6 (n=2), TG (n=3) and nonTG (n=3) were quantified via qPCR to measure the dCT between Line1 5'UTR (forward: 5'-GGCGAAAGGCAAACGTAAGA-3'; reverse: 5'-GGAGTGCTGCGTTCTGATGA-3')(66) and *Gapdh* gene (forward: 5'-ACTTTGGCATTGAAGGGCTG-3'; reverse: 5' TGGAAGAGTGGAGTTGCTGTTG-3') (Students t-test, p. = 0.25.) (c) Number of tunnel positive tubules and (d) cells per 100 tubules in testes from adult C57BL/6 (n=4) and TG¹ males (n=4).

Supplementary Figure 2.4. Variable offspring outcome in litters sired by individual males and analyzed between PND 0-21. Genotypes and generations are: a) C57BL/6 b) Line 1 TG² c) Line 1 TG³ d) Line 1 nonTG³ e) Line 2 TG², f) Line 2 TG³, and g) Line 2 TG⁴

Supplementary Figure 2.5. Wild-type C57BL/6 skeletal anatomy and normal staining patterns at embryonic day 18.5. (a-d) Alcian blue stains cartilage (blue), and alizarin red stains ossified bone (red).

Supplementary Figure 2.6. (a) Pair-wise scatter plots show the correlation of total library size normalized read counts (log2) at the region ± 250 bp surrounding TSS for H3K4me2 in sperm of C57BL/6 males (two replicates), KDM1a TG³ and nonTG³ littermates. Genes which have more

than or equal to 5 reads (log2 scale) in the regions analyzed in both replicates of H3K4me2 BL6 are shown in red. (b) Density plot showing the ratio of H3K4me2 in sperm of KDM1A TG³ males over C57BL/6 males in regions \pm 250 bp surrounding transcriptional start sites (TSS). The plot area is divided into 3 zones by green lines. The zone on the left specifies the genes with reduced H3K4me2, the middle zone specifies the genes with unchanged H3K4me2, and the right zone shows the genes gaining H3K4me2. For the specification of the green lines, please see materials and methods.

Supplementary Figure 2.7. Heatmaps showing nucleosome occupancy around TSS (\pm 3kb) in sperm of C57BL/6 WT, KDM1A TG³ and nonTG³ littermates. Genes were classified in three groups according to H3K4me3 and H3K27me3 states in WT sperm(17). H3K4me2 occupancies around TSS of corresponding chromatin samples are shown in Figure 5b and 5c. Mono-nucleosomes isolated from MNase digested sperm chromatin (input).

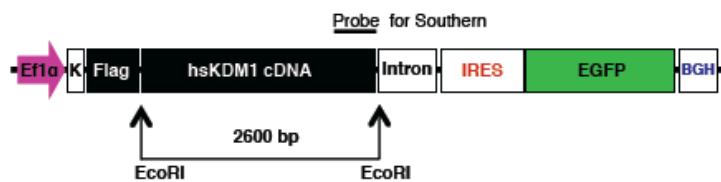
Supplementary Figure 2.8. DNA methylation analysis by quantitative Sequenom MassARRAY, Mean \pm SEM of methylation (%) at individual CpGs for regions of high H3.3, CpG density and major depletion of H3K4me2 (ratio of TG vs WT < -3). No differences were detected between methylation levels in transgenic (TG³), non-transgenic (nonTG³) and controls sperm.

Supplementary Figure 2.9. Analysis of RRBS in sperm of TG³, nonTG³ and control animals.

Heatmap of Pearson's correlation coefficients between all pairs of samples, calculated using percent methylation of CpGs with a minimal coverage of 10 reads ($n = 0.64$ million). Similar samples were grouped by hierarchical clustering as shown by dendrogram. To identify genotype associated methylation, the methylation status of each individual CpG was summarized in a contingency table with the rows corresponding to methylated and unmethylated reads, and columns to three genotypes. A P-value was calculated for each CpG using Fischer's exact test. The figure shows the number of P-values below a given cut-off (p_0) for the observed data (black dots) and expected by chance (red line). At $p_0 = 1 / (\text{number of CpG})$ we find 629 significant CpGs. (c) Heatmap of percent methylation for 629 CpGs with significant genotype associations. CpGs with similar methylation levels across multiple samples were grouped by hierarchical clustering as shown by dendrogram. The average difference of methylation between groups is typically small (< 20%). Furthermore, there is no overrepresentation of methylation patterns that correlate with transgenerational phenotypes.

Supplemental Figure 2.1

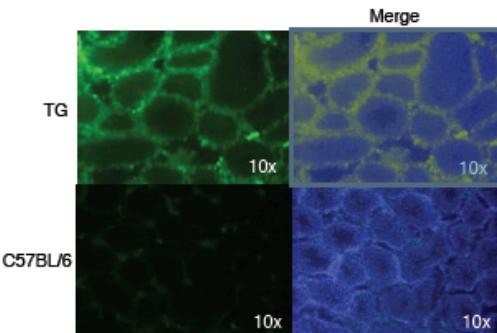
A)



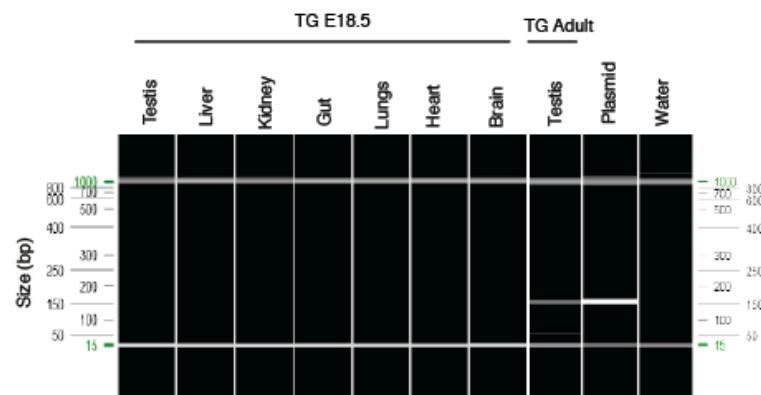
B)



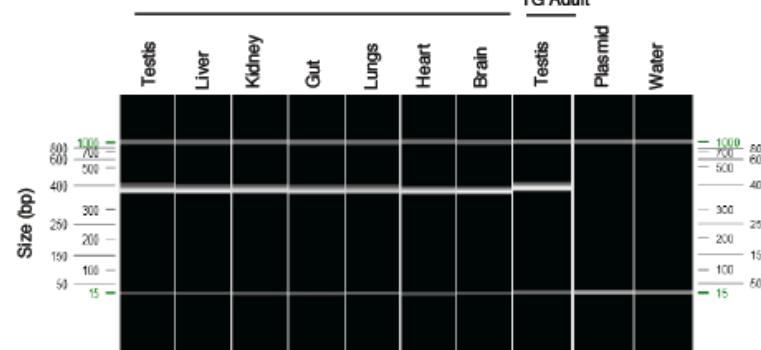
C)



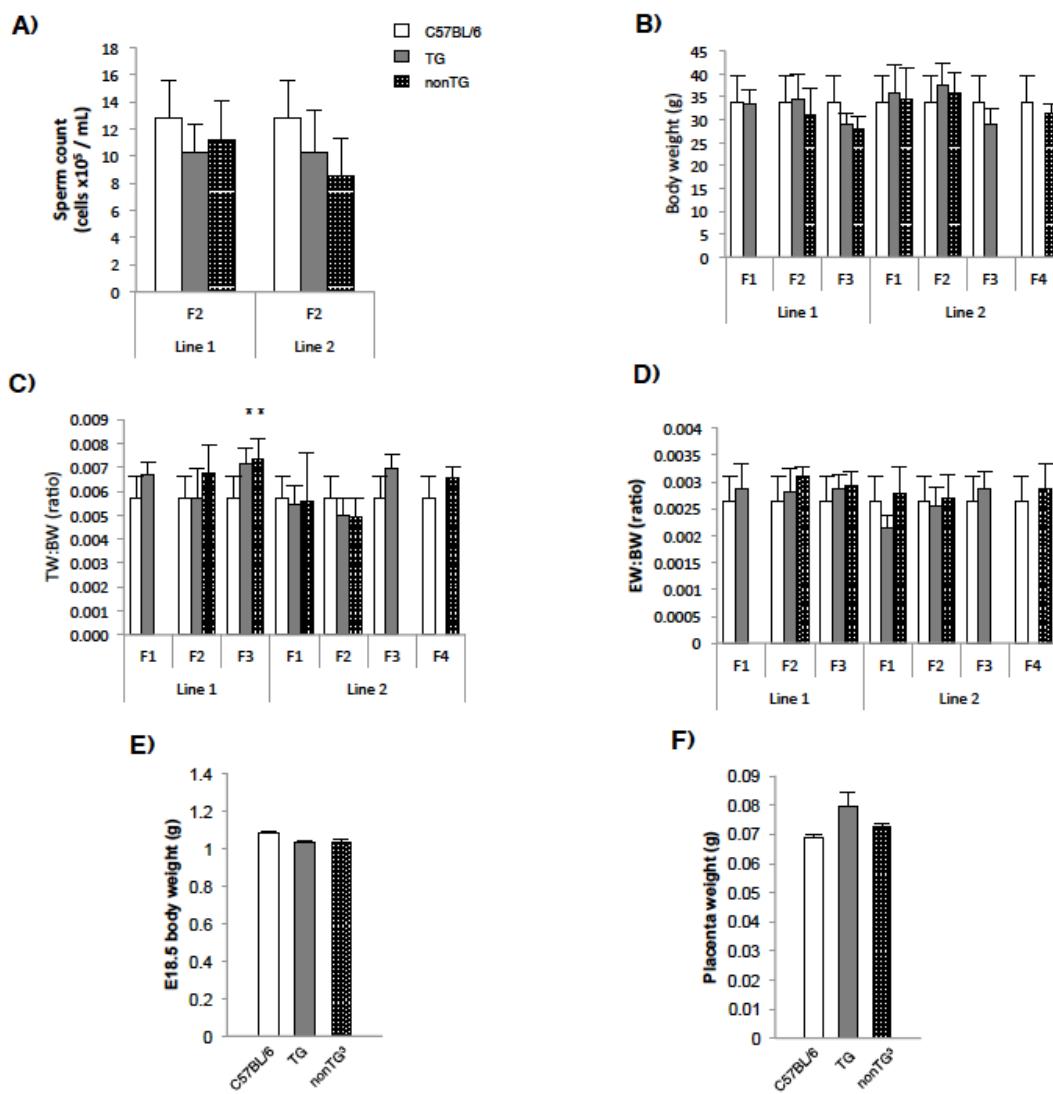
D) hsKDM1A



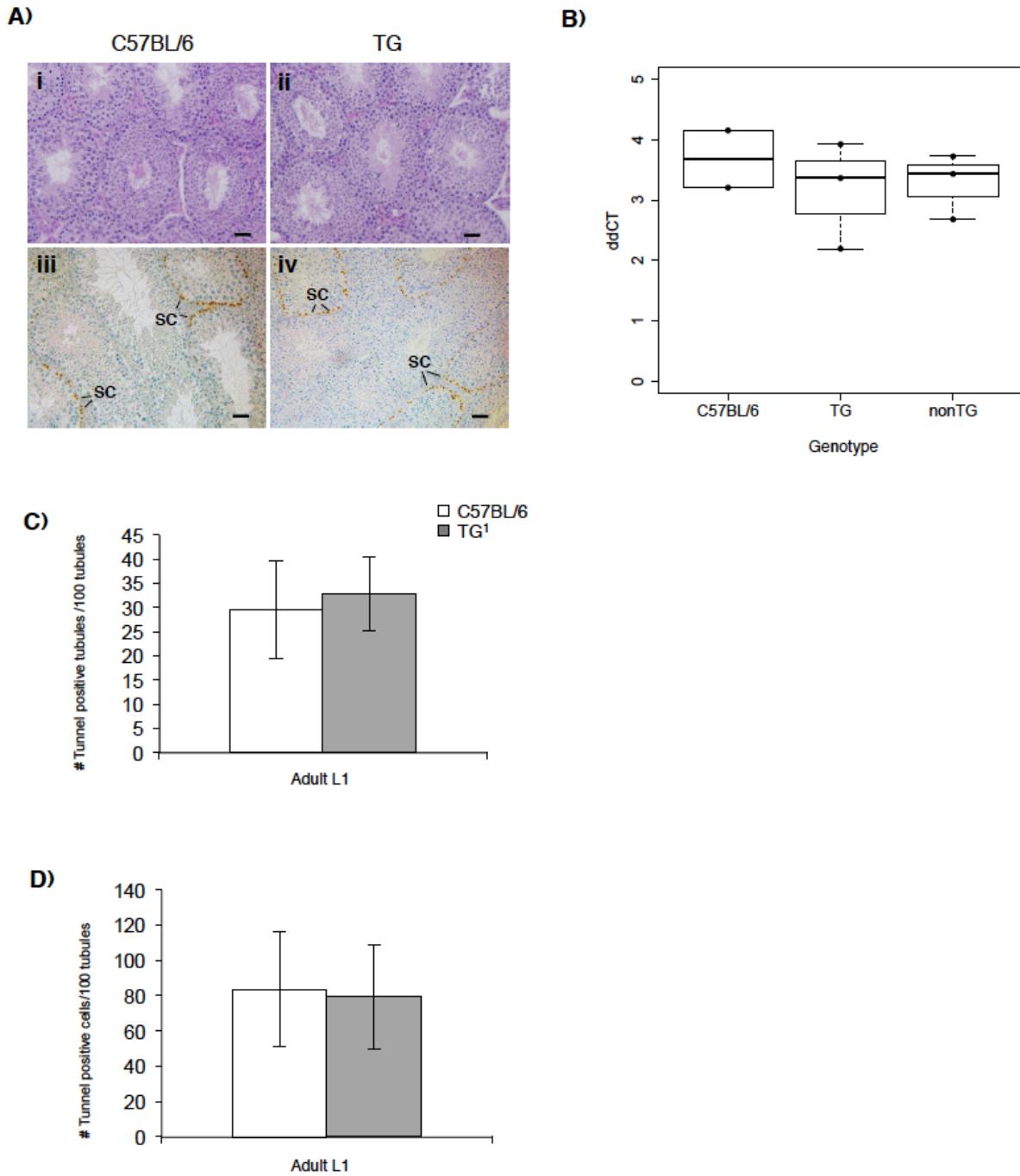
E) Gapdh



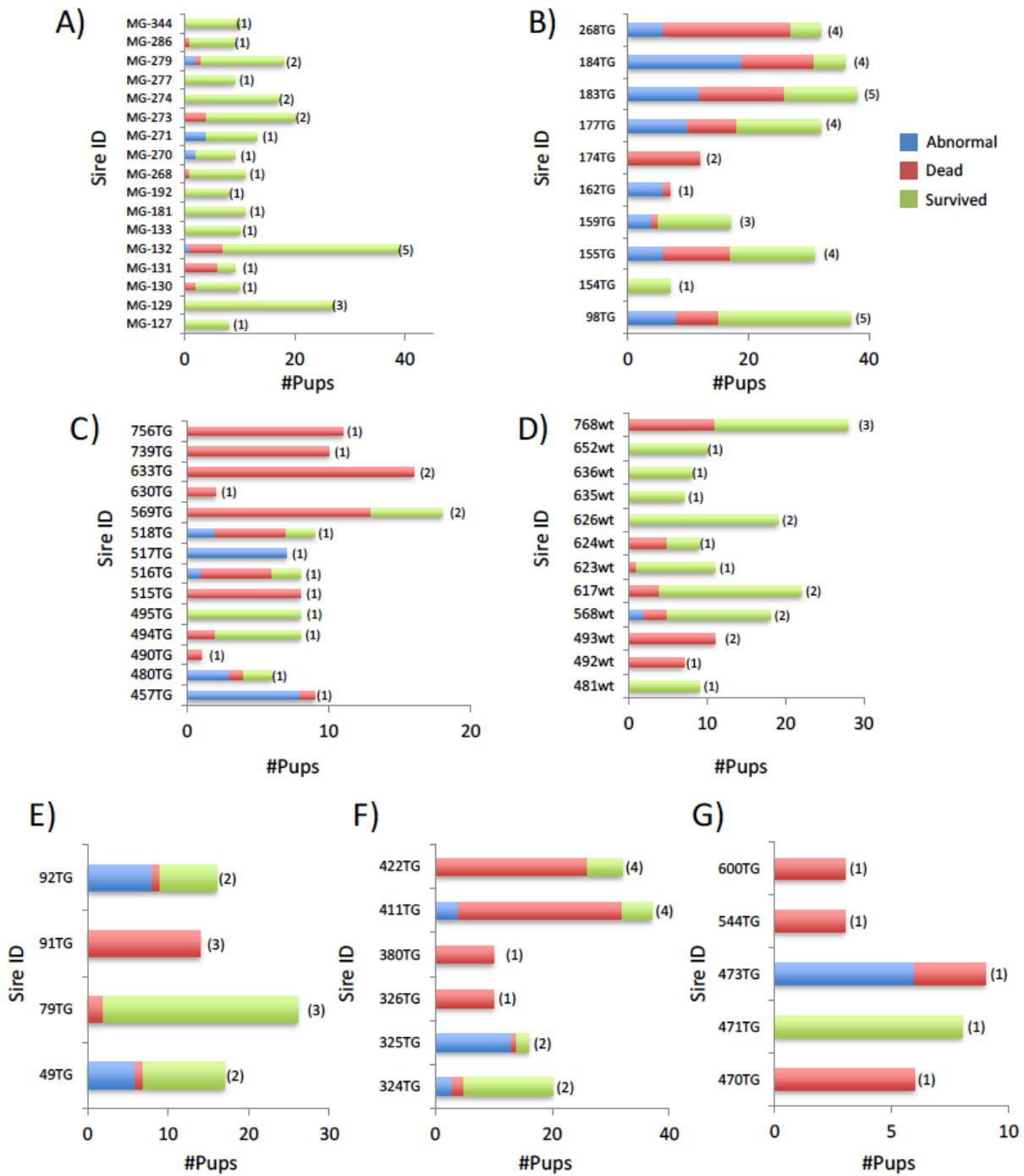
Supplemental Figure 2.2



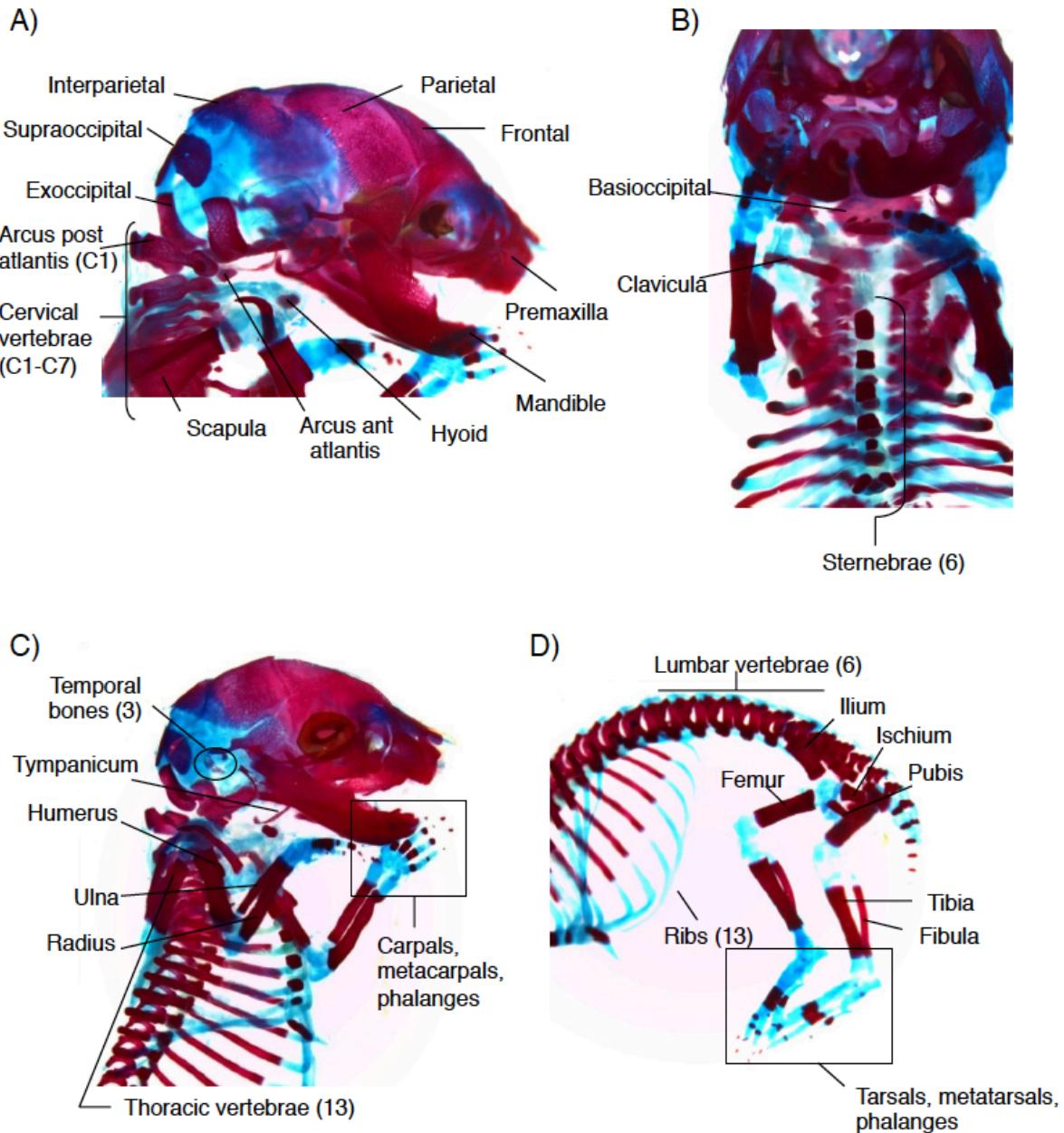
Supplemental Figure 2.3



Supplemental Figure 2.4

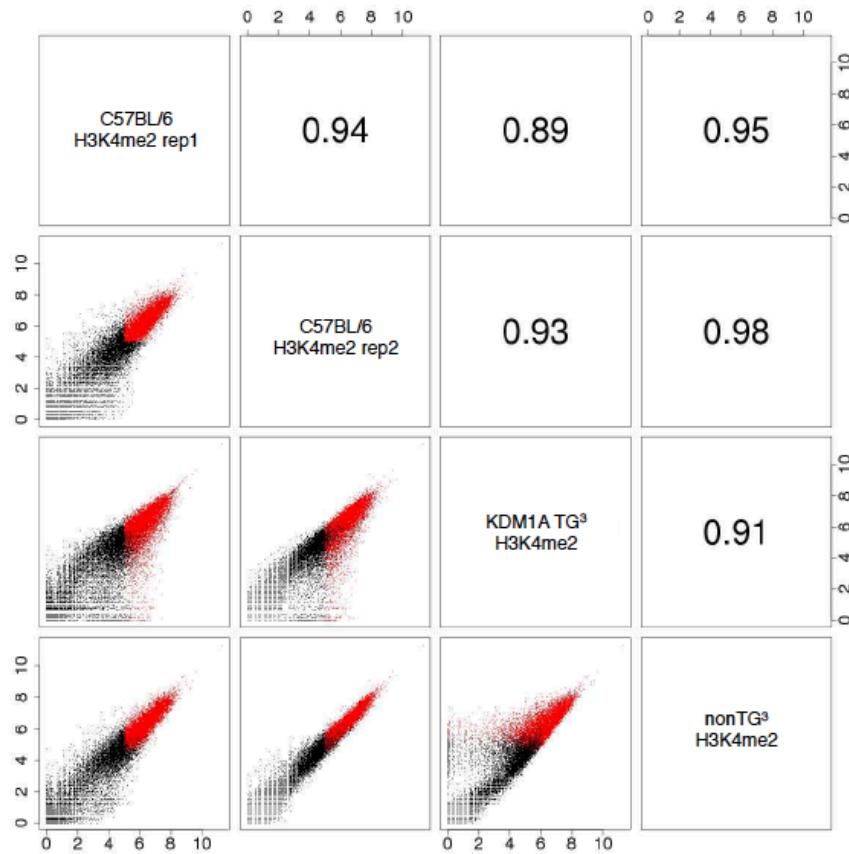


Supplemental Figure 2.5

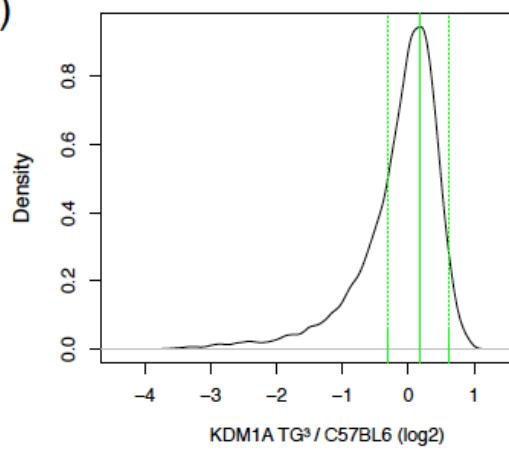


Supplemental Figure 2.6

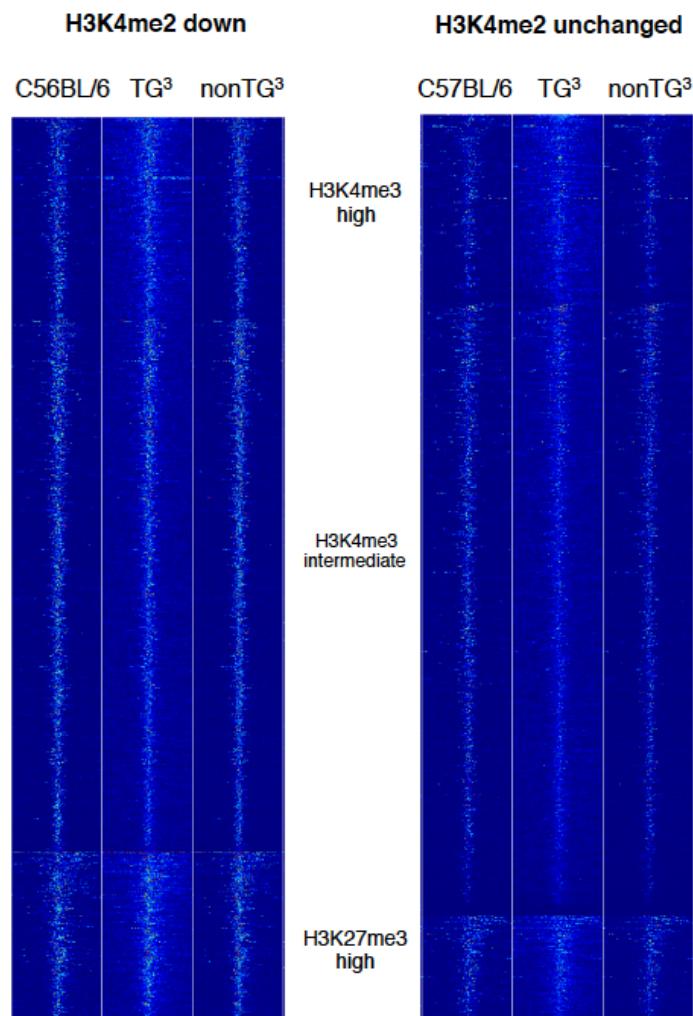
A



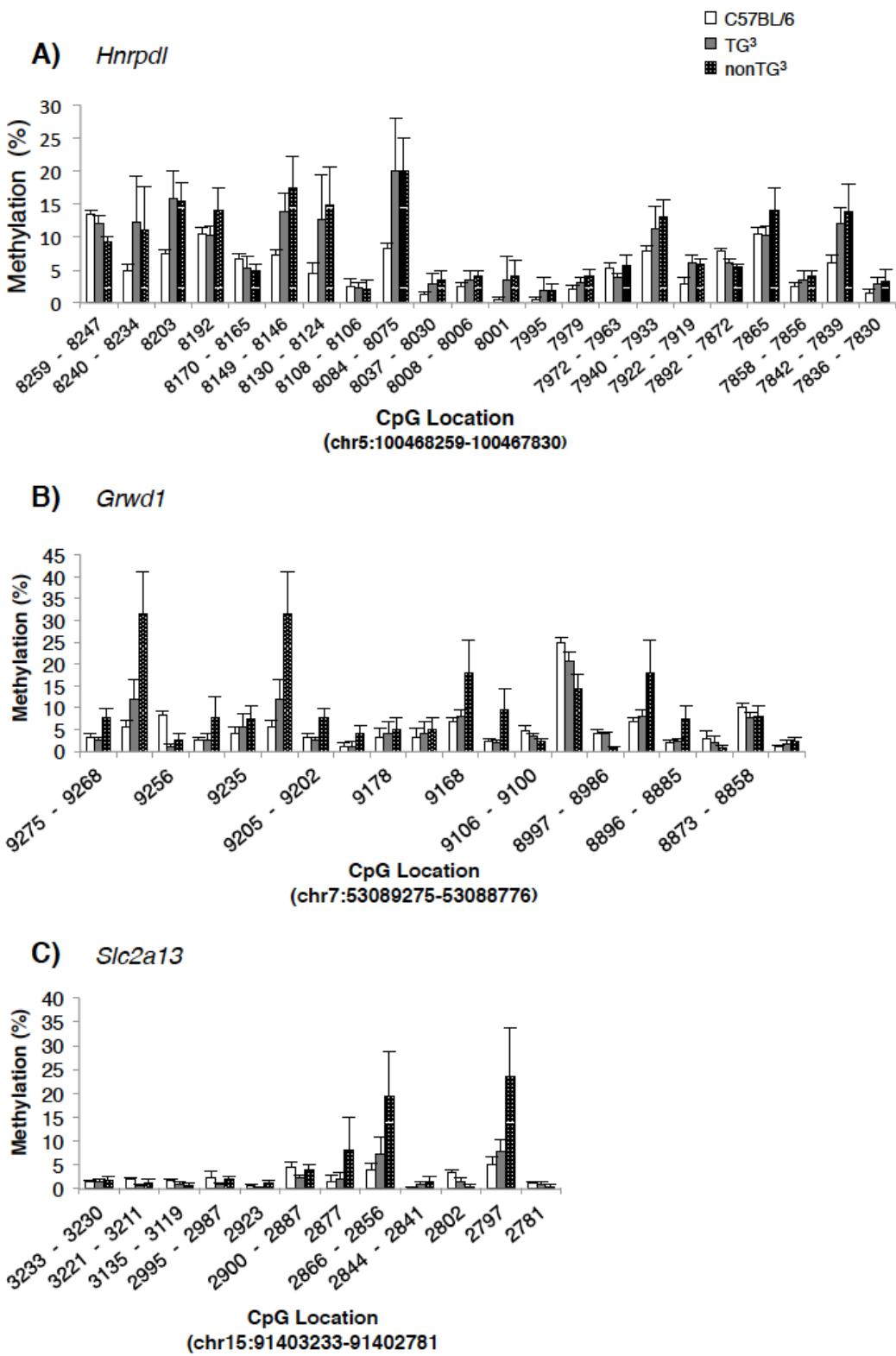
B)



Supplemental Figure 2.7



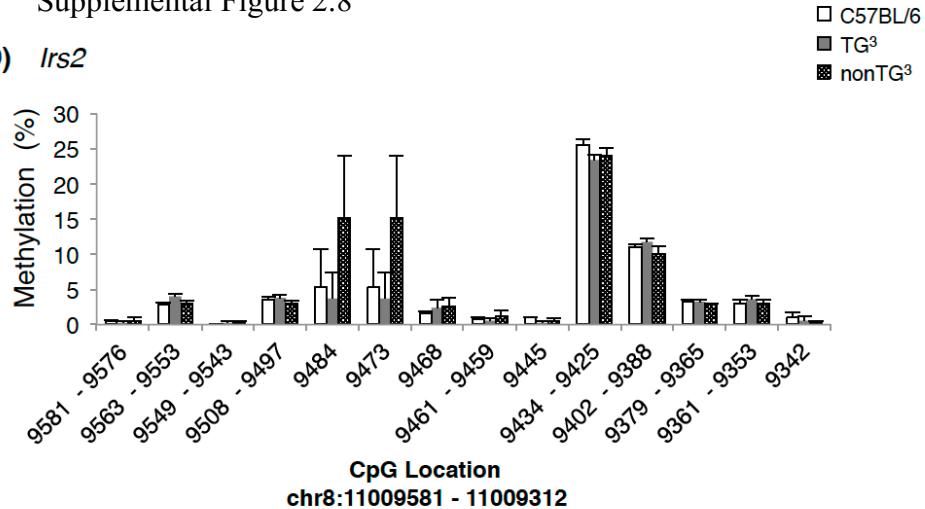
Supplemental Figure 2.8



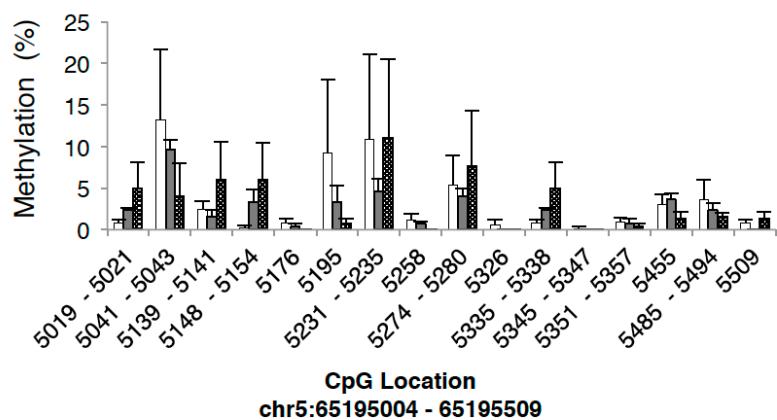
Supplemental Figure 8

Supplemental Figure 2.8

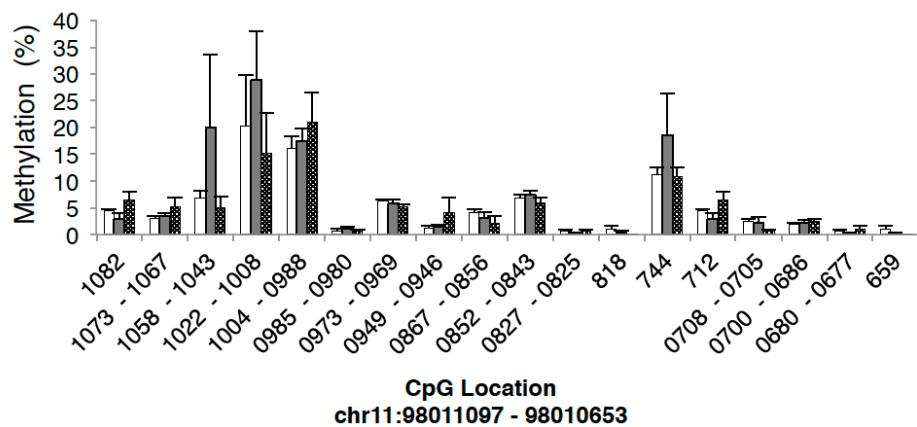
D) *Irs2*



E) *Klf3*



F) *Fbxl20*



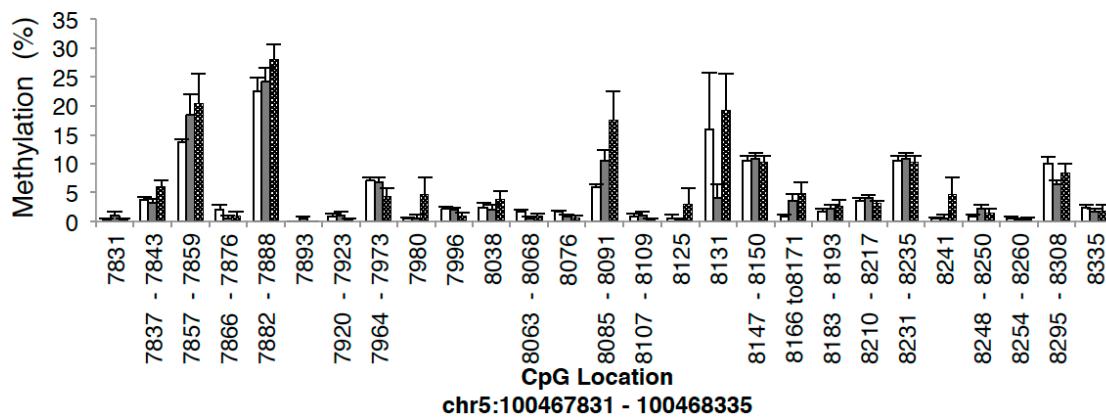
Supplemental Figure 2.8

G) *Enoph1*

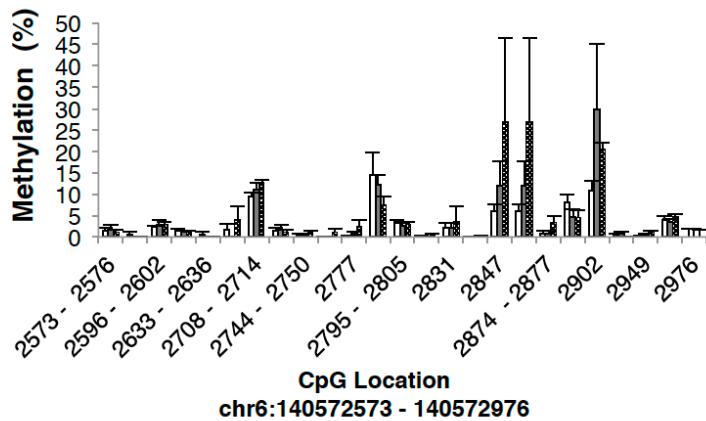
C57BL/6

TG³

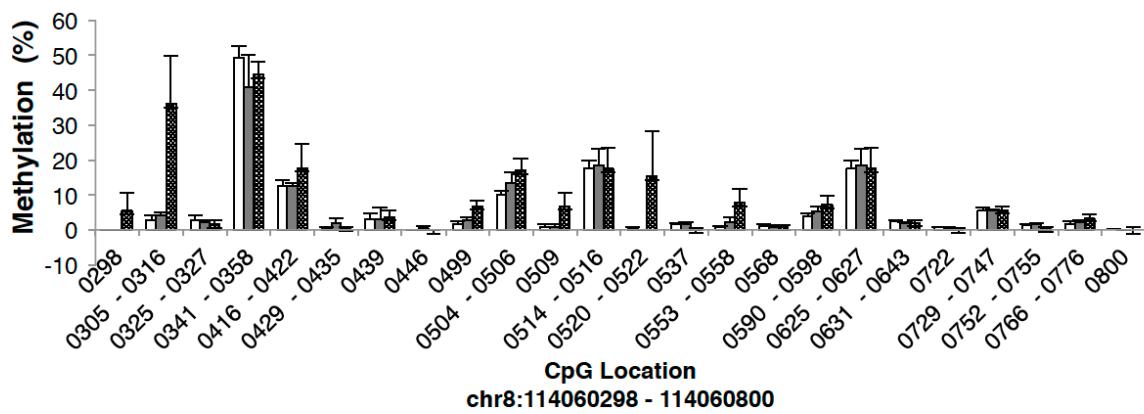
nonTG³



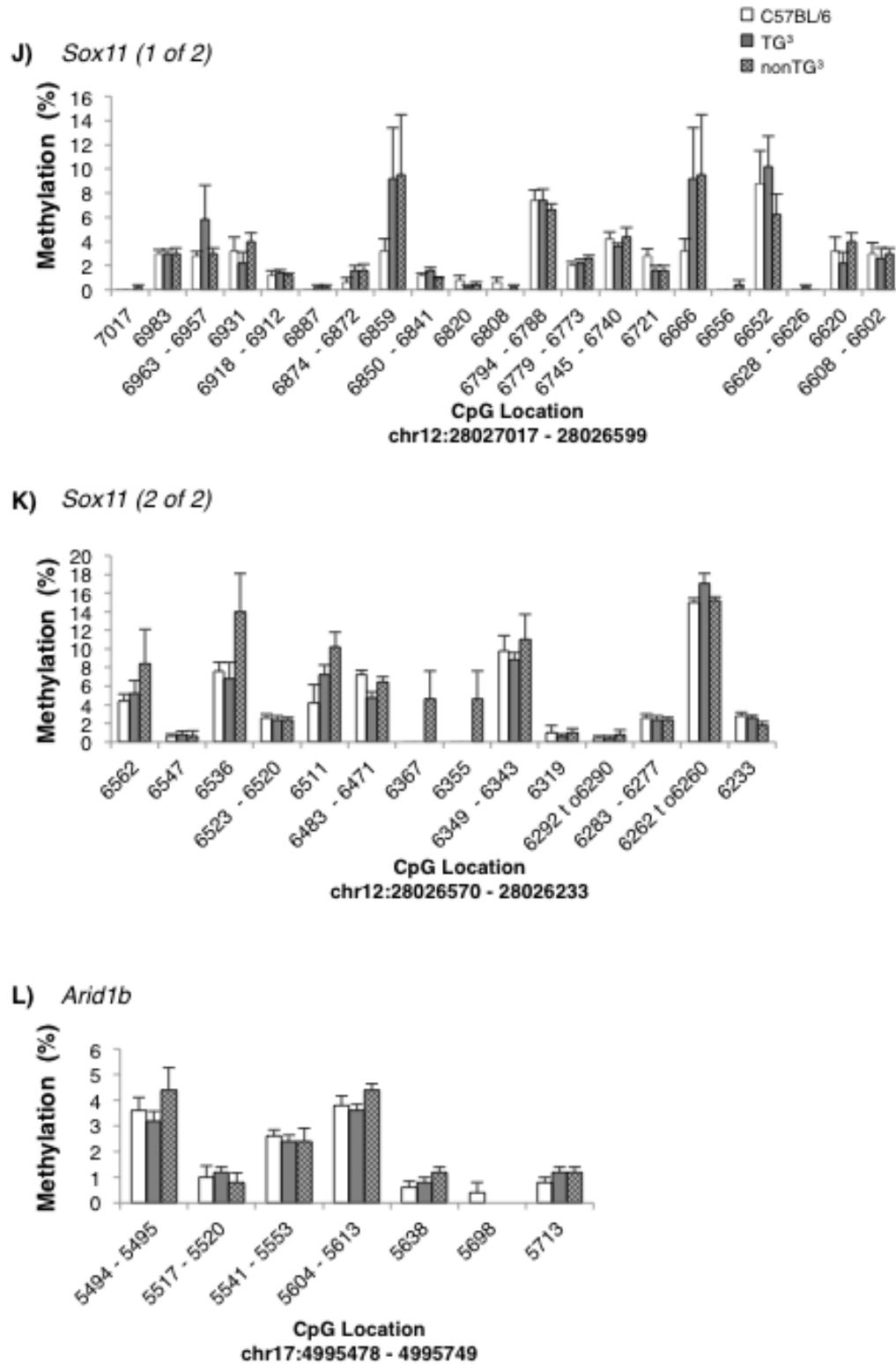
H) *Aebp2*



I) *Znrf1*

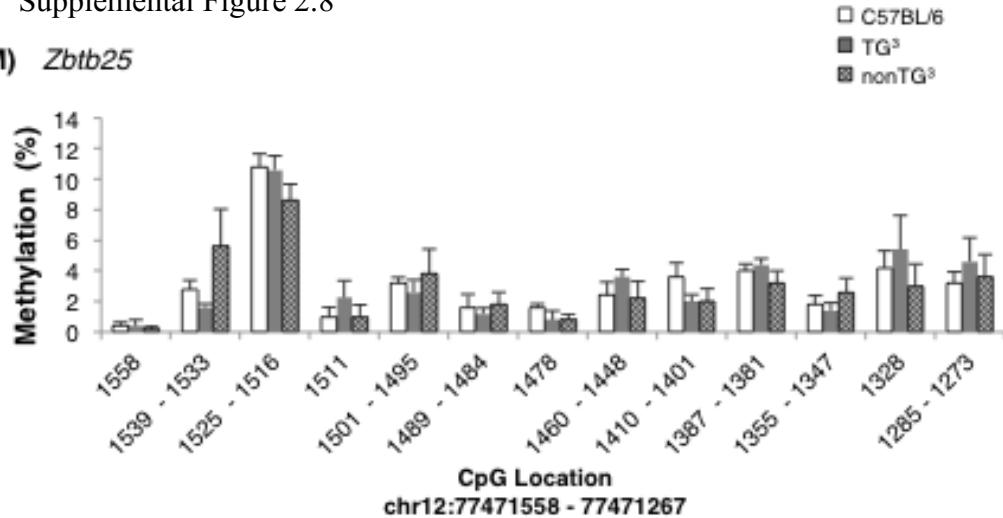


Supplemental Figure 2.8

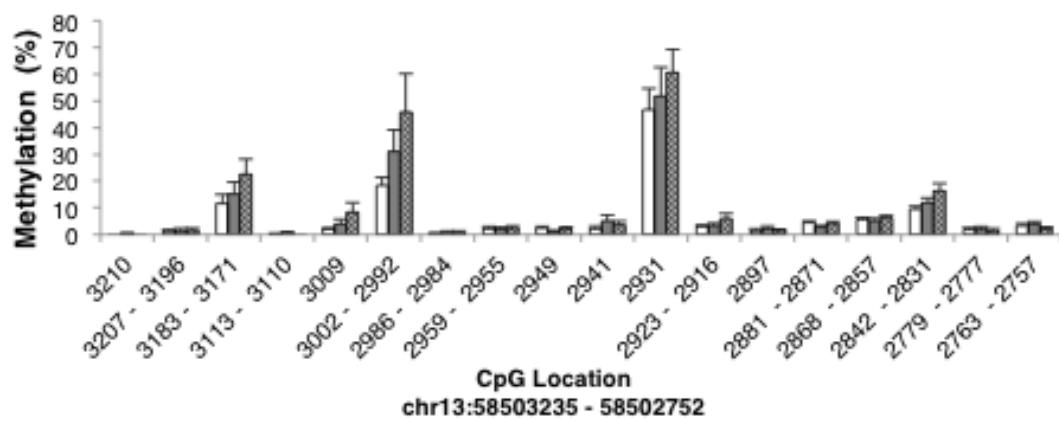


Supplemental Figure 2.8

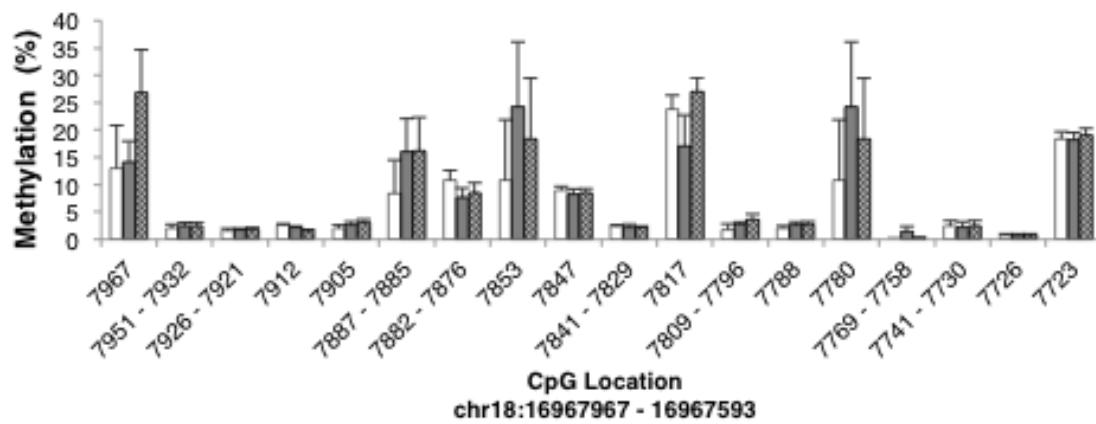
M) *Zbtb25*



N) *Hnmpk*

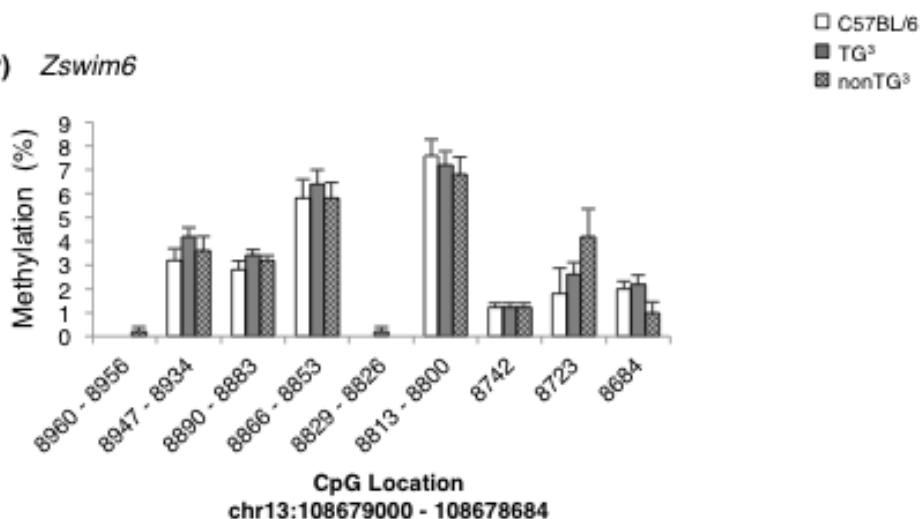


O) *Cdh2*

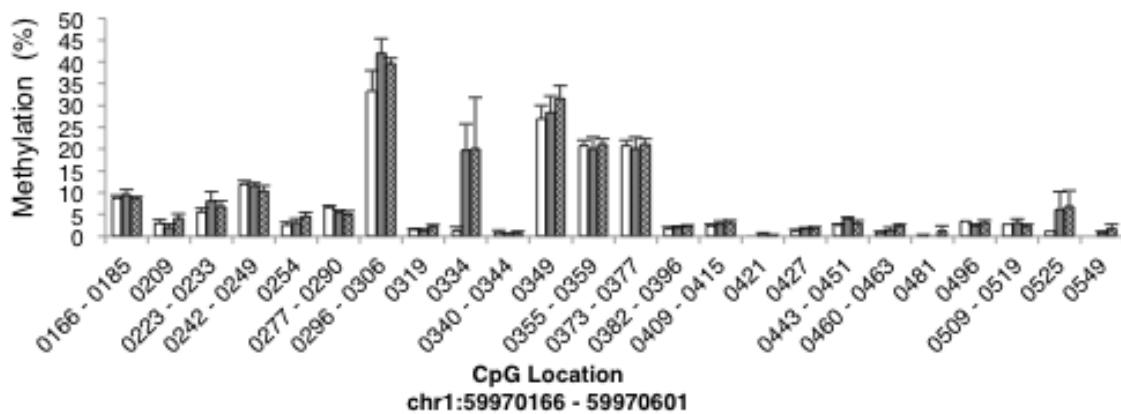


Supplemental Figure 2.8

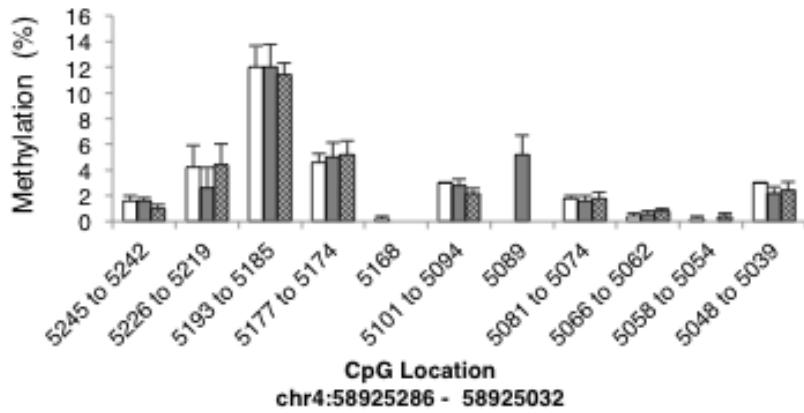
P) *Zswim6*



Q) *Fam117b*



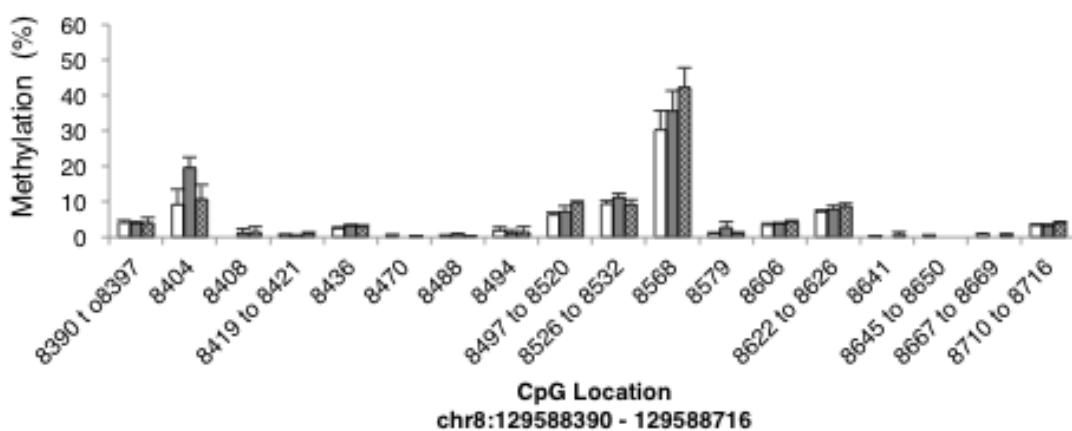
R) *A1314180*



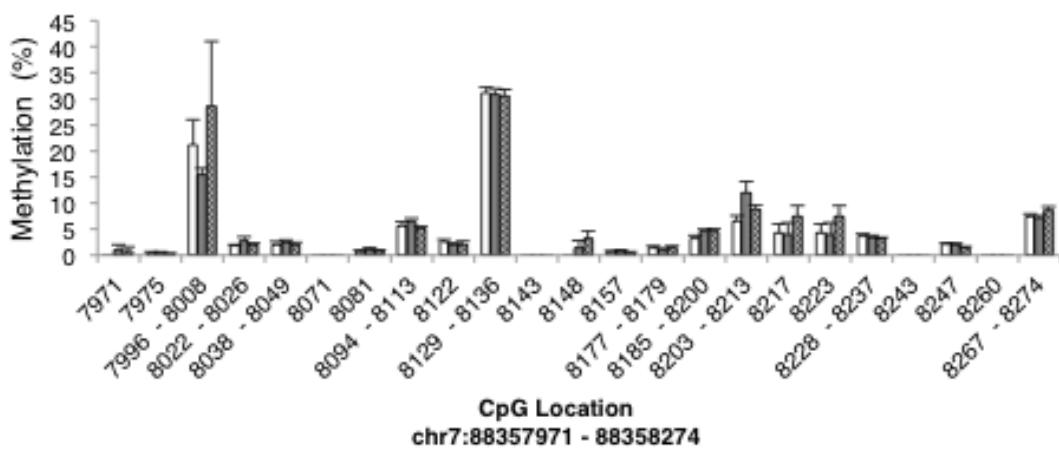
Supplemental Figure 2.8

□ C57BL/6
■ TG³
▨ nonTG³

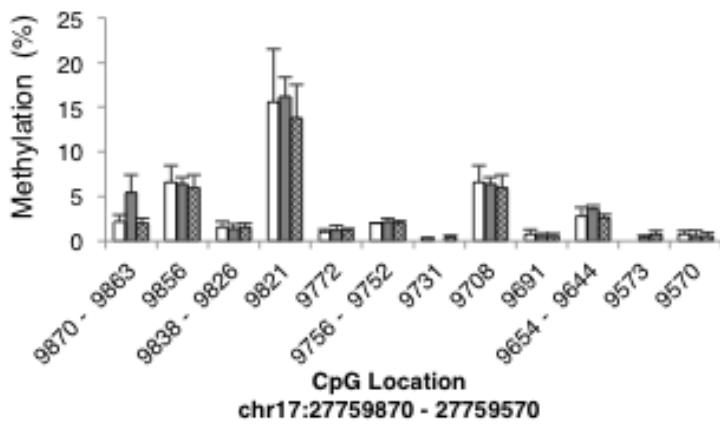
S) *Pard3*

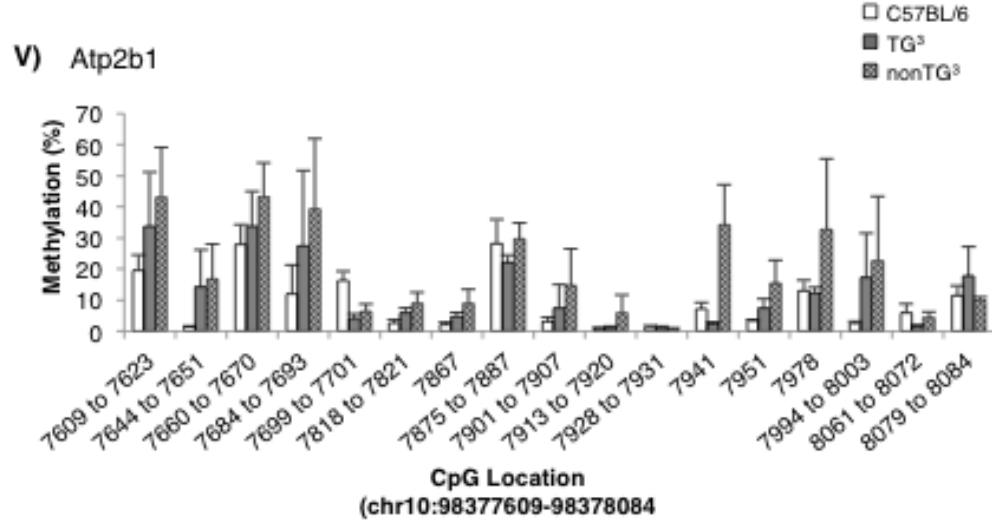


T) *Pde8a*



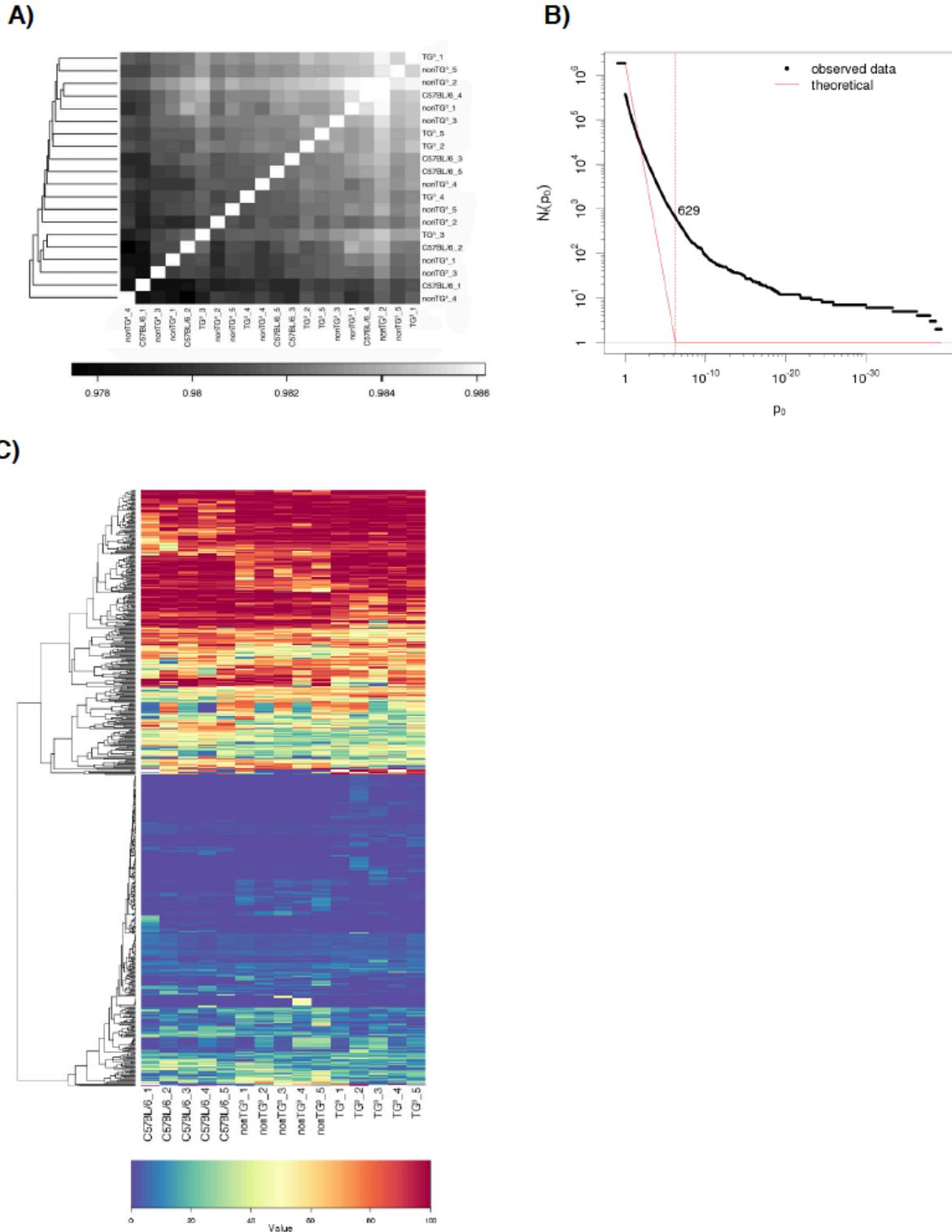
U) *Nudt3*





Supplemental Figure 8

Supplemental Figure 2.9



Supplemental Figure 9

2.9 Connecting Text

Although this work has described reduced levels of H3K4me2 in transgenic (TG) mouse sperm we did not detect persistent histone methylation changes in the non-transgenic (nonTG) littermates. Moreover, the failure to detect any changes in sperm DNA methylation at CGIs or regions of differential histone-modification suggests a model of transgenerational inheritance that is distinct from what has been described before. We did detect differentially abundant RNA transcripts in sperm of TG and nonTG which are associated with a number of genes marked by the related sperm histone modifications H3K4me3 and H3K27me3. It's likely that this is the result of misregulated gene expression during meiosis. In addition, we and others have shown that the class of genes marked by H3K4me3 and H3K27me3 in sperm involve developmental processes related to our phenotype. Therefore, we hypothesized that the activity of KDM1A would more broadly influence the sperm chromatin methylation landscape, and in turn implicate H3K4me3 and H3K27me3 in transgenerational epigenetic inheritance. The goal of the next chapter was to examine histone H3K4me3 and H3K27me3 in the sperm of TG and nonTG mice to identify stable methylation differences between generations.

Chapter 3: Sperm histone H3K4me3 is implicated in paternal epigenetic inheritance

3.1 Abstract

The underlying molecular mechanisms of epigenetic inheritance remain poorly understood. Recently we've shown that disruption of histone-H3 lysine 4 di-methylation (H3K4me2) by the overexpression of KDM1A in mouse testis leads to transgenerational reduction in survivability and increased abnormal offspring development. However, the sperm chromatin alterations underlying the transgenerational phenomena remained unresolved. We hypothesized that KDM1A overexpression, and the resulting reduction in H3K4me2, would also impact surrounding histone modifications. Therefore, we applied an intergenerational examination of sperm H3K4me3 and H3K27me3. We show that, independently of H3K27me3, H3K4me3 is increased in sperm of transgenic and wild-type littermates (nonTG). Thus, we have described a set of paternally inherited histone-modifications likely resistant to epigenetic reprogramming.

3.2 Introduction

Complex disease is rising at rates that cannot be explained by genetics alone. Epidemiological studies of human cohorts have shown *in utero* and early childhood environment may predispose individuals to intergenerational late-onset disease (1, 2). Ensuing research related to the developmental origins of health and disease (DOHaD) has thus been focused on the maternal contribution while the role of the father remains underappreciated and unresolved. Paternal pre-conception exposure to toxicants, and conditions of under or over-nutrition are

associated with metabolic disorders, birth defects, and disease in the offspring of mouse and men (3-6). Remarkably, paternal effects can span multiple generations and may be mediated by epigenetic signals inherited through the sperm (3, 7-13). Identifying the mechanisms of epigenetic inheritance in sperm is central to understanding how a father's environment can impact the health and development of the next generation.

Spermatogenesis involves the most dynamic chromatin remodeling event of any cell type. Testis specific histone variants are incorporated throughout meiosis and the vast majority of nucleosomes are evicted and replaced by protamines in the spermatid (14, 15). By the end of spermiogenesis, the spermatozoa of mice and men retain between 1 and 15% of total nucleosomes (16-18). Nevertheless, a growing body of work suggests that this retention is non-random and instructive to the developing embryo (7, 18-24). In fact, the retained nucleosomes contain various post-translational modifications on their histone tails and are preferentially enriched at CpG rich gene promoters with functional roles in development and cellular process (18-21). These biological-signposts are well positioned to contribute to the developmental programme of the next generation. Therefore, we and others hypothesize that sperm histones are implicated in paternal epigenetic inheritance (7, 18, 22, 23).

Current examples of paternal epigenetic inheritance implicate sperm non-coding RNA or DNA methylation as the primary mediator of this process (4, 6, 11-13, 25-27). How RNA molecules participate in the stable transmission of transgenerational epigenetic phenotypes remains unknown. In contrast, a recent discovery in human primordial germ cells (PGCs) has shown that some enhancer and promoter regions are able to avoid DNA demethylation during gametic reprogramming (28). The identification of these epigenome "reprogramming escapees" contributes to the mechanistic understanding of the propagation of epimutations between

generations (28). Moreover, we have previously identified that sperm histone methylation plays a central role in the initiation of transgenerational phenotypes that include severe developmental abnormalities and post-natal death (7).

In this model, transgenic (TG) male mice were designed to overexpress histone demethylase KDM1A in the germ cells of the testis (7, 29). KDM1A exposure throughout spermatogenesis culminates in an altered sperm epigenome that was characterized by a reduction in H3K4me2 at the transcriptional start site (TSS) of many developmental genes (7). Hemizygous paternal transmission of the transgene generates both transgenic and wildtype-descendants (nonTG) that are derived from sperm with an abnormal epigenome (Figure 1). The nonTG lineage provides an opportunity to examine molecular mechanisms of intergenerational epigenetic inheritance. Our previous analysis in this model failed to detect changes to H3K4me2 in nonTG sperm (7). Moreover, we observed no changes to sperm DNA methylation at CpG rich regions in TG or nonTG sires. As a result, we were left perplexed by what possible epigenetic mechanisms were underlying the transgenerational phenotype.

We had observed that regions of reduced H3K4me2 in TG sperm were associated with regions of enriched H3K4me3 and H3K27me3 in wild-type sperm (7, 21). Although KDM1A does not have demethylase activity for these marks, these methylation states were enriched at genes with functional roles related to development. Moreover, H3K4me3 and H3K27me3 were associated with genes that displayed intergenerational differences in sperm RNA abundance (7). Hence, we suspected that the activity of hsKDM1A during spermatogenesis had directly or indirectly altered the histone modifications surrounding regions of reduced H3K4me2 in sperm. Therefore, our goal was to identify intergenerational differences in sperm H3K4me3 and H3K27me3 using chromatin immunoprecipitation and whole genome sequencing (ChIP-seq) of

C57BL/6J (control), TG, and nonTG mouse sperm. We identified regions of differentially methylated H3K4me3 (dmH3K4me3) in comparison to control at the TSS of genes in both TG and nonTG littermates. These loci were predominantly independent of sperm H3K27me3, and H3K27me3 remained unchanged between groups. An integrative analysis revealed common associations between genes with dmH3K4me3 and regions of murine KDM1A binding in spermatocytes, altered H3K4me2 in TG sperm, and H3K4me3 in the early embryo (7, 22, 30). Taken together, this study implicates H3K4me3 in mammalian paternal epigenetic inheritance while exploring the relationship between sperm histone methylation and biological processes in the next generation.

3.3 Results

Characteristics of H3K4me3 and H3K27me3 in sperm

Our implementation of the ChIP-seq protocol described by Hisano *et al.*, produced highly correlated, reproducible datasets from mono-nucleosome chromatin fragments (SI Figure 2) (31). In comparison to prior examples, the ChIP-sequencing effort presented in this study provides improved genomic coverage of sperm H3K4me3 and H3K27me3 by utilizing over 30 million sequencing reads for each of the 1-3 biological replicates (SI Table 1) (18-21, 32, 33). Genome wide quantification of 1000bp tiles showed how H3K4me3 enrichment between samples is globally similar ($R \geq 0.96$) (SI Figure 2). These robust signals improved our ability to detect histone methylation enrichment at low-levels. To identify H3K4me3 peak regions we used MACS2 (34) with customized sperm-chromatin parameters to account for broad peaks (see methods). We detected no major differences in peak count or distribution between groups, and report a total of 30445 peaks from C57BL/6J control, 31125 from TG, and 34685 from nonTG (Figure 1c; SI Table 1). As expected, H3K4me3 was preferentially enriched at promoter regions

in mouse sperm (55%, 15708), but a large number of peaks were also enriched at intergenic regions of the genome (30%; 9255) (Figure 1c). Although some intergenic peaks may be associated with unannotated genes, 457 were found to overlap regions annotated as sperm enhancers or super enhancers (18). Nevertheless, due to the specific depletion of sperm H3K4me2 previously described in TG males (7) we focused on regions of TSS (\pm 1kbp) for all known-genes (n=24506). We identified 15708 TSS with enriched H3K4me3 in mouse sperm using a read-count threshold (see methods, SI Figure 4), and confirmed a strong association between H3K4me3 enrichment and CpG ratio (21) (Figure 2a). We further classified H3K4me3 positive TSS into categories of high, intermediate, or low levels of enrichment (Figure 2a). GO analysis of these genes revealed distinct functional categories between each enrichment level (35, 36). High-H3K4me3 was comprised of mainly metabolic and spermatogenesis functions, intermediate enriched for regulation of transcription and embryonic developmental processes, while low consist of transport, signalling, and immune system functions (Figure 2C).

TG and nonTG sperm gained H3K4me3 in comparison to C57BL/6J controls

Between-group comparisons were used to determine whether over-expression of KDM1A indirectly altered sperm-H3K4me3 in TG and nonTG mice when compared to control. To do this, we applied two analytical approaches available in R/Bioconductor: DiffBind (37, 38) and csaw (37). Both methods identified a relative increase in H3K4me3 at gene TSS in TG and nonTG sperm. However, we favoured the increased power of csaw over DiffBind to detect and control statistical differences with a window based, regional approach (37). Therefore, we elected to focus on the top ranked TSS regions (\pm 1kbp) with increased H3K4me3 as determined by csaw. The differentially methylated promoter regions identified using DiffBind showed a

similar specificity (67% overlap) and pattern to those identified with csaw, and are displayed in the supplemental (SI Figure 1; SI Table 2).

In TG-sperm, we report 2064 TSS regions with increased H3K4me3 signal in comparison to control (Figure 3). In nonTG sperm, 201 TSS regions showed increased H3K4me3 over control, of which, 161 (80%) were the same regions to gain H3K4me3 in TG sperm (Figure 3). Characterization of genes gaining H3K4me3 showed the 70% were within the intermediate or low category of enrichment (Figure 2a). The protein-protein interaction network of differentially methylated genes shared between TG and nonTG sperm enriched for functional roles such as embryonic development, transcriptional regulation and chromatin modification, which are consistent with the embryonic development phenotype reported in offspring of these mice (7, 39) (Figure 4)

No detectable changes in H3K27me3 between generations

We compared enrichment of H3K4me3 and H3K27me3 in control sperm and detected 4665 promoter regions were co-enriched with both H3K4me3 and H3K27me3 (Figure 3g,h). These regions constitute a population of genes that corresponds to bivalent genes set early in spermatogenesis and maintained in sperm (19, 40). Notably, there was a striking lack of association between genes with increased dmH3K4me3 and genes we defined as bivalent (Figure 3c, g, h). Only 5% of all TG and 1% of both TG+nonTG dmH3K4me3 were detected at bivalent loci (Figure 3c). Moreover, we detected no differences in H3K27me3 between genotypes or across generations (data not shown). This suggest that the mechanism of intergenerational inheritance is independent of H3K27me3 in this model.

Gains in H3K4me3 occurred at specific genes correlated with spermatocyte Kdm1a binding sites

Recent evaluations of sperm chromatin have revealed that H3K4me3 and 2 are abundant in mouse sperm, with many of the same regions enriched for both modifications (7, 19, 21). By integrating publically available datasets with the sperm H3K4me3 ChIP-seq performed in this study, we were able to ask how TG and nonTG differences in H3K4me3 relate to the surrounding histone methylation landscape. We first confirmed that 15391/15708 (98%) of the TSS bound by H3K4me3 are also marked by H3K4me2 in control sperm (Figure 5B) (7). By extension, 95% of genes with dmH3K4me3 in TG and nonTG sperm were also found to be associated with H3K4me2 in control sperm (Figure 5B). We then asked how well dmH3K4m3 coincides with murine Kdm1a binding in the pachytene spermatocyte and examined this relationship using the ChIP-seq dataset from Zhang et al., 2013 (30). Analysis of this dataset using our established thresholds (SI Figure 6) reported endogenous Kdm1a binding in pachytene spermatocytes occurs at 68% of all TSS marked by H3K4me2 (11012 of 16236) or H3K4me3 (10796 of 15708) in the sperm (SI Figure 7). In addition, we confirmed that nearly all of the 11012 TSS bound by Kdm1a in spermatocytes are associated with H3K4me2 and H3K4me3 in sperm (98%) (Figure 5B). Most of the dmH3K4me3 detected in TG (90%) and both TG and nonTG (89%) sperm occurred at regions associated with endogenous Kdm1a binding in spermatocyte (Figure 5B). As we saw for H3K4me2 (7), dmH3K4me3 occurs at a specific subset of all endogenous Kdm1a binding sites. These differentially methylated regions are further specified in that they occur at gene promoters with higher levels of spermatocyte Kdm1a enrichment than unchanged H3K4me3 (SI Figure 7b). To address how dmH3K4me3 related to the reduction of H3K4me2 in TG-sperm, we compared the genes with altered methylation in

both datasets (Figure 5A,B) (7). Most of the dmH3K4me3 occurred at a different set of genes than those we previously detected as reduced H3K4me2; however, 523 (26%) and 48 (31%) of the TSS were shared with either TG dmH3K4me3 or both TG & nonTG dmH3K4me3, respectively (Figure 5B, green chord).

dmH3K4me3 in TG and nonTG sperm is correlated with H3K4me3 and gene expression in the 2-cell embryo

To determine a link between sperm H3K4me3 and its potential for gene regulation in embryonic development we compared our datasets to ChIP-seq and RNA-seq data from pre-implantation mouse embryos (22, 23, 41). Most TSS marked by H3K4me3 in sperm are present at the 2-cell stage (90%; 14183/15708) (Figure 5b). Moreover, TSS with dmH3K4me3 were almost completely associated with H3K4me3 in the 2-cell embryo (97% of TG-dmH3K4me3; 94% of TG/nonTG-dmH3K4me3) (Figure 5a,b). This suggests a likely paternal contribution to embryonic chromatin state. Moreover, our computational analysis of publically available allele-specific maps of H3K4me3 in pre-implantation embryos (23) showed similar patterns of enrichments of H3K4me3 for the parental allele at the 2-cell stage (SI Figure 5a,b).

Genes with dmH3K4me3 in TG-sperm are also highly expressed in pre-implantation embryos (23, 41), and many are active as early as the PN5 zygote (23) (Figure 6B, SI Figure 5). A linear model fit through zero showed that genes with altered expression in TG-sired 2-cell embryos trend towards higher expression in the ICM when compared to all genes (Figure 6A).

3.4 Discussion

The molecular mechanism underlying paternal epigenetic inheritance remains poorly defined. However, sperm histones and their post-translational modifications may play a role in

the transgenerational health and development of offspring (7, 18, 21, 42). Here we examined whether H3K4me3 and H3K27me3 differed in sperm from KDM1A overexpressing TG mice and their nonTG wildtype-littermates in comparison to control. We anticipated that disruption of H3K4 di-methylation would reduce the abundance of H3K4me3 by removing a necessary intermediate in the establishment of tri-methylation (43, 44). Unexpectedly, we showed through two independent analyses that the H3K4me3 signal was primarily increased at the TSS of TG sperm when compared to control (Figure 2a, SI Figure 1). Moreover, a set of these genes remained differentially methylated in the sperm of nonTG littermates (Figure 3).

KDM1A in the germline is critical for major cellular processes such as transcriptional regulation, maintenance of differentiation and meiotic entry, and chromatin remodelling (30, 45-49). This range of function is illustrated by the over 11000 gene promoters bound by KDM1A in the spermatocyte, most of which are associated with H3K4me2 and H3K4me3 (30). Our data suggests there is a strong correlation between sperm dmH3K4me3 and a subset of these KDM1A binding sites (SI Figure 7b; Figure 5b), but it remains unclear how an increase in H3K4me3 was first established. The specificity of dmH3K4me3 observed in this study mimics the specificity of reduced H3K4me2 that was previously described in the TG sperm of this model (7). From this relationship, we infer that regions of dmH3K4me3 in sperm are initiated during spermatogenesis by the targeting of transgenic KDM1A to a subset of endogenous Kdm1a binding sites. The genes marked by dmH3K4me3 can be stratified based on their relationship with surrounding chromatin characteristics (Figure 5b). Although there are genes with dmH3K4me3 that overlap regions of reduced H3K4me2 in TG sperm, a large proportion are independent of this change (Figure 5b). Indeed, this result may be caused by insufficient power to detect methylation

differences in either study but may also allude to a scenario where transgenic KDM1A overexpression results in context dependent chromatin changes.

How transgenic KDM1A identifies and influences these discrete regions is unknown but may be driven by KDM1As various binding partners and protein complex associations (50-52). For example, in addition to a primarily repressive role in transcription (53, 54), KDM1A has been found to associate with the methyltransferase-containing super-complex ALL-1 that's involved in H3K4me3 methylation and transcriptional activation (55). Likewise, in prostate tumour cells, direct methylation of KDM1A by the protein-methyltransferase Ehmt2 facilitates an interaction between the chromatin remodelling protein CHD-1 and KDM1A that's necessary for androgen receptor (AR) dependent transcriptional activation (56). The dual chromodomains of CHD-1 have been shown to selectively target H3K4me3 that are independent of any co-enrichment for H3K27me3 (57, 58); therefore, this interaction would support the underrepresentation of H3K27me3 at gene promoters with dmH3K4me3 detected in this model (Figure 3, Figure 5B).

Once bound to a target region, the presence of a secondary domain on the surface of KDM1A is capable of stabilizing complex binding to chromatin across observable timescales (59, 60). The stability of this interaction may allow Kdm1a to act as a platform for the recruitment of additional chromatin remodelers and transcription factors to these sites (53, 61), and even serve as an anchor point for higher order chromatin looping. If the kinetics of these different mechanisms are shifted due to abnormal levels of KDM1A protein, TG spermatocytes and spermatids may have increased opportunity to interact with alternative effector proteins to indirectly increase H3K4me3 at target sites. Indeed, the aforementioned hypotheses require testing via the specific characterization of transgenic KDM1A binding partners and locations

across different stages of spermatogenesis to clearly demonstrate how dmH3K4me3 is established.

The discovery of preserved regions of abnormal histone methylation in sperm of nonTG mice unexposed to transgenic KDM1A suggests that the abnormal epigenome is paternally inherited by the zygote and maintained into the male germ cells. At this time, we do not understand how dmH3K4me3 can escape epigenetic reprogramming in both embryonic and germline development. The recent observation of intergenerational H3K27me3 inheritance in yeast and *Drosophila* provides the first evidence that histone modification can behave as a true epigenetic signal to mediate non-genetic inheritance between generations (62-65). The observation of persistent methylation differences in the nonTG sperm, the altered H3K4me signal may also feed-forward by reinforcing some of the abnormal patterns throughout embryonic remodelling events. These chromatin dynamics have recently been discussed in breakthrough papers which map pre-implantation mouse embryo chromatin signature (22-24). While it's unclear if paternal chromatin is erased or simply undetectable at the PN5 zygote, the consensus is that H3K4me3 signal not only grows in strength but spreads across broad domains as development progresses towards genome activation in the late 2-cell embryo. The correlation between dmH3K4me3 and the chromatin signature of preimplantation embryos highlights the likelihood that paternal histones contribute to the early embryo (18, 22-24). Our data showed no preference for broad ($\geq 5\text{ kbp}$) H3K4me3 being altered in sperm (SI Figure 7); however, promoter associated peaks observed in sperm appeared to mark the boundaries of broad H3K4me3 (24). This "book-end" like behaviour may serve as paternally-derived targets for epigenetic and transcriptional machinery to contain chromatin remodelling events within a specific region and establish the proper zygotic chromatin signature. Although the immediate

role of paternal nucleosomes after fertilization is unclear, it's also possible that paternal histone methylation facilitates the asymmetrical paternal transcription occurring between pronuclei and before zygotic activation (66). Therefore, we propose that the paternal contribution of H3K4me3 and H3K4me2 may tune the timing and magnitude of embryonic transcription (Figure 6a; SI Figure 5a).

3.5 Methods

Breeding pedigree and transmission of transgenic KDM1A mice

KDM1A TG mice were generated as previously described (7). Briefly, full-length human KDM1A was overexpressed from a modified pIRES-EGFP vector (Clontech, USA; #6064-1) that contained a germ-cell specific isoform of the human EF1 α promoter (29). Heterozygous transgenic males from the F3 generation (TG³) (7) were bred to C57BL/6J females to generate TG⁴ and nonTG⁴ offspring (Figure 1). All C57BL/6J control mice were age matched and obtained from Charles River Laboratories International (Wilmington, MA). Mice were provided with water and standard mouse chow ad libitum and housed in a 12-hour light, 12-hour dark cycle at 21°C with 50% humidity. All animal procedures were approved by the Animal Care and Use Committee of McGill University, Montreal, Canada.

Sperm isolation

To isolate sperm for chromatin immunoprecipitation followed by next generation sequencing (ChIP-seq), cauda epididymides were cut 4-8 times for sperm to escape, then placed in Donners medium (135 mM NaCl; 5mM KCl; 1mM MgSO₄; 2mM CaCl₂; 30mM HEPES; 60% lactate syrup; 1mM sodium pyruvate; 20mg/mL BSA; 25 mM NaHCO₃) with incubation and gentle

agitation at 37C for 1 hour. The media containing sperm was then filtered through a 20 μ m mesh strainer and washed up with cold phosphate buffered saline (PBS) until somatic cells were removed. Sperm purity and quantity was assessed via hemocytometer counts following the World Health Organization (WHO) guidelines and was at minimum 99.5% pure (0 somatic cells in 200 cells counted). Recovery ranged from 5 – 8 million cells per mouse and samples were pooled from 3-4 mice.

Chromatin Immunoprecipitation followed by next generation sequencing

A total of 12 million sperm cells pooled from 3-4 TG, nonTG and C57BL/6J males (4-6 months). For H3K4me3 ChIP-seq, n=3 biological replicate ChIP were performed per group. For H3K27me3 ChIP-seq, n=1 ChIP was performed per group. Sperm was prepared as previously described (31) with some modifications. First, pooled sperm was made highly accessible through pre-treatment with 50mM DTT. Mono-nucleosome chromatin fragments were generated from 6 aliquots of 2 million cells through treatment with non-ionic detergents in lysis buffer (0.5% nonident-P40, 1% sodium deoxycholate, 15mM Tris-HCl (pH 7.5), 60mM KCl, 5mM MgCl₂, 0.1mM EGTA, 1mM DTT, and 0.3M sucrose), and a subsequent enzymatic digestion with 15U of MNase (Roche, Switzerland, #10107921001) for exactly 5 minutes at 37°C. Soluble chromatin was recovered after centrifugation at 17000 x g and pre-cleared with blocked-dynabeads for 1 hour. Genomic DNA (gDNA) was prepared from 4 million sperm cells. After cell lysis, isolated chromatin was fragmented using 5x 30s/30s on-off cycles on a cup-horn sonicator to generate 300-600bp fragments. 5 μ g of ChIP-grade primary antibody (Anti-H3K27me3 (Cell Signalling Technologies; USA, #CST9733); Anti-H3K4me3 (Cell Signalling Technologies, USA; #CST9751) was bound to 50 μ L dynabeads protein-A (Invitrogen, USA;

Cat#10001D) in PBS + 0.5% BSA through incubation with rotation at 4°C for 7 hours. Bead-antibody complex was washed 1x with PBS+BSA, followed by 3x with combined buffer (15mM Tris-HCl (pH 7.5), 60mM KCl, 5mM MgCl₂, 0.1 mM EGTA, 85mM Tris-HCl, pH 7.5, 3mM MgCl₂ and 2mM CaCl₂), and combined with precleared chromatin for incubation with rotation overnight at 4°C. Bead-antibody-chromatin complex was collected using a magnetic bar and washed once with Wash Buffer A (50mM Tris-HcL (pH 7.5), 10mM EDTA and 75mM NaCl), then twice with Wash Buffer B (50mM Tris-HCl (pH 7.5), 10mM EDTA and 125mM NaCl) for 5 minutes each. After the third wash, the bead-complex was transferred to a low-bind 1.5mL tube (Eppendorf, Germany; Cat# 0030108051) and subject to two repeated elutes using 125µL elution buffer (0.1M NaHCO₃, 0.2% SDS and 5mM DTT) and shaken for 10 minutes at 65°C. Pooled elutes were then treated with RNaseA and Proteinase K as previously described (31). Immunoprecipitated DNA was recovered using Zymogen ChIP-Clean and Concentrator Kit (Zymo Research, USA; Cat# D5205). DNA quality and concentration was assessed using Bioanalyzer high-sensitivity DNA kit (Agilent, USA; Cat# 5067-4626) 3-5ng of 147bp mononucleosomal DNA were size selected using Agencourt AMPure XP magenetic beads (Beckman Coulter, USA; Cat#A63880) following standard protocol.

Library prep and Next-Generation Sequencing

H3K27me3, H3K4me3 and input libraries were prepared using Kapa BioSystems HTP Library Preparation kit (Roche, Switzerland; Cat# 07138008001). Sequencing was performed on Illumina HiSeq 2500 and multiplexed with 6 samples per lane. H3K27me3 was sequenced using paired-end 125bp reads while H3K4me3 libraries were sequenced using single-end 100bp reads.

Processing and Alignment of Raw Data

Fastq files were checked for quality with FastQC (v0.11.5; (67)) and, where required, reads were trimmed from the 3' end by Trimmomatic (v0.32; (68)) using the parameters TRAILING = 20, and MINLEN = L, where L = read length x 0.80. Reads were aligned to the mouse reference genome (UCSC version mm10, December 2011) using BWA (v0.7.15) (69) with the settings MEM -MP -t 10 -v 2 -c 100 for paired-end, or by bowtie (v1.1.2) (70) with the settings -t -v 3 -m 100 -p 10 for single-end. These parameters were implemented to reduce reads with multiple alignments and allow up to 3 base-pair mismatches between read and reference (71). Samtools (v 1.3) (72) was used to filter out unaligned reads, sort the alignment file and convert it to BAM format. Read duplicates were marked using PicardTools (v2.4.1; Broad Institute), but not discarded. Alignment statistics and quality control were generated using R package ChIPQC (73).

Peak Calling

The peak callers HOMER (v 4.8.2) (74), MACS (v 1.4.2) (34) and MACS2 (v 2.1.1.2) (34) were compared for suitability to peak calling of the unique sperm chromatin profiles. Biological replicates were merged and tested against a sonicated genomic DNA control sample. It was determined that for H3K4me3, MACS2 callpeak with the options --nomodel -q 0.000001 --broad --broad-cutoff 0.000001 gave the best balance between type I and type II errors, and visual assessment of tracks indicated they were the most accurate defining peak breadth.

Quantification of chromatin signatures at promoter regions

Where appropriate, biological replicates were merged before analysis. Reads were counted in 1-kbp windows (\pm 1-kb) surrounding the transcriptional start site of the longest transcript for all known-genes (n=24506; refseq TxDb accessed October 2016). The ChIP-seq datasets used for analysis included sperm H3K4me3, H3K27me3, H3K4me2 (7), spermatocyte Kdm1a (30), and pre-implantation embryo H3K4me3 and H3K27me3 (22, 23). To allow for comparison of chromatin features between experiments the read counts were then normalized by library size using the following formula (21, 31): $\text{Log2}((\text{Cnt}_{\text{smp}} / \text{LSize}_{\text{smp}}) * \min(\text{LSize}_{\text{smp}})) + C$, where Cnt_{smp} = reads for each 1-kb window per sample, $\text{LSize}_{\text{smp}}$ = the sum of counts for all 1-kb windows per sample, and $C = 8$, a constant value used to stabilize counts of low abundant gene windows.

Classification of chromatin signature at gene TSS

To identify the total number of gene promoters that were enriched for a given chromatin mark we determined the log2 read count value that separates the windows containing only background from those with an enriched signal. The bimodal distribution of the density pattern clearly indicated two populations of promoter windows. The value corresponding to the local minimum between the two populations of window counts served as the lower threshold for each chromatin state analyzed (Figure 2; SI Figure 3). TSS were classified as associated with a feature if the window counts surpassed the threshold. We defined genes as co-enriched for two histone modifications if the window counts in both datasets surpassed both thresholds. Classification of gene promoters with low, intermediate, or high levels of H3K4me3 or H3K27me3 followed a

similar density distribution approach, where threshold values were determined based on natural features of the density function (Figure 2).

Differential Methylation analysis with csaw

We applied csaw's sliding window approach to identify differential methylation at promoter regions (37). In brief, reads outside of blacklist regions (75) with mapq ≥ 20 were counted inside 150bp sliding windows genome wide. Windows were filtered out if their read count abundance fell below the fold enrichment vs background was below $\log_2(3)$ in order to select only high-confidence enrichment regions (SI Figure 3). An MA-plot (76) was used to compare the \log_2 ratio of counts per bin (M) against the average abundance (A) of the bin, between all samples. Comparisons between M values at high abundance appeared non-linear and suggests the presence a slight bias in the immunoprecipitation efficiency. These types of bias cannot be dealt with through linear-scaling normalization techniques and thus Loess-normalization was applied to the data under the assumption that the majority of methylated windows were not differentially methylated (77, 78). PCA plots were generated using normalized read counts and identified TG-A as a significant outlier that was subsequently removed from further analysis. EdgeR was used to compare TG to CRwt and nonTG to CRwt (77, 79). To increase statistical power, only those windows which overlap +/- 1kbp around the TSS of known-genes were selected for analysis (n = 24506; Refseq TxDB accessed October 2016). The FDR was further improved by aggregating adjacent promoter-windows into a single region. P-values of the adjacent windows were combined using Sime's method (80) to generate one statistic to explain each region. Any region that contained at least 1 window with $\log FC \geq 0.5$ with an $FDR \leq 0.1$ was called as

dmH3K4me3. All of the above methods were implemented using the R/Bioconductor (v 3.4) package csaw (v 1.8.1) (37).

Differential Methylation analysis using DiffBind

H3K4me3 peaks for TG nonTG and control sperm were identified using MACS2 (see above). The peaks that were common between at least 2 groups were combined into a union-set for differential methylation analysis with DiffBind (v 2.2.12) (38). ChIP-seq reads within each peak region were counted using the summarizeOverlaps function (GenomicAlignments, v 1.10.1) and include duplicate reads. Total read counts for each group were normalized using DiffBind's default parameters of full library size. PCA plots were generated using normalized read counts and identified TG-A as a significant outlier that was subsequently removed from further analysis. Normalized read counts in peaks were compared using DESeq2. The comparisons made were TG vs Control and nonTG vs Control, and we focused our results on H3K4me3 differences occurring in peaks at annotated promoters. Significant differentially methylated H3K4me3 peak regions (dmH3K4me3) were called if $\log FC \geq 0.5$, $FDR \leq 0.05$. All of the above methods were implemented using R/Bioconductor (v3.4).

GO enrichment analysis

GO analysis of genes classified as high, intermediate or low enrichment of H3K4me3 or H3K27me3 were defined with PantherGO (35, 36). GO analysis of genes identified as containing differentially methylated H3K4me3 in TG and nonTG sperm was performed using DAVID (81, 82). GO analysis of protein-protein interaction network was performed using topGO as part of networkanalyst.ca (39, 83-85).

Network Analysis

Networks were generated using NetworkAnalyst.ca (39, 84, 85). Briefly, the gene list containing differentially methylated H3K4me3 TSS in both TG and nonTG sperm were mapped to the genes the protein-protein interaction (PPI) database InnateDB (input = 157; mapped seeds = 35) (86). This database contains manually curated protein interactions that are retrieved from the published literature and datasets within the International Molecular Exchange Consortium (IMEx)(87). Then, first order interactions are returned for each mapped gene resulting in subnetworks of nodes (n = 57) connecting to their interaction partners via edges (n=70). Subnetwork1 was used for visualization while GO analysis was performed on all subnetworks. For details see www.networkanalyst.ca.

Coverage Tracks

Differentially methylated promoter regions in TG and nonTG H3K4me3 sperm were visualized using the R/bioconductor package GViz (v 1.18.2). The read counts from merged replicates were scaled to counts per million (CPM) for comparison purposes using the following formula: $1e6 * \text{Cov}_{\text{smp}} / \text{LSize}_{\text{smp}}$). Datasets visualized were H3K4me3 in C57BL/6J, TG, and nonTG sperm, H3K4me2 in TG and C57BL/6J sperm, H3K27me3 in C57BL/6J sperm, KDM1A in C57BL/6J pachytene spermatocyte (30) and H3K4me3 in 2-cell preimplantation embryo (22)

3.6 References

1. D. J. Barker, The fetal and infant origins of adult disease. *BMJ* **301**, 1111 (1990).
2. C. N. Hales, D. J. Barker, The thrifty phenotype hypothesis. *Br Med Bull* **60**, 5-20 (2001).
3. M. D. Anway, A. S. Cupp, M. Uzumcu, M. K. Skinner, Epigenetic transgenerational actions of endocrine disruptors and male fertility. *Science* **308**, 1466-1469 (2005).

4. R. Lambrot *et al.*, Low paternal dietary folate alters the mouse sperm epigenome and is associated with negative pregnancy outcomes. *Nature communications* **4**, 2889 (2013).
5. L. Ly *et al.*, Intergenerational impact of paternal lifetime exposures to both folic acid deficiency and supplementation on reproductive outcomes and imprinted gene methylation. *Molecular Human Reproduction* **23**, 461-477 (2017).
6. E. J. Radford *et al.*, In utero effects. In utero undernourishment perturbs the adult sperm methylome and intergenerational metabolism. *Science* **345**, 1255903 (2014).
7. K. Siklenka *et al.*, Disruption of histone methylation in developing sperm impairs offspring health transgenerationally. *Science* **350**, aab2006 (2015).
8. B. G. Dias, K. J. Ressler, Parental olfactory experience influences behavior and neural structure in subsequent generations. *Nature neuroscience* **17**, 89-96 (2014).
9. K. Gapp *et al.*, Implication of sperm RNAs in transgenerational inheritance of the effects of early trauma in mice. *Nature neuroscience* **17**, 667-669 (2014).
10. B. R. Carone *et al.*, Paternally induced transgenerational environmental reprogramming of metabolic gene expression in mammals. *Cell* **143**, 1084-1096 (2010).
11. Y. Wei *et al.*, Paternally induced transgenerational inheritance of susceptibility to diabetes in mammals. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 1873-1878 (2014).
12. Q. Chen *et al.*, Sperm tsRNAs contribute to intergenerational inheritance of an acquired metabolic disorder. *Science* **351**, 397-400 (2016).
13. U. Sharma *et al.*, Biogenesis and function of tRNA fragments during sperm maturation and fertilization in mammals. *Science* **351**, 391-396 (2016).
14. S. Kimmins, P. Sassone-Corsi, Chromatin remodelling and epigenetic features of germ cells. *Nature* **434**, 583-589 (2005).
15. S. Rousseaux *et al.*, Epigenetic reprogramming of the male genome during gametogenesis and in the zygote. *Reprod Biomed Online* **16**, 492-503 (2008).
16. N. Tanphaichitr, P. Sobhon, N. Taluppeth, P. Chalermisarachai, Basic nuclear proteins in testicular cells and ejaculated spermatozoa in man. *Exp Cell Res* **117**, 347-356 (1978).
17. R. Balhorn, B. L. Gledhill, A. J. Wyrobek, Mouse sperm chromatin proteins: quantitative isolation and partial characterization. *Biochemistry* **16**, 4074-4080 (1977).
18. Y. H. Jung *et al.*, Chromatin States in Mouse Sperm Correlate with Embryonic and Adult Regulatory Landscapes. *Cell Rep* **18**, 1366-1382 (2017).
19. U. Brykczynska *et al.*, Repressive and active histone methylation mark distinct promoters in human and mouse spermatozoa. *Nature structural & molecular biology* **17**, 679-687 (2010).
20. S. S. Hammoud *et al.*, Distinctive chromatin in human sperm packages genes for embryo development. *Nature* **460**, 473-478 (2009).
21. S. Erkek *et al.*, Molecular determinants of nucleosome retention at CpG-rich sequences in mouse spermatozoa. *Nature structural & molecular biology* **20**, 868-875 (2013).
22. X. Liu *et al.*, Distinct features of H3K4me3 and H3K27me3 chromatin domains in pre-implantation embryos. *Nature* **537**, 558-562 (2016).
23. B. Zhang *et al.*, Allelic reprogramming of the histone modification H3K4me3 in early mammalian development. *Nature* **537**, 553-557 (2016).

24. J. A. Dahl *et al.*, Broad histone H3K4me3 domains in mouse oocytes modulate maternal-to-zygotic transition. *Nature* **537**, 548-552 (2016).
25. M. Rassoulzadegan *et al.*, RNA-mediated non-mendelian inheritance of an epigenetic change in the mouse. *Nature* **441**, 469-474 (2006).
26. K. Gapp *et al.*, Early life stress in fathers improves behavioural flexibility in their offspring. *Nature communications* **5**, 5466 (2014).
27. A. B. Rodgers, C. P. Morgan, N. A. Leu, T. L. Bale, Transgenerational epigenetic programming via sperm microRNA recapitulates effects of paternal stress. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 13699-13704 (2015).
28. W. W. Tang *et al.*, A Unique Gene Regulatory Network Resets the Human Germline Epigenome for Development. *Cell* **161**, 1453-1467 (2015).
29. T. Furuchi, K. Masuko, Y. Nishimune, M. Obinata, Y. Matsui, Inhibition of testicular germ cell apoptosis and differentiation in mice misexpressing Bcl-2 in spermatogonia. *Development* **122**, 1703-1709 (1996).
30. J. Zhang *et al.*, SFMBT1 functions with LSD1 to regulate expression of canonical histone genes and chromatin-related factors. *Genes & development* **27**, 749-766 (2013).
31. M. Hisano *et al.*, Genome-wide chromatin analysis in mature mouse and human spermatozoa. *Nature protocols* **8**, 2449-2470 (2013).
32. B. R. Carone *et al.*, High-resolution mapping of chromatin packaging in mouse embryonic stem cells and sperm. *Developmental cell* **30**, 11-22 (2014).
33. B. Samans *et al.*, Uniformity of nucleosome preservation pattern in Mammalian sperm and its connection to repetitive DNA elements. *Developmental cell* **30**, 23-35 (2014).
34. Y. Zhang *et al.*, Model-based analysis of ChIP-Seq (MACS). *Genome biology* **9**, R137 (2008).
35. H. Mi *et al.*, PANTHER version 7: improved phylogenetic trees, orthologs and collaboration with the Gene Ontology Consortium. *Nucleic acids research* **38**, D204-210 (2010).
36. P. D. Thomas *et al.*, PANTHER: a library of protein families and subfamilies indexed by function. *Genome research* **13**, 2129-2141 (2003).
37. A. T. Lun, G. K. Smyth, csaw: a Bioconductor package for differential binding analysis of ChIP-seq data using sliding windows. *Nucleic acids research* **44**, e45 (2016).
38. C. S. Ross-Innes *et al.*, Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature* **481**, 389-393 (2012).
39. J. G. Xia, M. J. Benner, R. E. W. Hancock, NetworkAnalyst - integrative approaches for protein-protein interaction network analysis and visual exploration. *Nucleic Acids Research* **42**, W167-W174 (2014).
40. S. S. Hammoud *et al.*, Chromatin and transcription transitions of mammalian adult germline stem cells and spermatogenesis. *Cell Stem Cell* **15**, 239-253 (2014).
41. L. Wang *et al.*, Programming and inheritance of parental DNA methylomes in mammals. *Cell* **157**, 979-991 (2014).
42. M. Teperek *et al.*, Sperm is epigenetically programmed to regulate gene transcription in embryos. *Genome research* **26**, 1034-1046 (2016).
43. A. Wood *et al.*, Ctk complex-mediated regulation of histone methylation by COMPASS. *Mol Cell Biol* **27**, 709-720 (2007).

44. A. Kirmizis *et al.*, Arginine methylation at histone H3R2 controls deposition of H3K4 trimethylation. *Nature* **449**, 928-U917 (2007).
45. D. J. Katz, T. M. Edwards, V. Reinke, W. G. Kelly, *C. elegans* LSD1 demethylase contributes to germline immortality by reprogramming epigenetic memory. *Cell* **137**, 308-320 (2009).
46. R. Lambrot, C. Lafleur, S. Kimmins, The histone demethylase KDM1A is essential for the maintenance and differentiation of spermatogonial stem cells and progenitors. *Faseb Journal* **29**, 4402-4416 (2015).
47. H. Sun *et al.*, Lysine-specific histone demethylase 1 inhibition promotes reprogramming by facilitating the expression of exogenous transcriptional factors and metabolic switch. *Sci Rep* **6**, 30903 (2016).
48. K. Ancelin *et al.*, Maternal LSD1/KDM1A is an essential regulator of chromatin and transcription landscapes during zygotic genome activation. *Elife* **5**, (2016).
49. D. A. Myrick *et al.*, KDM1A/LSD1 regulates the differentiation and maintenance of spermatogonia in mice. *PLoS One* **12**, e0177473 (2017).
50. R. Baron, C. Binda, M. Tortorici, J. A. McCammon, A. Mattevi, Molecular mimicry and ligand recognition in binding and catalysis by the histone demethylase LSD1-CoREST complex. *Structure* **19**, 212-220 (2011).
51. A. P. Barrios *et al.*, Differential properties of transcriptional complexes formed by the CoREST family. *Mol Cell Biol* **34**, 2760-2770 (2014).
52. Y. J. Shi *et al.*, Regulation of LSD1 histone demethylase activity by its associated factors. *Molecular cell* **19**, 857-864 (2005).
53. F. Forneris, C. Binda, A. Adamo, E. Battaglioli, A. Mattevi, Structural basis of LSD1-CoREST selectivity in histone H3 recognition. *J Biol Chem* **282**, 20070-20074 (2007).
54. C. T. Foster *et al.*, Lysine-specific demethylase 1 regulates the embryonic transcriptome and CoREST stability. *Mol Cell Biol* **30**, 4851-4863 (2010).
55. T. Nakamura *et al.*, ALL-1 is a histone methyltransferase that assembles a supercomplex of proteins involved in transcriptional regulation. *Molecular cell* **10**, 1119-1128 (2002).
56. E. Metzger *et al.*, Assembly of methylated KDM1A and CHD1 drives androgen receptor-dependent transcription and translocation. *Nature structural & molecular biology* **23**, 132-139 (2016).
57. R. J. Sims *et al.*, Human but not yeast CHD1 binds directly and selectively to histone H3 methylated at lysine 4 via its tandem chromodomains. *Journal of Biological Chemistry* **280**, 41789-41792 (2005).
58. J. J. Lin *et al.*, Mediator coordinates PIC assembly with recruitment of CHD1. *Genes & development* **25**, 2198-2209 (2011).
59. J. M. Burg, J. J. Gonzalez, K. R. Maksimchuk, D. G. McCafferty, Lysine-Specific Demethylase 1A (KDM1A/LSD1): Product Recognition and Kinetic Analysis of Full-Length Histones. *Biochemistry* **55**, 1652-1662 (2016).
60. R. Baron, N. A. Vellore, LSD1/CoREST is an allosteric nanoscale clamp regulated by H3-histone-tail molecular recognition. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 12509-12514 (2012).
61. F. Forneris *et al.*, A highly specific mechanism of histone H3-K4 recognition by histone demethylase LSD1. *J Biol Chem* **281**, 35289-35295 (2006).

62. L. J. Gaydos, W. Wang, S. Strome, Gene repression. H3K27me and PRC2 transmit a memory of repression across generations and during development. *Science* **345**, 1515-1518 (2014).
63. R. T. Coleman, G. Struhl, Causal role for inheritance of H3K27me3 in maintaining the OFF state of a Drosophila HOX gene. *Science* **356**, (2017).
64. F. Zenk *et al.*, Germ line-inherited H3K27me3 restricts enhancer function during maternal-to-zygotic transition. *Science* **357**, 212-216 (2017).
65. A. Inoue, L. Jiang, F. Lu, T. Suzuki, Y. Zhang, Maternal H3K27me3 controls DNA methylation-independent imprinting. *Nature* **547**, 419-424 (2017).
66. F. Aoki, D. M. Worrad, R. M. Schultz, Regulation of transcriptional activity during the first and second cell cycles in the preimplantation mouse embryo. *Developmental biology* **181**, 296-307 (1997).
67. S. Andrews, FastQC: a quality control tool for high throughput sequence data. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>, (2010).
68. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120 (2014).
69. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760 (2009).
70. B. Langmead, C. Trapnell, M. Pop, S. L. Salzberg, Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology* **10**, R25 (2009).
71. H. Royo, M. B. Stadler, A. Peters, Alternative Computational Analysis Shows No Evidence for Nucleosome Enrichment at Repetitive Sequences in Mammalian Spermatozoa. *Developmental cell* **37**, 98-104 (2016).
72. H. Li *et al.*, The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009).
73. T. S. Carroll, Z. Liang, R. Salama, R. Stark, I. de Santiago, Impact of artifact removal on ChIP quality metrics in ChIP-seq and ChIP-exo data. *Front Genet* **5**, 75 (2014).
74. S. Heinz *et al.*, Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Molecular cell* **38**, 576-589 (2010).
75. E. P. Consortium, An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74 (2012).
76. S. Dudoit, Y. H. Yang, M. J. Callow, T. P. Speed, Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments. *Statistica Sinica* **12**, 111-139 (2002).
77. M. D. Robinson, A. Oshlack, A scaling normalization method for differential expression analysis of RNA-seq data. *Genome biology* **11**, R25 (2010).
78. K. V. Ballman, D. E. Grill, A. L. Oberg, T. M. Therneau, Faster cyclic loess: normalizing RNA arrays via linear models. *Bioinformatics* **20**, 2778-2786 (2004).
79. D. J. McCarthy, Y. Chen, G. K. Smyth, Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic acids research* **40**, 4288-4297 (2012).
80. R. J. Simes, An Improved Bonferroni Procedure for Multiple Tests of Significance. *Biometrika* **73**, 751-754 (1986).

81. D. W. Huang, B. T. Sherman, R. A. Lempicki, Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic acids research* **37**, 1-13 (2008).
82. D. W. Huang, B. T. Sherman, R. A. Lempicki, Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols* **4**, 44-57 (2009).
83. A. Alexa, J. Rahnenfuhrer, topGO: enrichment analysis for gene ontology. *R package version*, (2010).
84. J. Xia, E. E. Gill, R. E. Hancock, NetworkAnalyst for statistical, visual and network-based meta-analysis of gene expression data. *Nature protocols* **10**, 823-844 (2015).
85. J. G. Xia *et al.*, INMEX-a web-based tool for integrative meta-analysis of expression data. *Nucleic Acids Research* **41**, W63-W70 (2013).
86. K. Breuer *et al.*, InnateDB: systems biology of innate immunity and beyond--recent updates and continuing curation. *Nucleic acids research* **41**, D1228-1233 (2013).
87. S. Orchard *et al.*, Protein interaction data curation: the International Molecular Exchange (IMEx) consortium. *Nat Methods* **9**, 345-350 (2012).

3.7 Figure Legends

Figure 3.1 Mouse model and genomic annotation of H3K4me3 and H3K27me3 in sperm

(a) Summary of transgenic mouse model used in this study (1). We overexpressed KDM1A in testis of C57BL/6J transgenic male mice to alter the sperm epigenome and previously observed a reduction of H3K4me2 at the TSS of developmental genes, and differential abundance of sperm RNA content in comparison to control. Intergenerational effects include reduced offspring survivability and increased developmental abnormalities at E18.5. Transgenerational developmental abnormalities were observed in the F2-F3 generations. For this study, TG and nonTG sperm from the F1 generation were analyzed. (b) Bar chart depicting genomic annotation of H3K4me3 and H3K27me3 peak regions in sperm of C57BL/6J control, TG and nonTG mice.

Figure 3.2 ChIP-seq analysis of H3K4me3 and H3K27me3 enrichment in control sperm

The total number of reads from all control sperm replicates of (a) H3K4me3 or (b) H3K27me3 ChIP-seq libraries were count at transcriptional start site (TSS +/- 1kbp) of all known-genes and plot against the CpG ratio of the region (observed CpG / expected CpG). The solid red line depicts the minimum count threshold that separates enriched windows (blue) from background (black). Overlaid in orange and red are all dmH3K4me3 TSS in either TG or both TG and nonTG sperm, respectively. (ii-iii) The distribution of counts within windows can be further clustered into three categories of methylation enrichment that we defined as low, intermediate, or high. (c-d) Enrichment of the top selected GO biological-process terms for gene TSS marked by high, intermediate, and low levels of H3K4me3 and H3K27me3 respectively.

Figure 3.3 Differential methylation analysis of sperm from TG and nonTG males in comparison to C57BL/6J controls.

(a) H3K4me3 enrichment for all samples represented by loess normalized log₂ read counts within the best window (150bp) of each differentially methylated region in TG or nonTG sperm versus control. Significance of each gene promoter region (merged windows intersecting ± 1 -kbp TSS) was determined using csaw and EdgeR with regional p.value ≤ 0.05 and FDR ≥ 0.1 . **(b)** Boxplots for the distributions of H3K4me3 log₂ read counts at the TSS (± 1 -kbp) of genes categorized by unchanged (n=13433) and differentially methylated H3K4me3 (n=2265). **(c)** Venn diagram showing association between gene regions with differential H3K4me3 in TG vs CRwt and nonTG vs CRwt in contrast to genes enriched for H3K27me3 in control sperm. **(d,e)** Enriched GO biological process determined using DAVIDs over-representation analysis for gene regions with differentially methylated H3K4me3. **(d)** TG sperm (n = 2016) or **(e)** both TG and nonTG sperm (n=157). Enrichments include general themes of transcriptional or translational regulation, chromatin modification, and development. **(f-h)** Scatter plots depicting log₂ read counts at all known genes (+/- 1kbp). Dashed lines represent count threshold that a given TSS must reach to be classified as enriched (blue dots; quadrant. Genes with altered H3K4me3 in TG sperm are overlaid in orange, while genes with altered H3K4me3 in both TG and nonTG sperm are overlaid in red. **(f)** Merged counts for TG-H3K4me3 vs C57BL/6J H3K4me3; **(g-h)** Merged counts for C57BL/6J H3K4me3 vs H3K27me3.

Figure 3.4 Set of dmH3K4me3 genes common in TG and nonTG sperm are involved in a network of regulatory proteins

First order protein-protein interaction (PPI) network visualization of genes with differentially methylated H3K4me3 in both TG and nonTG sperm when compared to control (n seeds = 36).

Edges denote protein interactions. Nodes enriched for biological processes of embryonic development are highlighted in blue ($p.\text{val}$ 1.6e-05). All other node colours and size represent the degree and betweenness centrality within the network. Visualization prepared using www.networkanalyst.ca (2-4)

Figure 3.5 Visualizing the relationship between genes with altered sperm chromatin and the surrounding epigenetic landscape **(a)** Snapshots of chromatin occupancy from CPM normalized read counts at selected genes with increased H3K4me3 detected through csw. Occupancies of TG (dark orange), nonTG (purple), and control sperm (dark green) were taken from this study. Additional tracks highlight the surrounding chromatin landscape and include TG H3K4me2 (1) (pink) and C57BL/6J sperm (1) (light green), H3K27me3 in C57BL/6J sperm (this study; orange), KDM1A binding in pachytene spermatocyte (5) (brown) and H3K4me3 in 2-Cell Embryo (6) (grey). Gene functions include gene regulation, chromatin remodeling, and development. **(b) Parallel set for all genes with differentially methylated H3K4me3 at the TSS in either TG or nonTG sperm.** We used a visual approach to simultaneously represent all relationships between genes with dmH3K4me3 and their associated chromatin features using multiple datasets. Each level represents the presence or absence of the specified histone modification, or binding of KDM1A for the designated cell type with respect to all regions of dmH3K4me3 in TG or nonTG sperm ($n=2265$). A positive association is represented by a red horizontal line (label “YES”). A negative association is represented by a grey horizontal line (label “NO”). dmH3K4me3 genes are represented by coloured vertical lines called chords (blue and green). The green chords highlight the proportion of dmH3K4me3 genes which overlap regions of reduced H3K4me2 in TG-sperm. The blue chords represent all other dmH3K4me3

genes. The thickness of each chord denotes the relative number of genes associated with each state. For clarity of display only co-associations ≥ 10 were selected. The panels are interpreted from top to bottom and feature the dmH3K4me3 genes associated with: reduced H3K4me2 in TG sperm (*I*); sperm H3K4me2 (*I*) & H3K4me3 (this study); Kdm1a in pachytene spermatocyte (5); sperm H3K27me3 (this study); and H3K4me3 in the 2-cell embryo (6).

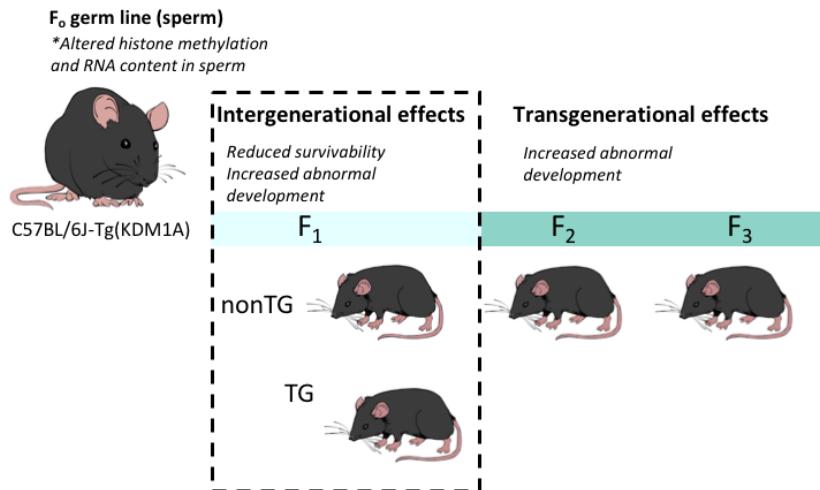
Figure 3.6 dmH3K4me3 are correlated with gene expression in early preimplantation embryo

Scatterplots of paternal allele-specific RNA-seq datasets (7) highlight how dmH3K4me3 in TG and nonTG sperm are **(a)** wildtype control sired 2-cell embryos vs. the inner cell mass (ICM). Paternally expressed genes at the 2-cell stage are represented as black dots. Overlaid in pink are the genes with altered expression in TG-sired 2-cell embryos (*I*) ($n=876$). Overlaid in yellow are the genes with both altered gene-expression in TG-sired 2-cell embryos and dmH3K4me3 in TG-sperm ($n=73$). Linear regression lines for each 2-cell embryo gene set (wildtype = black; TG = pink; TG + dmH3K4me3 = yellow) were plotted with intercepts set to zero. TG sired 2-cell gene expression was slightly favoured towards the ICM in comparison to bulk genes. **(b)** Genes with altered H3K4me3 in TG (orange) and nonTG (red) sperm are paternally expressed in the 2-cell embryo and throughout preimplantation development **(c)** Genes defined as bivalent in sperm (H3K4me3 and H3K27me3) are paternally expressed as early as the 2-cell embryo (blue), and include the few genes with differentially methylated H3K4me3 in TG sperm (orange; $n=103$).

Figure 3.1

a

Paternal epigenetic transmission



b

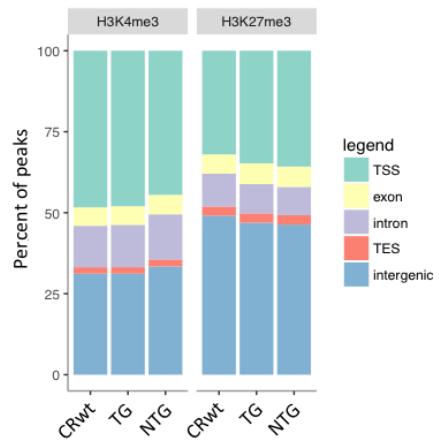


Figure 3.2

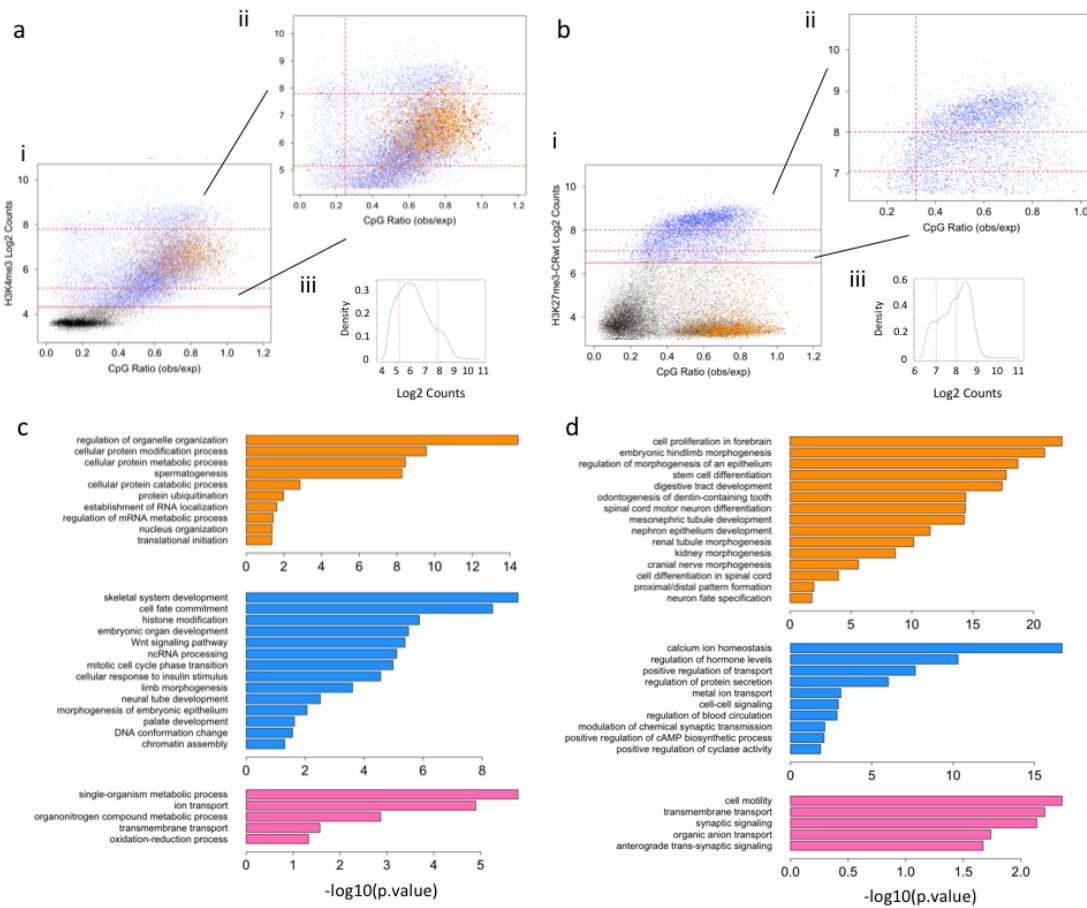


Figure 3.3

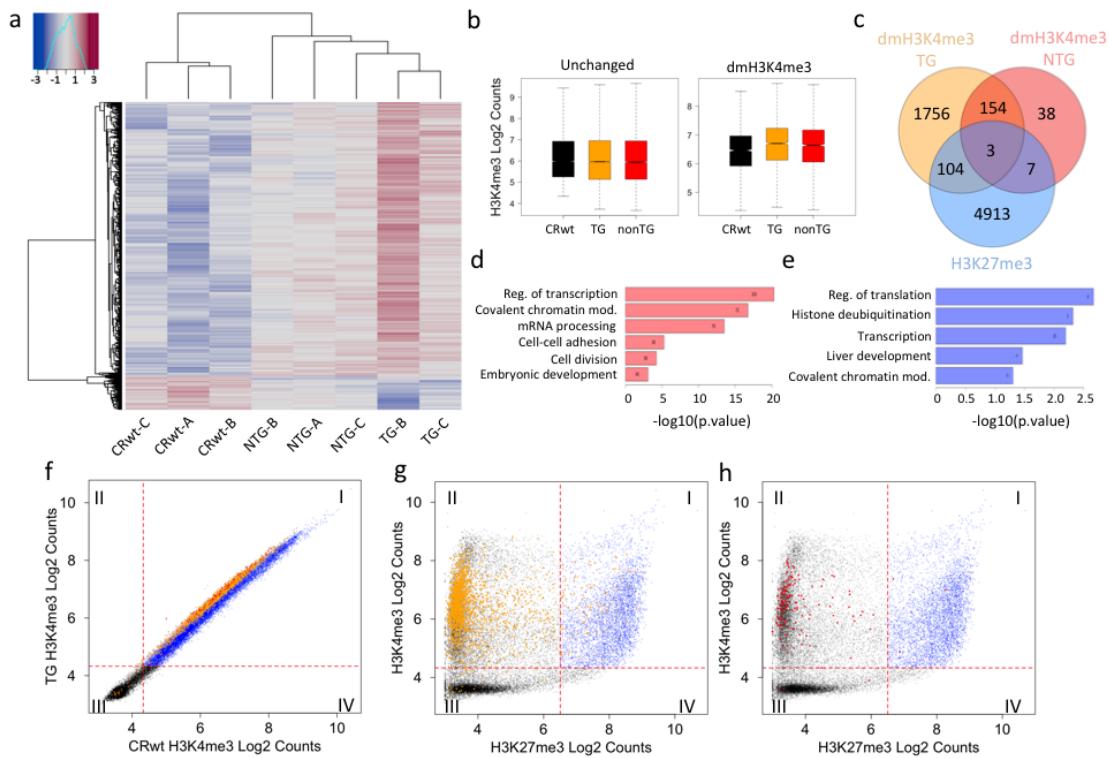


Figure 3.4

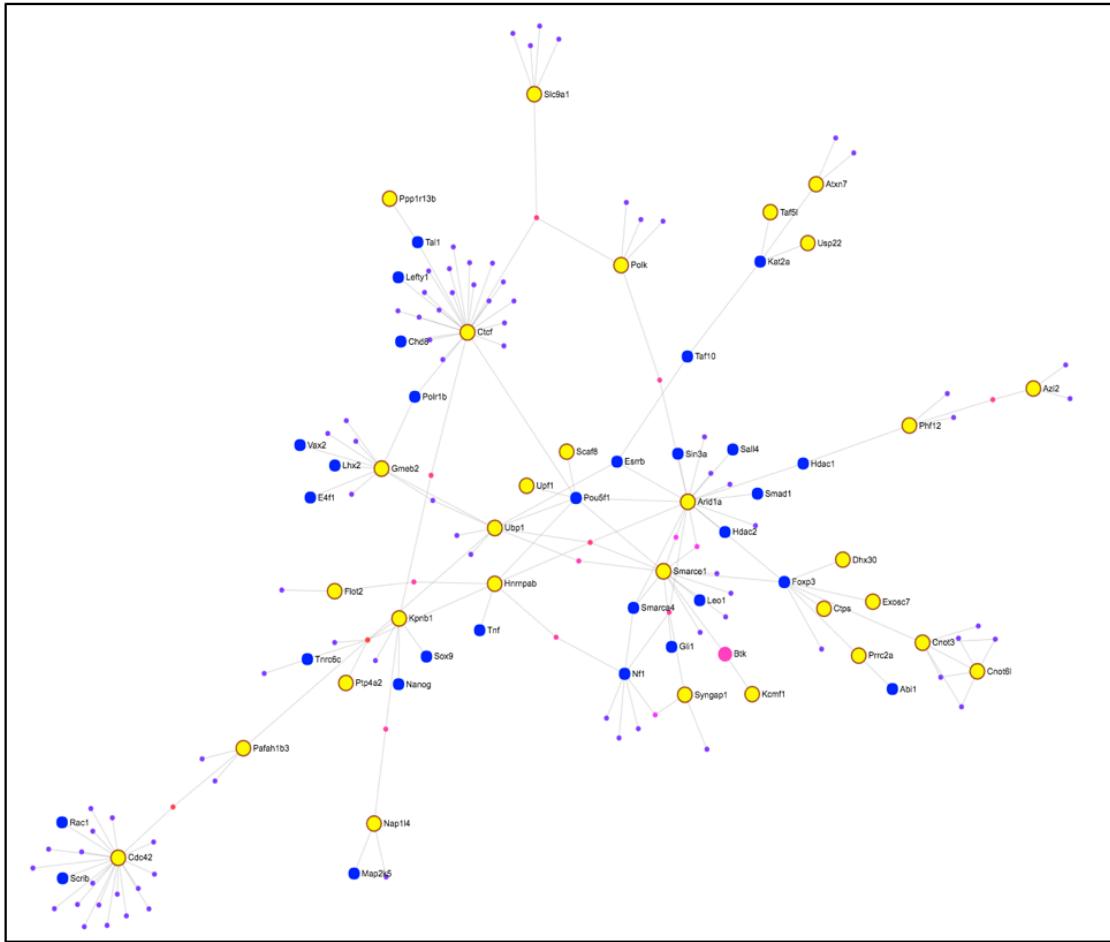
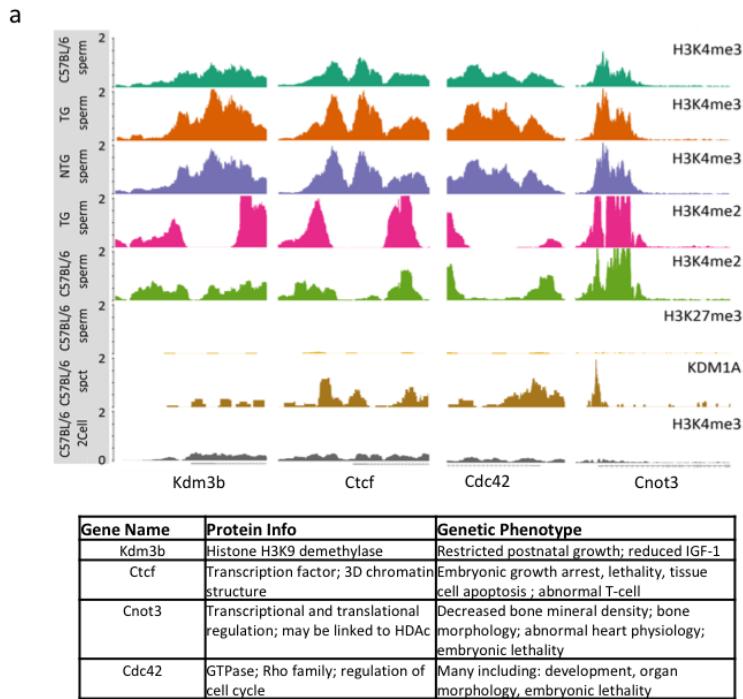


Figure 3.5



b

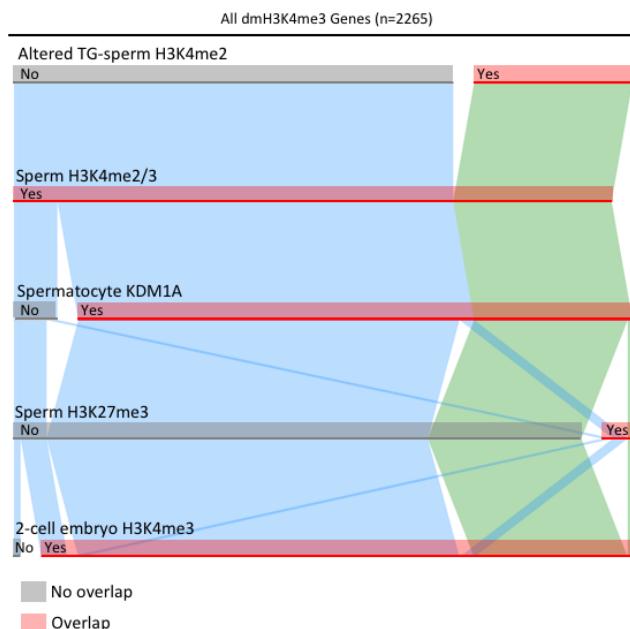
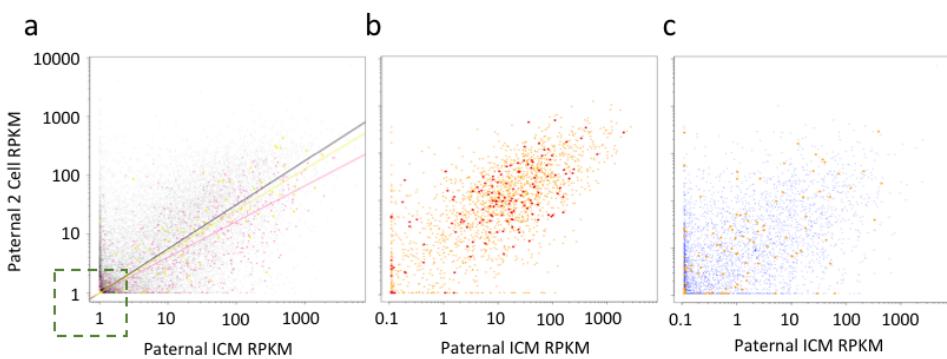


Figure 3.6



3.8 Appendix II

Supplementary Figure 3.1. Supplementary differential methylation analysis using DiffBind

(a) Heat map of between group sample correlation for reads at peak regions. **(b)** MA-plot showing differentially methylated H3K4me3 peak regions intersecting promoter regions in TG and nonTG sperm (red). **(c)** Venn Diagram depicting intersects between genes with differential H3K4me3 in TG-sperm vs control, and nonTG sperm vs control. **(d)** Scatter plot of control sperm H3K4me3 against control H3K27me3 overlaid with genes with differentially methylated H3K4me3 in TG sperm (green) and nonTG sperm (red).

Supplementary Figure 3.2. Assessing quality of replicates in the H3K4me3 ChIP-seq dataset.

(a) Pairwise comparisons of all H3K4me3 replicates using log2 read counts within genome wide tiling windows of 1000bp. **(b)** Boxplots to show the distribution of log2 read counts across the TSS of all RefSeq genes (n=24600) using varied window sizes ($\pm 250\text{bp}$; $\pm 500\text{bp}$; $\pm 1000\text{bp}$; $\pm 3000\text{bp}$) for merged H3K4me3 replicates. **(c)** Pairwise comparison of H3K4me3 enrichment in merged samples. **(d)** Boxplots of merged replicates again showing log2 read counts of across the TSS of all RefSeq genes with varying window sizes. **(e)** Complexity analysis of aligned ChIP-seq reads for H3K4me3 in control, nonTG, and TG sperm datasets. The number of uniquely aligned reads were plotted as a function of total reads in each sample (solid lines). Predicted values (dashed lines) were extended to represent complexity if 100M total reads were sampled.

Supplementary Figure 3.3. csaw sliding window analysis.

(a) Histogram of background read abundance as determined by the number of reads in 2000 bp windows tiled across the genome (red). An abundance threshold was set at $\geq \log_2(5)$ fold over background (blue). Windows below this threshold were filtered out for downstream analysis. **(b)** Principle component analysis of biological replicates for loess normalized H3K4me3 filtered windows showed TG-A is a significant outlier while remaining samples cluster distinctly from the control. **(c)** MA-plots compare H3K4me3 positive windows of CRwt-A and all biological replicates of each group. The log-fold difference between sample windows (M values) were plot against mean abundance (A values). Individual sample comparisons with raw counts highlight composition bias, as log fold-change differences are not centered on 0, and antibody efficiency biases, as windows with increased abundance have non-linear shape curve. **(d)** Individual comparisons with loess normalized counts are corrected.

Supplementary Figure 3.4. Establishment of background abundance threshold for all chromatin features. Density plots of \log_2 counts at TSS ($\pm 1\text{ kbp}$) revealed a bimodal distribution of counts for all chip-seq datasets used in this study. The local minimum count values (red line) of this distribution was used to define positively enriched genes (blue).

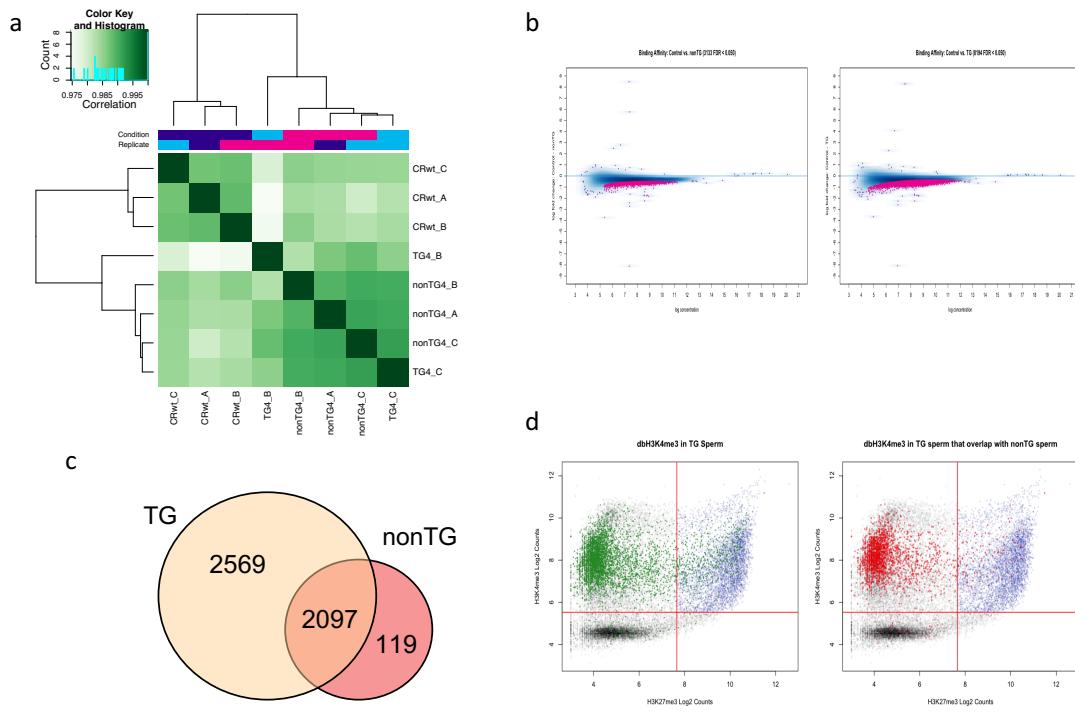
Supplementary Figure 3.5. Scatter plots of RNA seq data from control preimplantation embryos to highlight gene expression profile for dmH3K4me3 in early development(8).

Supplementary Figure 3.6. a) Genome coverage of an allele-specific ChIP-seq analysis of H3K4me3 enrichment in preimplantation embryos done on the dataset from Zhang et al., 2016

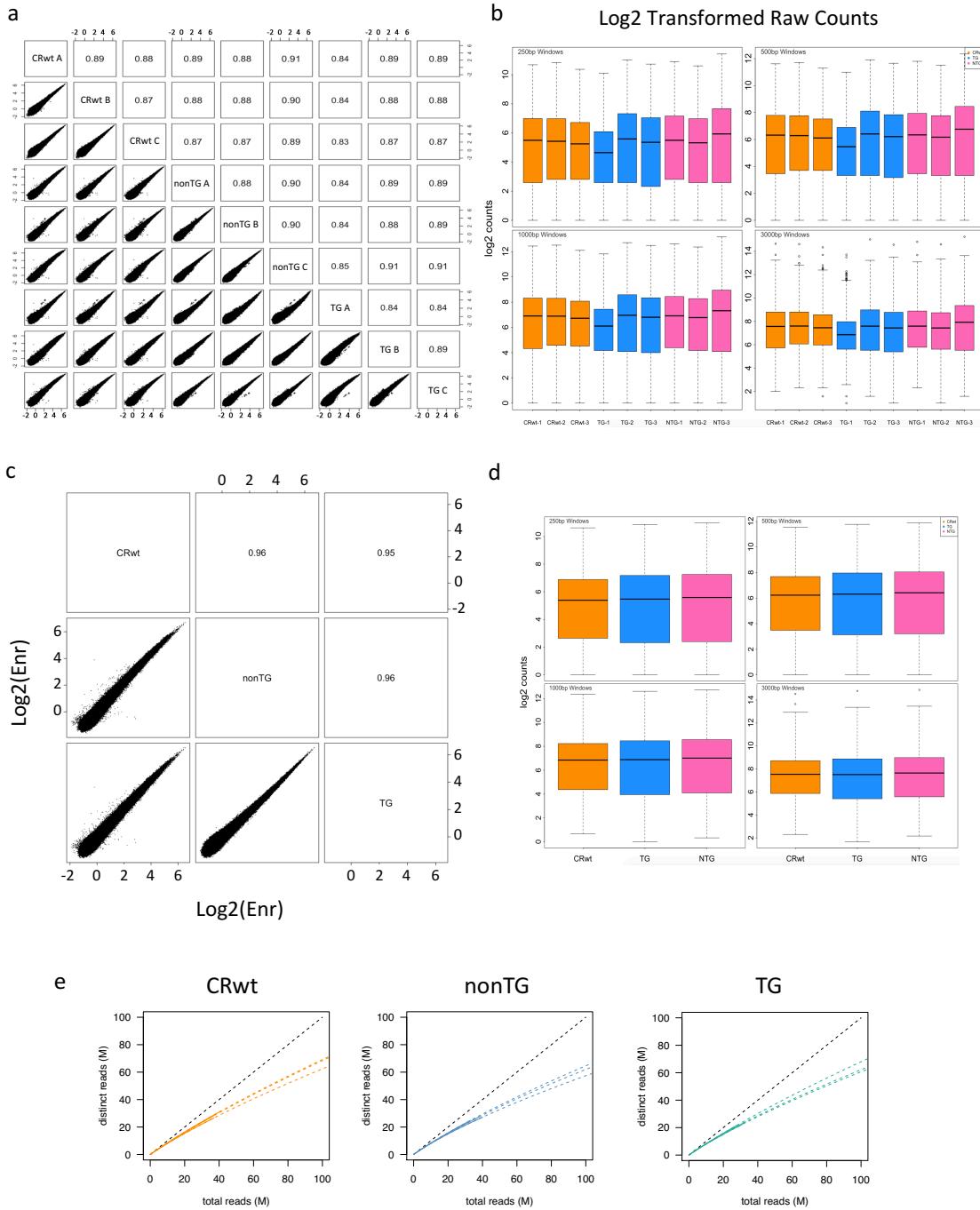
(8) and compared to sperm (this study). PWK (paternal genome) and C57BL/6 (maternal genome) reads were aligned to a C57BL/6 genome that was N-masked using the small nucleotide polymorphism variant coordinate file (VCF). All N-containing reads were then identified as paternal or maternal in origin and split into unique alignment files with SNP-Split. Read counts were then RPKM normalized and coverage tracks were generated with DeepTools. **b)** Count density of paternal H3K4me3 RPKM counts show paternal H3K4me3 read counts from 2-cell embryo are enriched over background. **c)** RPKM normalized read counts from both maternal and paternal chromosomes in 2-cell embryo (6) plotted against reads from 2-cell paternal chromosomes.

Supplementary Figure 3.7. a) Scatter plot to show KDM1A in the pachytene spermatocyte is found enriched at the same TSS ($\pm 1\text{ kbp}$) as H3K4me3 in sperm. **b)** Genes with dmH3K4me3 are found to contain more KDM1A in the spermatocyte and less H3K27me3 than those unchanged. **c-d)** H3K4me3 peak breadth is not preferentially altered in TG or nonTG sperm

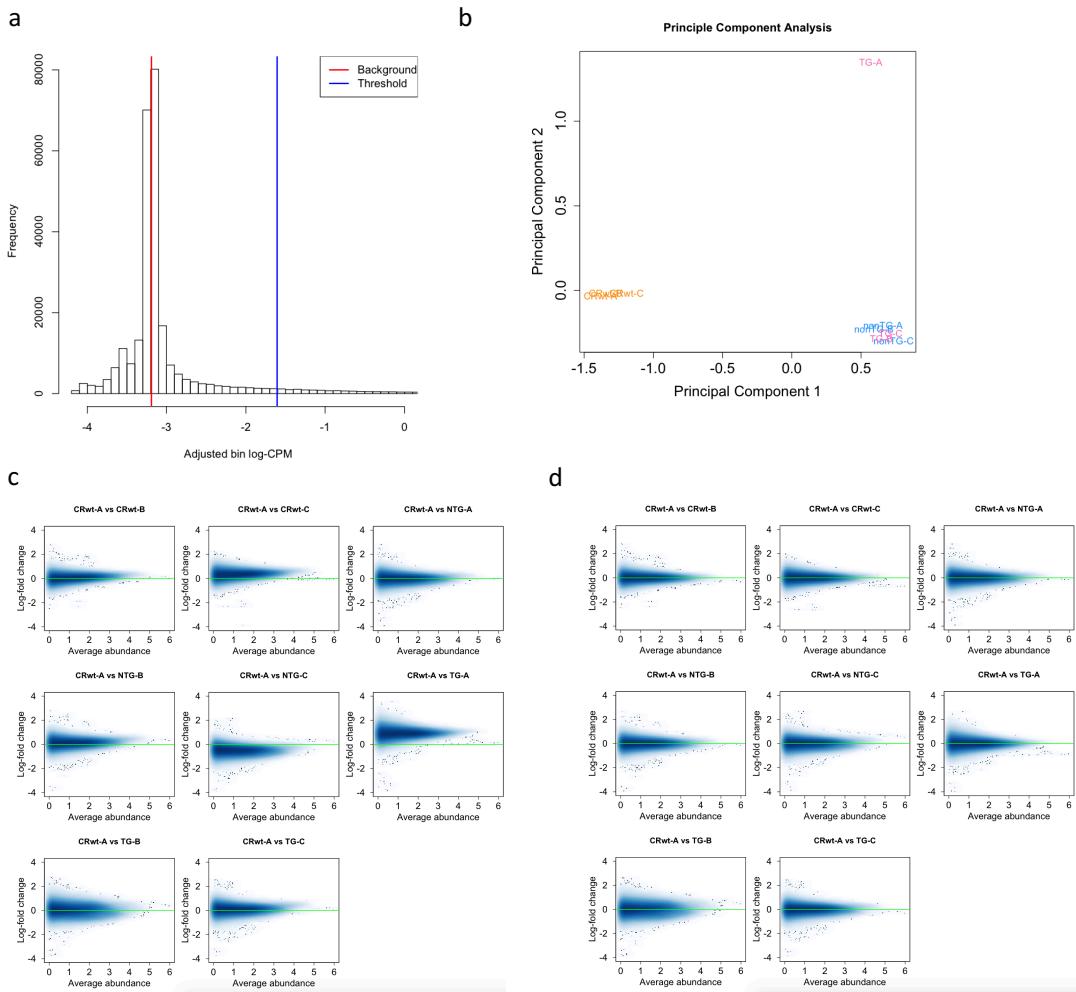
Supplemental Figure 3.1



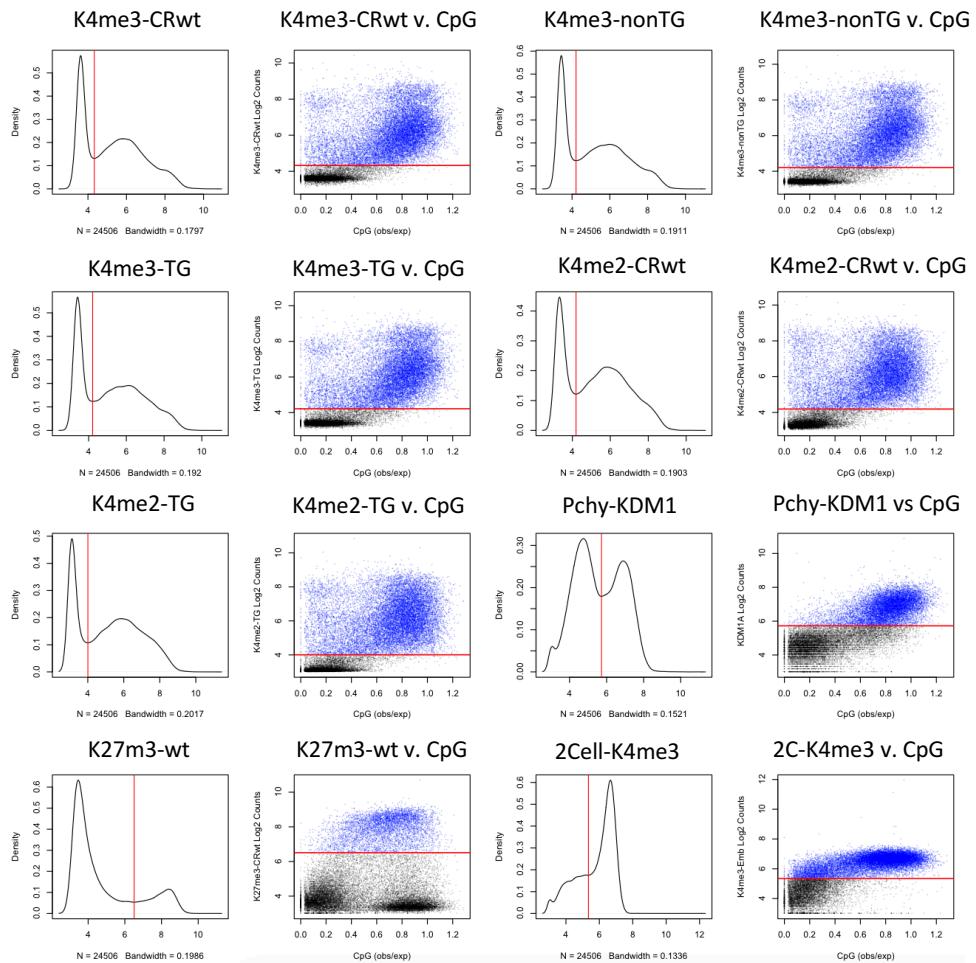
Supplemental Figure 3.2



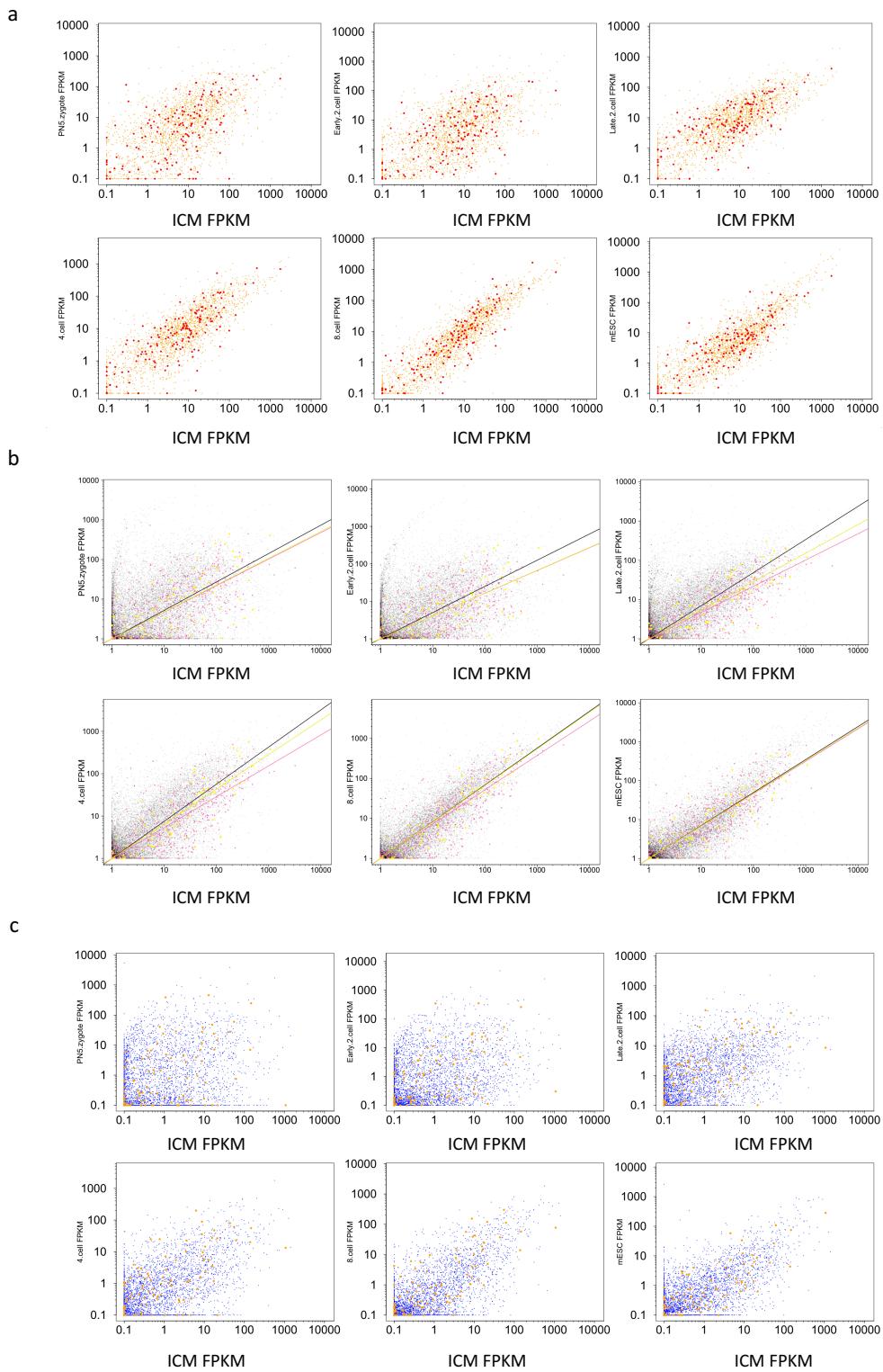
Supplemental Figure 3.3



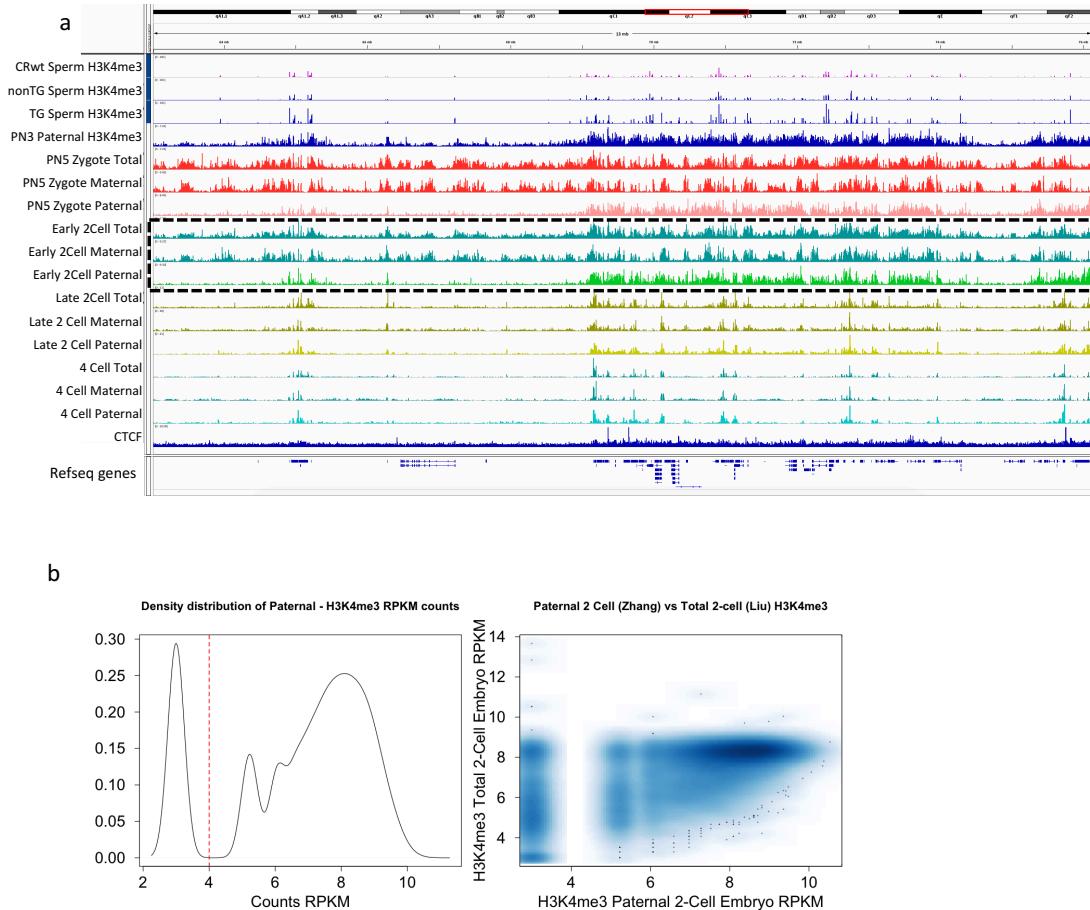
Supplemental Figure 3.4



Supplemental Figure 3.5



Supplemental Figure 3.6



Supplemental Figure 3.7

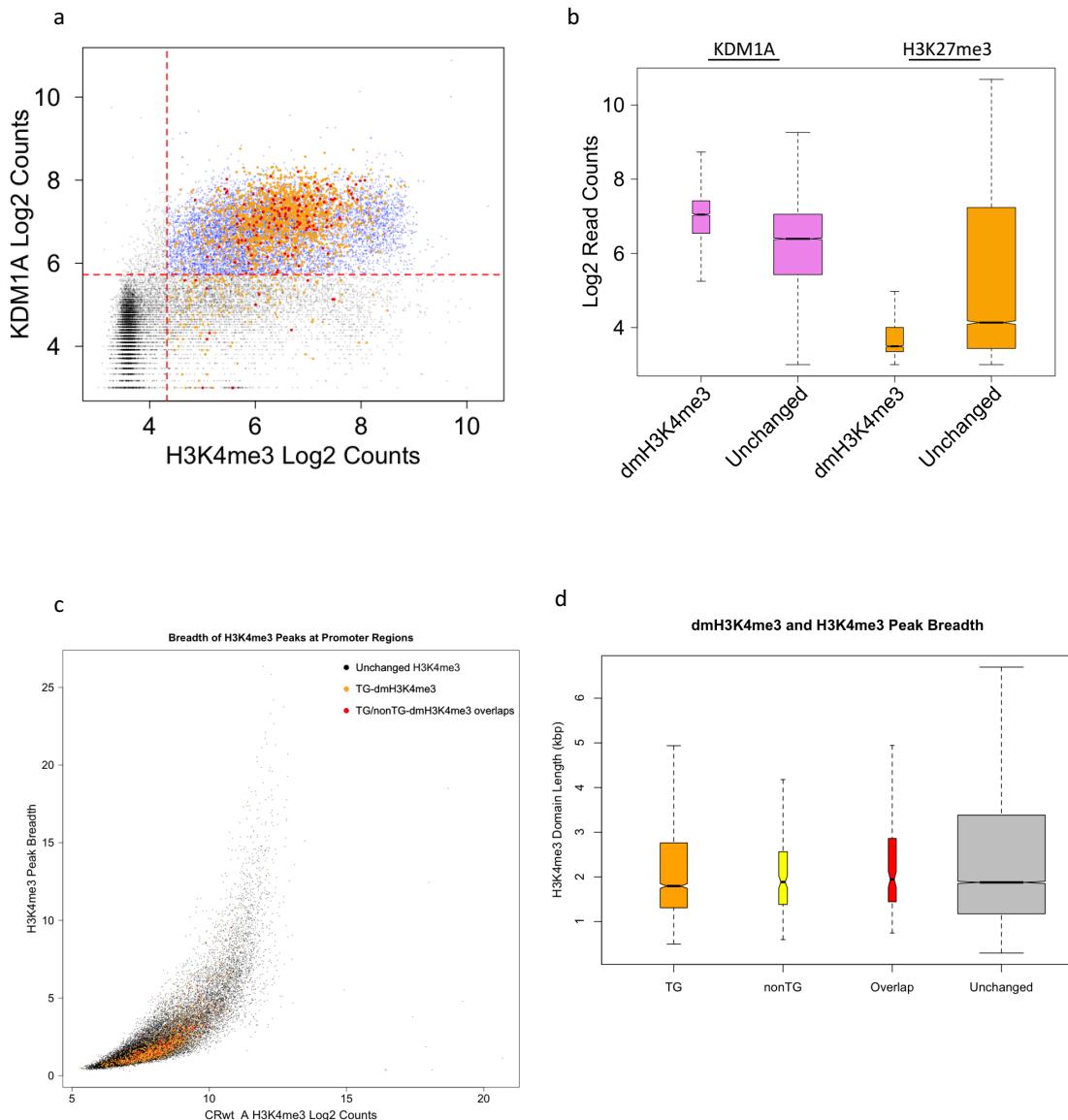


Table 1) Total sequencing effort and statistics for H3K4me3 and H3K27me3 sperm ChIP-seq

Sample ID	Factor	Aligned Reads	% Duplicate	Read Length	% Reads in Peaks	% Reads in Blacklist
CRwt-A	H3K4me3	37040546	28.4	100	49.8	13.5
CRwt-B	H3K4me3	41447857	23.4	100	40.3	10.9
CRwt-C	H3K4me3	37089922	23.8	100	39.8	13.3
NTG-A	H3K4me3	38265903	28.5	100	47.6	13.2
NTG-B	H3K4me3	33694577	25.2	100	47	11.7
NTG-C	H3K4me3	40050485	32.3	100	59.7	10.8
TG-A	H3K4me3	28528809	22.6	100	36.5	14.6
TG-B	H3K4me3	33318848	27.3	100	49.9	10.8
TG-C	H3K4me3	31946545	28.5	100	53	14
CRwt	H3K27me3	53897843	37.1	100	49.00	0.729
TG	H3K27me3	43629648	38.7	100	48.80	0.705
nonTG	H3K27me3	52114128	30.2	100	36.00	0.668
gDNA	Genomic DNA	83881151	11.4	100	N/A	3.98

3.9 Supplemental References

1. K. Siklenka *et al.*, Disruption of histone methylation in developing sperm impairs offspring health transgenerationally. *Science* **350**, aab2006 (2015).
2. J. G. Xia *et al.*, INMEX-a web-based tool for integrative meta-analysis of expression data. *Nucleic Acids Research* **41**, W63-W70 (2013).
3. J. G. Xia, M. J. Benner, R. E. W. Hancock, NetworkAnalyst - integrative approaches for protein-protein interaction network analysis and visual exploration. *Nucleic Acids Research* **42**, W167-W174 (2014).
4. J. Xia, E. E. Gill, R. E. Hancock, NetworkAnalyst for statistical, visual and network-based meta-analysis of gene expression data. *Nature protocols* **10**, 823-844 (2015).
5. J. Zhang *et al.*, SFMBT1 functions with LSD1 to regulate expression of canonical histone genes and chromatin-related factors. *Genes & development* **27**, 749-766 (2013).
6. X. Liu *et al.*, Distinct features of H3K4me3 and H3K27me3 chromatin domains in pre-implantation embryos. *Nature* **537**, 558-562 (2016).
7. L. Wang *et al.*, Programming and inheritance of parental DNA methylomes in mammals. *Cell* **157**, 979-991 (2014).
8. B. Zhang *et al.*, Allelic reprogramming of the histone modification H3K4me3 in early mammalian development. *Nature* **537**, 553-557 (2016).

Chapter 4: General Discussion

4.1 Transgenerational Effects from KDM1A Overexpression in Mouse Testis

4.1.1 Male role on offspring survivability

The transgenerational inheritance of paternal effects has been observable for decades, although examples of mammalian models remain infrequent and poorly described. Recent studies suggest DNA methylation, non-coding RNA and nucleosomes in sperm are possible epigenetic mediators of these effects, yet to date no clear mechanism has been agreed upon. The body of work done in this thesis has established a transgenic mouse model for the study of transgenerational epigenetic inheritance. In our model, we have demonstrated that the overexpression of a histone demethylase KDM1A in the early germ cells of murine testis is the major initiating event towards the establishment of an altered sperm epigenome. As a result, we have shown intergenerational molecular changes in histone methylation and sperm RNA content and have correlated these changes with transgenerational offspring phenotypes. More specifically, we observed poor offspring development two familial generations away from the initial exposure to transgenic KDM1A. These findings link histone methylation in mammalian sperm to a transgenerational role in offspring development, health, and disease.

Multiple paternal factors such as age, poor semen parameters, and DNA damage have been associated with embryonic loss in mouse and men (1-3); however, few studies have associated changes in sperm chromatin with poor pregnancy outcomes (4). The results of our model suggest that early and late stage development processes may have a significant paternal epigenetic component. TG and nonTG sired embryos at E18.5 showed evidence of major organ developmental problems with signs of edema and hydrocephaly (Figure 2.2b). Moreover, underdevelopment, and skeletal abnormalities of the cranium and mandible at E18.5 were

frequently observed across multiple generations. The dramatic reduction in offspring survivability at PND2 from both TG and nonTG sires was likely due to these phenotypes (Figure 2.1d). The observed early embryonic loss, or the human equivalent of preterm birth, further connects an abnormal sperm epigenome to poor pregnancy outcomes.

Remarkably, the pre-weaning survivability of TG-sired offspring deteriorated with successive TG generations (TG2-TG4; Figure 2.1). Cumulative paternal effects in the TG lineage are likely due to the repeated inheritance of an altered sperm epigenome caused by overexpression of KDM1A. In contrast, we observed a gradual dilution of the offspring phenotype at E18.5 among successive nonTG generations (nonTG3-nonTG5; Figure 2.3). The existence of these two phenomena is strong evidence for an epigenetic effect which escapes embryonic and germ-line reprogramming for multiple generations. The accumulation of an altered epigenome and a progressive phenotype across generations has been shown in *C. elegans* where the *Kdm1a* homologue, *spr5*, serves as a mediator of epigenetic reprogramming (5). But, the transgenic lineage of our model represents one of the first known observations of a progressive paternal inheritance in mammals.

Phenotypic analysis of this model included the examination of 15 litters, 150 embryos and 10 embryonic skeletons per group; however, it is likely that more penetrable phenotypes would be detected if an even more comprehensive examination was used. For example, an in-depth histopathological analysis of major organs may identify paternal effects that were previously undetectable by observation of gross embryo morphology. Moreover, due to the connection between an altered sperm epigenome and the affected gene processes being enriched for metabolic functions it is likely that metabolism of the offspring has been disrupted. This disruption may manifest as complex phenotypes such as diabetes, and cardiovascular disease (6).

This may also suggest that transgenic mice and their descendants may be more susceptible to nutritional challenge of high fat, low protein, or folate deficient diets. In addition to metabolism, some abnormalities in the sperm caused by aging, DNA damage and epimutations have recently been linked with increased incidence of psychological disorders and disease in children, such as autism spectrum, schizophrenia, anxiety, stress and depression (7-10). Anecdotal evidence of similar phenotypes was collected through basic observation of the transgenic mouse colony. As a result, it would be relevant to perform formal behavioural assessments on TG and nonTG sired offspring. With these possibilities in mind, our transgenic mouse model inspires a new research direction whereby we study the effects of diet and lifestyle on the offspring in fathers with an already altered sperm epigenome. This model would then represent a relevant population of men who live in regions of high-risk exposure to toxicants, mutagens, and nutritional instability.

4.2 Mechanisms Mediating Epigenetic Inheritance of Paternal Effects

4.2.1 The role of Transgenic KDM1A

The changes to the sperm epigenome reported in this model generate many questions relating to KDM1A's role in their establishment. Indeed, it is likely that the major reductions to H3K4me2 in the TG sperm are a direct result of KDM1A increased activity while overexpressed. But how then are relative gains of H3K4me3 produced by the activity of a demethylase? And how are they then maintained into the sperm of the next generation? To begin to answer these questions we must consider the biochemical mechanisms of KDM1A targeting and activity, and the effect that KDM1A overexpression may have on the biology of the germ cells as a whole. Epigenetic machinery with opposing functions is expressed and is active simultaneously in most cells; however, the distinct patterns of epigenetic features and the coordination of gene

expression implies the proteins involved are in an intricate regulatory network. The overexpression of transgenic KDM1A may disrupt this cellular balance, and, in turn, impact the sperm epigenome through direct or indirect means. In wildtype mouse spermatocytes, *Kdm1a* is dynamically regulated at both the transcriptional and protein level, with a peak in abundance during meiotic prophase in the pachytene spermatocyte (11, 12). In theory, transgenic overexpression of KDM1A would supersede any endogenous regulatory dynamics and maintain abnormal protein levels and activity throughout spermatogenesis.

However, KDM1A requires special consideration as its ability to target chromatin, and thus its biological role, relies heavily on its non-catalytic domains. A long tower domain which extends out from near its active site enables KDM1A to interact with a number of different protein complexes (13). For example, the pairing of KDM1A with CoREST has been well established as a requirement for targeted demethylation of nucleosomal H3K4me1/2 (14-16). KDM1A/CoREST binding to DNA occurs through weak electrostatic interactions between the Sant2 domain of CoREST and nucleosomal DNA (13, 15, 17). The low binding specificity of this interaction facilitates a KDM1A/CoREST “random walk” across the genome in search of suitable substrates (15, 18, 19). This method for substrate recognition by *Kdm1a*’s active site is dependent on selective electrostatic forces, and thus KDM1A/CoREST will probe regions of the epigenome until these conditions are satisfied. Therefore, the specificity of interaction is achieved through local modulation of the electrostatic chromatin environment at KDM1A target sites (15, 20, 21). The successful binding of nucleosome substrate induces an allosteric rotation of KDM1A/CoREST complex analogous to a nanoscale clamp (13, 17, 22). A secondary binding domain on the surface of the KDM1A enzyme may further enhance substrate specificity, and help induce stable interaction with the H3 tail (17). This multi-factor interaction provides added

stability where KDM1A can remain bound across timescales consistent with chromatin remodeling events (17, 20, 23). Stably bound KDM1A may then serve as a binding platform to recruit transcription factors, remodeling chaperones, ncRNA or other gene-regulatory elements to the target site. This biology, in the context of differential methylation detected at both direct (H3K4me2) and indirect (H3K4me3) substrates of KDM1A, suggests a secondary, non-enzymatic role in the cell that's governed by multiple protein and RNA interactions. To identify a clear mechanism for how histone methylation becomes altered prior to formation of the spermatozoon it would be important to understand these binding interactions and their timing in the context of our model.

The strong correlation between regions of sperm H3K4me2 depletion and endogenous KDM1A binding suggests that these changes are primarily driven through the KDM1A/CoREST interaction. However, due to the homogenous reduction of H3K4me2 from pooled sperm of heterozygous TG mice, the majority of transgenic KDM1A activity must have been initiated prior to the segregation of chromosomes in meiosis. If the action took place in post-meiotic spermatids, 50% of the mature sperm would not carry the transgenic KDM1A alleles and the resulting epigenetic phenotype would be undetectable when samples were pooled. Therefore, the stable binding and prolonged activity of transgenic KDM1A at promoter regions in the spermatocyte may have aberrantly reduced transcription and turnover at its endogenous target sites. The resulting suppression could reduce histone turnover and replacement at KDM1A bound regions, while stable binding would preserve the altered state throughout spermiogenesis and into the mature sperm. Indeed, a similar rational was used when deciding to assess sperm H3K27me3 for stable changes between generations. Genes co-enriched with H3K27me3 and H3K4me3 in sperm are associated with the canonical histone H3.1/2 and this implies that they

undergo low levels of nucleosome turnover during spermatogenesis (24). Any changes to H3K27me3 made in the spermatogonia and spermatocytes may have been maintained through to the sperm. Although we detected no changes to H3K27me3 in this model, a conditional UTX overexpressor mouse under the same control of our truncated EF1a promoter may answer this question directly.

Conversely, regions of increased sperm H3K4me3 also correlate with endogenous Kdm1a binding sites, but only partially intersect regions of H3K4me2 depletion (Figure 3.5B). These results suggest the histone methylation changes observed in this model are not directly related, and may point to an expanded, context dependent, role for KDM1A that is not limited to demethylation alone. In support of this hypothesis, KDM1A/CoREST has been associated with large protein complexes composed of opposing functional subunits. For example, the H3K4me3 methyltransferase ALL-1 forms a super-complex containing a mixture of 29 proteins that includes the KDM1A demethylase, CoREST and HDACs yet functions to remodel nucleosomes and promote transcription at promoter regions (25). It's unknown what role KDM1A plays in this major complex, but KDM1A overexpression may shift the dynamics of this relationship to dysregulate ALL-1 activity at its target sites. Additionally, KDM1A may be positioned to regions of H3K4me3 through a direct interaction with the chromodomain containing zinc finger protein CHD1 (26). Further KDM1A targeting is conferred through the activity of transcription factors, such as the SNAIL family involved in differentiation, specifically palate development in mice (27-29). These proteins contain histone H3-tail peptide mimics capable of engaging directly with the KDM1A binding pocket to guide remodeling activity. Moreover, non-histone proteins such as DNMT and lncRNAs such as *Hotair* or *Terra* can bind directly to KDM1A and direct it to telomeres and heterochromatin regions (30-32). Collectively, these interactions expand

KDM1A's possible roles to the regulation of gene expression through recruitment of transcription factors and chromatin compartmentalization. Chimeric KDM1A/B proteins which have been constructed to remove the tower domain but retain enzymatic activity may prove to be useful in teasing apart these interactions (33). We could also modify the chromatin binding domain on KDM1A's surface to characterize if stable binding to histone targets is sufficient to alter the epigenome. Moreover, overexpression of an enzymatically inactive KDM1A may help infer if epigenetic effects are a direct response to enzymatic activity, or abnormal physical interactions with separate effector proteins. Indeed, identifying the specific binding partners of transgenic KDM1A would be possible with co-immunoprecipitation of hsKDM1A followed by mass-spectrometry of purified proteins using cells isolated from different stages of spermatogenesis. Products of the pull-down assays could also be used in-vitro to develop high throughput methyltransferase activity experiments that help characterize the activity and kinetics of hsKDM1A and its binding partners. Although this would not necessarily be representative of the germ-line *in vivo*, it would give direction to begin implicating additional proteins, targets, and their combinatorial effects in the establishment of the inherited sperm epigenome.

4.2.2 Modes of intergenerational Inheritance

Knowledge of the molecular influence germ-line chromatin has on mammalian embryogenesis has been limited by cell-input requirements imposed by standard NGS techniques such as ChIP-seq. Recently, micro-scale ChIP protocols have successfully generated reproducible sequencing data from as little as 200 mouse preimplantation embryos (34-37). Analysis of active marks H3K4me3 (35-37) and H3K27ac (35) and repressive H3K27me3 (36, 37), H3K9me3 (unpublished) in the oocyte and embryo has revealed an unique chromatin landscape unlike that observed in somatic cells. For example, broad H3K4me3 chromatin

domains span partial or hypomethylated DNA regions in the oocyte, zygote, and early to late 2-cell preimplantation embryos. The establishment of this non-canonical pattern of H3K4me3 (ncH3K4me3) begins in the D7 oocyte and reaches maximum breadth by the MII oocyte. Upon fertilization, the ncH3K4me3 pattern is maternally inherited by the PN5 zygote and infiltrates the paternal chromatin immediately (37). In the zygote, the broad signal persists until the 2-cell stage, just prior to mouse ZGA (35-37), where it's then trimmed away to resume a canonical form that marks gene promoters in the 4-cell embryo (35-37). When compared to the tight, narrow, and isolated H3K4me3 peaks that we and others have described in sperm, the genomic regions marked by zygotic ncH3K4me3 appear to be in opposite phases (Supplemental Figure 3.6) (24, 38). In other words, the broad ncH3K4me3 peak boundaries of the zygote and early two-cell are associated with CpG islands and TSS and appear restricted by the regions of retained H3K4me3 in sperm. Moreover, allele specific ChIP-seq for paternally inherited H3K4me3 showed a similar broad methylation pattern as early as the PN5 zygote (37); however, the sequencing depth used was not sufficient to rule out technical artifacts as a possible cause.

Hi-C chromatin capture assays performed in single cell zygotes have shown a distinct lack of interaction between chromosomes from both paternal and maternal pronuclei and portray allelic differences in reprogramming of topically associating domains (TAD) (39). Allele-specific sequencing also highlights that the paternal and maternal chromosomes remain spatially distinct past the 8-cell stage (39). As a result, the eviction of paternal protamines, chromatin infiltration by naïve, maternally, supplied nucleosomes, and subsequent spreading of the broad methylation signal must be guided solely by factors contributed by the sperm. The retained paternal nucleosomes may function as book-ends, or sign posts, that limits the spread of maternally acquired nucleosomes, and serve as templates for the type of modification that is

written to them. A template mechanism would then propagate alterations to the paternal chromatin and effectively edit the epigenetic information for the development of the organism. It's possible that the altered paternal chromatin establishes a modified epigenetic ground-state that's carried forward through cell divisions, although ChIP-seq in embryos sired by TG and nonTG males would be required to test this hypothesis.

4.3 Regulatory Mechanisms for Epigenetic Information in Sperm

To date, most studies which examine the link between histone modifications and gene regulation are based on association and have typically showed distinct differences between chromatin states at active or repressed genes (40). Here we showed that changes in sperm histone methylation resulted in gene expression changes in the TG and nonTG sired embryos. Indeed, the selected genes were strongly related to developmental processes and correlate well with our observed phenotype. However, there has been little evidence describing a causal link between histone methylation status and transcriptional control. Ablation of a specific H3K4me3 histone methyltransferase in yeast will reduce global methylation levels but result in little change to global gene expression (41-43). Moreover, context specific changes to histone methylation have been associated with either transcriptional activation or repression (44). The removal of SET1A's catalytic activity in mESC induced localized reduction in gene signals with a corresponding reduction in gene expression (45). What effect does a change in chromatin signal then have on the regulation of underlying gene expression? In some cases, residual levels of H3K4me2/3 have been linked to an epigenetic memory of historically active transcription. High levels of H3K4me2/3 have been shown to persist hours after transcriptional shutdown, and sometimes span across cell divisions (46, 47). This stability may provide the cell with an ability

for rapid activation in response to environmental queues. Indeed, this may be linked to the inheritance of a transcriptional state (48, 49). In sperm, we've shown that retained H3K4me2/3 and H3K27me3 clustered by intensity of signal into groups of distinct functional pathways. It's unknown if these sperm signals facilitate an adaptive response in embryonic gene expression; nevertheless, within the detectable regions of altered TG sperm methylation are a subset of genes associated with increased expression in the two-cell embryo. The genes with altered two-cell expression are normally expressed at high levels in later stages of development (37). This suggests some precocious gene activation as a response to the altered sperm epigenome in this model. These results are one of the first examples of sperm histone methylation affecting gene expression in the next generation.

Recent analysis of the oocyte and early embryo show that in these transcriptionally quiescent cell stages, nearly 80% of the genes involved in the maternal-zygote transition are associated with the broad non-canonical form of H3K4me3 (35). Consequently, ncH3K4me3 may act as a novel regulatory mark involved in maintaining transcriptional quiescence in oocyte and the embryo until its removal at genome activation. In support of this opposing role, the conditional depletion of the histone methyltransferase KMT2B in primary follicles results in embryonic arrest at the 1 to 4-cell stage embryo (50). Moreover, reduction of H3K4me3 in the oocyte leads to reactivation of transcription (37). In many cases, the depletion of histone demethylases in the embryo resulted in a failure to develop to the blastocyst stage (35-37). Specifically, embryos depleted of KDM5A and KDM5B display downregulation at ZGA genes due, in part, to the spreading of broad H3K4me3 domains across embryonic promoters (35, 36).

The understanding of these context specific roles for both histone H3K4me3 and histone methylation remodelers in the sperm, oocyte and preimplantation embryo is just beginning.

Tools like deactivated recombinase Cas9 (dCas9) epigenetic editing are now capable of definitively explaining the role of histone methylation at targeted loci by positioning chromatin remodelers at target sites (51, 52). Recent studies have used dCas9 to target a H3K4me3 methylase to transcriptionally repressed loci and have observed context dependent activation of endogenous transcription that was directly associated with a gain of H3K4me3 signal (52). Future in-vivo applications of this powerful epigenetic tool will quickly begin to revolutionize the way we approach the question of causality of gene regulation (53-55). Using these techniques, many of the hypothesis that have arisen from our work can be directly tested with precision. For example, adopting an in vivo germ cell model with dCas9-KDM1A (51), would allow validation of candidate dmH3K4me3 and dmH3K4me2 regions as causal in the epigenetic inheritance observed in this model. Additionally, the targeting of histone modifications at germline enhancers or lncRNA known to be involved in craniofacial development (56) may directly test how epigenetic marks at sperm regulatory elements control embryonic development.

We've shown that changes to the sperm chromatin landscape also disrupted the TG and nonTG sperm RNA content. In human spermatozoa, highly expressed sperm RNAs correlate positively with genomic regions that retain nucleosomes and H3K4me3 (24, 57-59). For example, the sperm borne transcripts, EVX1 and EVX2 show evolutionary similarity to HOX genes and are involved in early embryonic development (60, 61), and map to regions of sperm-retained histone methylation (57). These sperm-RNA may serve as paternally contributed transcripts for the early embryo to seed the first stages of development. In our model, the differentially abundant transcripts were highly enriched for gene ontology terms associated with spermatogenesis and were likely the result of past meiotic pachytene spermatocyte and round

spermatid activity; however, some regulatory ncRNA were also identified. The differential abundance of lncRNA transcripts offer the fascinating possibility of RNA mediated disruption of developmental processes in the next generation. Indeed, specific embryonic knockouts of lncRNA has been associated with developmental abnormalities similar to those observed in this model (56). Moreover, modern RNA-seq methods have determined that approximately 25,000 of these transcripts exist in the sperm (62). Therefore, lncRNA in the sperm may serve as regulators of the early embryo by functioning as structural scaffolds to recruit epigenetic and transcriptional machinery (44). Moreover, enhancer RNA (eRNA) is generally associated with lncRNA transcripts. Sperm enhancers have been recently described in sperm, however it remains unknown how enhancer function is altered in KDM1A males. KDM1A is specific for H3K4me2 and H3K4me1, and H3K4me1 is associated with enhancers when in combination with H3K27ac. Due to the specificity of H3K4me2 depletion and KDM1A binding, it would be likely that H3K4me1 is altered in a similar way. Alternatively, tRNA derived small ncRNA (tsiRNA) have been implicated in mediating paternal epigenetic inheritance phenotypes in mice (9, 63, 64). These studies addressed the mechanistic role of sperm ncRNA through microinjections into developing embryos, but they do not address the role of chromatin in their origin. Our results point towards a fundamental role of sperm chromatin modification in the regulation of non-coding transcripts that are stored in the sperm. Future experiments may adopt the micro injection technique to definitively address the how KDM1A-overexpression-induced sperm RNA differences affect the embryo. This step would be fundamental in connecting two major facets of epigenetic inheritance and help to unify the field.

4.4 Next Generation Sequencing Techniques

4.4.1 Chromatin Immunoprecipitation Sequencing

In ChIP-seq, DNA sequences are enriched from chromatin fragments using an antibody derived against the specific histone modification of interest. Enzymatic approaches have been used to prepare native chromatin for ChIP-seq to improve resolution and efficiency of immunoprecipitation (IP) in low abundance rare cell types. Due to the low abundance of sperm nucleosomes, digestion of sperm chromatin by micrococcal nuclease (MNase) has been widely implemented. However, the tightly packed chromatin functions in part to prevent DNA damage by exonuclease activity. As a result, chromatin fragmentation with MNase requires sperm to be pre-treated with reducing agents like DTT to relax the toroid structure and facilitate DNA accessibility. The MNase fragmentation method works under the assumption that nucleosomal DNA wrapped around the histone octamer is protected from digestion, while the exposed linker DNA is preferentially digested. In theory, a high concentration of MNase will result in thorough chromatin fragmentation and the retention of only mono-nucleosomal DNA (150 bp). On the other hand, low concentrations result in partial digestion and nucleosomal DNA lengths that correspond to the length of consecutive nucleosome monomers. As a result of these different treatments, a controversial debate surrounding the genomic positioning of retained sperm nucleosomes has become important to address (65). Treatment of sperm chromatin with mild-MNase digestion showed a propensity for nucleosome enrichment at intergenic sites, with reduced representation at promoters (65). However, many others (24, 38, 58, 59, 66), including this work, have identified sperm nucleosomes primarily at the transcriptional start site (TSS) using high-concentration MNase chromatin fragmentation.

Evidence for MNase bias is driven by nucleosome positioning studies in well-established somatic cell models in *Drosophila* and yeast (67-70). The MNase-seq maps generated from high or low level MNase treatment in *Drosophila* cells produced the same distinctly different nucleosomal enrichment patterns that have been described in sperm (65, 68, 69). Likewise, Carone et al. were able to re-establish a TSS enrichment after titrating the MNase concentration from low to high levels in mESC (65). These results cannot be attributed to an inherent enzymatic sequence bias as nucleosome enrichment can be similarly tuned by reducing incubation temperatures (67). Therefore, it is the thermodynamics of the enzyme that determine the cut site and thus suggests that nucleosomes exist in two sensitivity states (67, 71, 72). Indeed, MNase-resistant nucleosomes are enriched at CpG containing di-nucleotides, while MNase-sensitive nucleosomes are found at AT rich sequences that are commonly described as nucleosome depleted in high concentration conditions (67, 73). In the case of Carone et al., a combination of mild-MNase digestion and the unique packaging of the sperm chromatin may preferentially enrich MNase-sensitive nucleosomes at the AT rich sequences outside of gene regions. Moreover, their protocol includes the formaldehyde crosslinking of DNA-protein complexes prior to MNase treatment in the already compact sperm. This added stability may further contribute to MNase resistance of nucleosomes at CpG rich promoter regions and reduce their representation at these sites (65). At the same time, it's possible that only histone dimers, and tetramers, rather than a full nucleosomal octamer, are associated with DNA and are positioned within chromatin. The reduced stability of this interaction creates hyper sensitive MNase-state that would not be detectable with extensive digestion protocols. As a result, the positioning of sperm nucleosomes may very well be a combination of both resistant and sensitive states that are detectable only through careful consideration of endonuclease sensitivity.

ChIP-seq libraries are prepared by adding adaptors to the end of DNA fragments which are then amplified with PCR cycles. Sequencing reads of the same starting position and length are computationally marked as a duplicate read that was generated through the PCR amplification process. However, how duplicate reads are treated in downstream analyses is subject to a number of biological considerations. For example, control of duplicate reads is generally preferred for peak calling because the methods used do not depend on absolute height, but simple changes in state. Without the removal of duplicates, automated pipelines may erroneously define phantom peaks from artifacts generated by library preparation or genome features like repetitive elements. In contrast, a “duplicate” read may be biologically meaningful, and not an artifact of PCR, especially in native ChIP-seq datasets. The positioning of the nucleosome restricts MNase cutting to the same few sequences of DNA flanking either side. Moreover, in sperm datasets with high coverage, the majority of informative reads will be limited to the small subset of regions that retain nucleosomes (1% - 8% of the genome). These specifics limit the number of possible unique start sites to a very small proportion of the genome. Therefore, removal of duplicates would cap the information available at these peak-regions, and as a result reduce the statistical power to detect differential enrichments (74, 75). To truly identify a technical duplicate in MNase data one must apply a unique molecular barcode for every DNA fragment in the sample that are then incorporated to the sequencing read. The marking of duplicates would then rely on matching molecular IDs to distinguish between the reads generated from PCR and those with biological relevance.

4.4.2 Sample normalization and differential methylation analysis

In order to effectively compare differences between sample groups, the sequencing reads within each dataset must be normalized to account for variance in library composition and

technical efficiency. The most common biases introduced between samples are due to the composition of the reads of the library. Differences in library size, or amplification efficiency at GC regions can vary between sample to sample (74, 76). Highly enriched regions consume more sequencing resources and thus suppress the representation of other regions. If there are differences in the magnitude of suppression, then erroneous differential calls may result. Trended biases occur in situations such as differences in ChIP antibody efficiency. For example, differences in antibody binding efficiency would have little effect at low-abundance regions, but as the signal increases, binding efficiency plays a larger and larger role. These non-linear trends require special consideration to remove.

We took two approaches to determine regions of differential methylation between groups. First, the reads in all sample libraries were scaled by a factor of the smallest total library size to normalize all regions of the genome. This method assumes the only differences between samples occur from differences in sequencing depth rather than experimental variance. In ChIP-seq of sperm, where only a small fraction of the genome contains nucleosomes and informative reads, the majority of the library is composed of background noise. If the composition of the background was driving variance between groups, but the differential binding was present in high abundance peak regions, then differential regions would shift in one direction. Indeed, the results from DiffBind appear to have behaved in this manner, with all regions showing an increase in differential methylation. To work around this, one may perform normalization on only the reads present in peak-regions, defined as “effective library size”. Effective library size may be more appropriate when overall binding is expected to be directly comparable between samples. A more sophisticated approach is Trimmed Mean of M-Values (TMM), which relies on the assumption that a core-set of genes between all samples are not differentially methylated

(77). TMM functions by trimming away putative DB-bins by removing extreme M-values (top and bottom 30%). Then, scaling factors are calculated using the remaining information and are applied to the counts of each window. If two samples are distinctly different, then biologically meaningful methylation differences would correspond with upper M-values and may be removed with this approach (77). Moreover, trended biases cannot be accounted for using linear scaling methods like TMM and would require normalization with a non-parametric linear regression normalization method such as locally weighted scatter plot smoothing (lowess) (77, 78).

Second, we applied csaw as described by Aaron Lun (75). Normalization factors were calculated by tiling the genome into large bins (10kbp) to determine the offset for which enrichment was sufficiently higher than background. This approach helped determine that we did have a slight, non-linear trended bias that was likely due to immunoprecipitation differences between samples. Therefore, we applied a non-linear loess normalization to our data. Csa's major advantage over DiffBind was the ability to focus directly on the sites of biological significance, while maintaining statistical control of sample variance and multiple comparison tests (75).

However, the biggest critique in chromatin biology is the heavy reliance on antibodies to infer position and quantity of nearly all histone modifications. The recently developed nucleosome spike-in approach has been successfully implemented into ChIP-seq protocols to help calibrate enrichment calculations across experiments (79, 80). These methodologies work by adding a known amount of an internal standard, such as a chromatin prepared from an unrelated species or semi-synthetic modified nucleosome-DNA complexes, to each sample. Antibodies for the protein of interest will enrich moieties on chromatin from both the sample and standard. When the sample is sequenced, species differences in DNA sequence allows for

computational deconvolution of the sample from standard. These techniques allow for unbiased quantification of enrichment efficiencies and may mitigate many of the known problems with antibody inefficiencies and specificities (79). Although using standards to calibrate immunoprecipitations is exciting, the technology may be cost prohibitive as all sequencing must be done with paired end reads, at a very high depth of coverage.

4.5 Considerations of the Model

4.5.1 A True Example of Epigenetic Inheritance

The clearing of the offspring phenotypes after multiple generations of nonTG breeding provides the strongest evidence that the paternal effects are mediated by non-genetic means. Nevertheless, we checked for signs of DNA instability in the sperm and testis via TUNNEL for DNA damage and the activation and mobilization of the Long Interspersed Nuclear Elements (LINE1). LINE1 activation could result from disruption of the epigenetic mechanisms regulating this retrotransposon family member and could lead to increased genomic instability and the subsequent disruption of endogenous gene regions. However, the probability of this resulting in such varied phenotypes in two lines is low. Likewise, microdeletions or mutations on sex chromosomes are ruled out when considering our observed phenotypes are not sex-biased. Importantly, the observed phenotype and underlying chromatin changes have been reproduced using a heat-inducible KDM1A expression mouse model (81).

4.5.2 Variability of Phenotype

Dosage consistency of transgenic KDM1A could be a source of variation in the model. Random insertion of transgene throughout the genome may not result in all copies having the

same expression pattern. For each insertion, the surrounding epigenetic factors, such as DNA methylation or heterochromatin spreading at centromeres, may occur independently in each cell type. The transgene integration site may also induce variation of expression akin to the position effect variegation (PEV) mosaic effect seen in *Drosophila* eye colour in PEV mutants (Reviewed in 82). It's currently unknown where in the genome hsKDM1A cDNA is located, but we see evidence for mosaic expression within cell types using transgene expression visualization techniques such as *in situ* hybridization (Figure 2.1C) and GFP reporter signal (Supplemental Figure 2.1C). In contrast, the cytoplasmic bridges that connect adjacent spermatogonia are partially preserved across meiotic divisions and facilitate the flow of cellular components (e.g hsKDM1A mRNA) between neighbouring daughter cells. These channels may actually create a homogenous distribution of the transgenic KDM1A within the syncytium.

Other sources of variation in transgene expression patterns may occur when subject to multiple recombination events such as breeding over multiple generations. During meiosis, homologous recombination can separate an inactivated transgene copy from the active copies and effectively restrict its expression to a subset of descendants (83). Combined with the inherent negative selection of the transgene due to cumulative offspring mortality, this model struggles with phenotype bottlenecking. To mitigate these effects, greater control of transgene expression is needed. One approach would be to implement a CRE-lox expression system nested into the transgenic construct. A flanked by lox-p (floxed) transcriptional stop sequence, such as the bovine growth hormone (BGH) poly-A sequence, would be inserted between the EF1a promoter and first exon of hsKDM1A to repress promoter transcription until its removal by a maternally supplied CRE-recombinase. Liberation of transcriptional inactivation could occur at any point prior to germ-line establishment at E6.25 and would allow for the use of well validated CRE-

mice. One potential pitfall is the dependence on a 100% efficient CRE recombination of the lox-p sites for KDM1A expression. A modernized technique could take advantage of CRISPR-Cas9 to excise the transcriptional stop using maternally supplied sgRNA for this region. This new model would also improve the mouse colony management with respect to transgenerational generations, and allow experiments performed in different years to begin from the first generational exposure to KDM1A (F_0).

4.5.3 Technical Limitations

The current transgenic construct was designed to constitutively express hsKDM1a in the gonads. Specifically, a spontaneous mutation in the EF1a promoter restricts transgenic KDM1A expression to the early germ cells (Figure 2.1A) (84). Knowledge of the precise timing of exogenous KDM1A expression is important to understanding how TG and nonTG sperm acquire an altered epigenome. Unfortunately, technical limitations in construct design left a non-functional flag tag and an attenuated reporter signal. Nevertheless, a combination of in-situ hybridization and eGFP visualization from fresh testis cryosections showed hsKDM1A is expressed throughout the full duration of spermatogenesis, but most abundant in the early stages (Figure 2.1A). The high level of intersection between sperm H3K4me2 depletion and spermatocyte's endogenous Kdm1a binding is representative of transgenic KDM1A function, but has not yet been directly associated. Moreover, better control of transgenic KDM1A identification or immune precipitation would facilitate binding assays, and mass spectrometry experiments to determine complex partners in purified cell types. To do these experiments, a 3X flag tag has been shown in our lab to greatly improve hsKDM1A detectability in GC1 cell lines (unpublished). Moreover, subcloning hsKDM1A cDNA into a new piRES-EGFP reporter vector

has dramatically increased eGFP reporter signal and will enable cell-stage localization of the transgene *in-vivo* (unpublished).

4.5.3.1 RNA-content of Sperm

Sperm transcripts are typically of very poor quality; heavily fragmented, with extensive truncation at the 3' end (57). Moreover, they exist at extremely low levels, at nearly 200 fold lower than somatic cells (57, 85). Ribosomal RNAs (rRNAs) are also excessively degraded, further contributing to the poor overall quality of a sperm's RNA content (86). Consequently, multiple protocols have been used to overcome these technical limitations (64, 85, 87, 88). When using RNA of low quality and quantity, the amplification protocol chosen greatly influences the results (89, 90). In hindsight, microarray hybridization has a limited ability to quantify sperm transcript abundance due to the effect of excessive degradation on probe-hybridization efficiency. A 3' degradation bias strongly impacts the hybridization probe signal, raw data and normalized counts. Moreover, further enrichment biases may have been introduced from the use of PCR to amplify the low abundance sperm transcripts for cDNA hybridization to probes. Moreover, this same reliance on probe hybridization limits potential for discovery and detection due to the pre-defined nature of the array. Although the gene ST array technologies consider multiple transcript variants, many ncRNA are processed into unknown forms. Therefore, transcript differences detected through microarray may be just the tip of the iceberg. Modern RNA-seq libraries are capable of circumventing some of these pitfalls. Future experiments should make use of these recent advancements to more clearly determine composition and relative abundance of RNA classes in the sperm of TG and nonTG descendants.

4.5.3.2 DNA methylation of sperm

The assessment of genome-wide DNA methylation of sperm in this model was limited by the inherent bias of Reduced Representation Bisulphite Sequencing (RRBS) technique to identify 5mC inside CGIs. Moreover, targeted approaches focused on regions with H3K4me2 depletion, which are highly focused at the hypomethylated CGIs in gene promoters. Disruption of H3K4me2 at these regions should have predisposed them to aberrant DNA methylation patterns; however, methylation state may not be sufficient on its own to disrupt the highly conserved hypomethylation of promoters. Consequently, we now hypothesize that CpGs outside of hypomethylated CGIs are more susceptible to epigenetic insults and it's likely that technical limitations failed to detect these regions. A thorough analysis would require whole-genome-bisulphite-sequencing. This approach would allow for detection of DNA methylation changes at intergenic and repeat elements, cryptic promoters and other hypermethylated sperm regions. The rationale behind looking for changes at these sites is the tight association between H3K9me and DNA hypermethylation in the sperm. Moreover, cycle specific interactions between DNMT1 and KDM1A show their localization at heterochromatin regions during the S-phase (30). In addition, H3K9 methylation is a substrate for KDM1A and would be a great non-promoter candidate to determine if intergenic regulatory regions have been disturbed in this model.

4.6 Conclusions

The outcomes of this thesis show for the first time that mammalian sperm histones and their modifications play a role in embryo development. These epigenetic signals may be an underlying cause of birth defects and disease that persists for multiple generations. In addition, this work provides an improved understanding of the molecular mechanism behind transgenerational epigenetic inheritance. By showing that changes to sperm histone methylation

can be stably maintained between generations, we provide a framework from which future studies can expand on how these signals are inherited and escape epigenetic reprogramming. We hope that this research will change how we think about inheritance and paternal effects, and that it may open doors towards the development of pre-conception advising and healthier offspring.

4.7 References

1. R. Sharma *et al.*, Effects of increased paternal age on sperm quality, reproductive outcome and associated epigenetic risks to offspring. *Reproductive Biology and Endocrinology* **13**, 35 (2015).
2. G. M. Bareh, E. Jacoby, P. Binkley, R. S. Schenken, R. D. Robinson, Sperm deoxyribonucleic acid fragmentation assessment in normozoospermic male partners of couples with unexplained recurrent pregnancy loss: a prospective study. *Fertility and sterility* **105**, 329-336. e321 (2016).
3. T. G. Jenkins, K. I. Aston, T. Meyer, D. T. Carrell, in *The Male Role in Pregnancy Loss and Embryo Implantation Failure*. (Springer, 2015), pp. 81-93.
4. R. Azpiazu *et al.*, High-throughput sperm differential proteomics suggests that epigenetic alterations contribute to failed assisted reproduction. *Hum Reprod* **29**, 1225-1237 (2014).
5. D. J. Katz, T. M. Edwards, V. Reinke, W. G. Kelly, A C. elegans LSD1 demethylase contributes to germline immortality by reprogramming epigenetic memory. *Cell* **137**, 308-320 (2009).
6. P. D. Gluckman, M. A. Hanson, T. Buklijas, F. M. Low, A. S. Beedle, Epigenetic mechanisms that underpin metabolic and cardiovascular diseases. *Nature Reviews Endocrinology* **5**, 401-408 (2009).
7. J. I. Feinberg *et al.*, Paternal sperm DNA methylation associated with early signs of autism risk in an autism-enriched cohort. *Int J Epidemiol* **44**, 1199-1210 (2015).
8. B. G. Dias, K. J. Ressler, Parental olfactory experience influences behavior and neural structure in subsequent generations. *Nature neuroscience* **17**, 89-96 (2014).
9. A. B. Rodgers, C. P. Morgan, N. A. Leu, T. L. Bale, Transgenerational epigenetic programming via sperm microRNA recapitulates effects of paternal stress. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 13699-13704 (2015).
10. K. Gapp *et al.*, Early life stress in fathers improves behavioural flexibility in their offspring. *Nature communications* **5**, 5466 (2014).
11. J. Zhang *et al.*, SFMBT1 functions with LSD1 to regulate expression of canonical histone genes and chromatin-related factors. *Genes & development* **27**, 749-766 (2013).
12. M. Godmann *et al.*, Dynamic regulation of histone H3 methylation at lysine 4 in mammalian spermatogenesis. *Biol Reprod* **77**, 754-764 (2007).

13. R. Baron, N. A. Vellore, LSD1/CoREST is an allosteric nanoscale clamp regulated by H3-histone-tail molecular recognition. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 12509-12514 (2012).
14. M. G. Lee, C. Wynder, N. Cooch, R. Shiekhattar, An essential role for CoREST in nucleosomal histone 3 lysine 4 demethylation. *Nature* **437**, 432-435 (2005).
15. S. Pilotto *et al.*, Interplay among nucleosomal DNA, histone tails, and corepressor CoREST underlies LSD1-mediated H3 demethylation. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 2752-2757 (2015).
16. A. P. Barrios *et al.*, Differential properties of transcriptional complexes formed by the CoREST family. *Mol Cell Biol* **34**, 2760-2770 (2014).
17. J. M. Burg, J. J. Gonzalez, K. R. Maksimchuk, D. G. McCafferty, Lysine-Specific Demethylase 1A (KDM1A/LSD1): Product Recognition and Kinetic Analysis of Full-Length Histones. *Biochemistry* **55**, 1652-1662 (2016).
18. G. Al-Ani, S. S. Malik, A. Eastlund, K. Briggs, C. J. Fischer, ISWI remodels nucleosomes through a random walk. *Biochemistry* **53**, 4346-4357 (2014).
19. J. N. McKnight, K. R. Jenkins, I. M. Nodelman, T. Escobar, G. D. Bowman, Extranucleosomal DNA binding directs nucleosome sliding by Chd1. *Molecular and cellular biology* **31**, 4746-4759 (2011).
20. F. Forneris *et al.*, A highly specific mechanism of histone H3-K4 recognition by histone demethylase LSD1. *J Biol Chem* **281**, 35289-35295 (2006).
21. F. Forneris, C. Binda, M. A. Vanoni, E. Battaglioli, A. Mattevi, Human histone demethylase LSD1 reads the histone code. *Journal of Biological Chemistry* **280**, 41360-41365 (2005).
22. R. Baron, C. Binda, M. Tortorici, J. A. McCammon, A. Mattevi, Molecular mimicry and ligand recognition in binding and catalysis by the histone demethylase LSD1-CoREST complex. *Structure* **19**, 212-220 (2011).
23. F. Forneris, C. Binda, A. Adamo, E. Battaglioli, A. Mattevi, Structural basis of LSD1-CoREST selectivity in histone H3 recognition. *J Biol Chem* **282**, 20070-20074 (2007).
24. S. Erkek *et al.*, Molecular determinants of nucleosome retention at CpG-rich sequences in mouse spermatozoa. *Nature structural & molecular biology* **20**, 868-875 (2013).
25. T. Nakamura *et al.*, ALL-1 is a histone methyltransferase that assembles a supercomplex of proteins involved in transcriptional regulation. *Molecular cell* **10**, 1119-1128 (2002).
26. E. Metzger *et al.*, Assembly of methylated KDM1A and CHD1 drives androgen receptor-dependent transcription and translocation. *Nature structural & molecular biology* **23**, 132-139 (2016).
27. M. A. Nieto, The snail superfamily of zinc-finger transcription factors. *Nat Rev Mol Cell Biol* **3**, 155-166 (2002).
28. Y. Lin *et al.*, The SNAG domain of Snail1 functions as a molecular hook for recruiting lysine-specific demethylase 1. *The EMBO journal* **29**, 1803-1816 (2010).
29. S. A. Murray, K. F. Oram, T. Gridley, Multiple functions of Snail family genes during palate development in mice. *Development* **134**, 1789-1797 (2007).
30. C. Brenner *et al.*, The interplay between the lysine demethylase KDM1A and DNA methyltransferases in cancer cells is cell cycle dependent. *Oncotarget* **7**, 58939-58952 (2016).

31. M. C. Tsai *et al.*, Long noncoding RNA as modular scaffold of histone modification complexes. *Science* **329**, 689-693 (2010).
32. A. Porro, S. Feuerhahn, J. Lingner, TERRA-reinforced association of LSD1 with MRE11 promotes processing of uncapped telomeres. *Cell reports* **6**, 765-776 (2014).
33. J. M. Burg *et al.*, A rationally-designed chimeric KDM1A/KDM1B histone demethylase tower domain deletion mutant retaining enzymatic activity. *FEBS letters* **589**, 2340-2346 (2015).
34. J. Brind'Amour *et al.*, An ultra-low-input native ChIP-seq protocol for genome-wide profiling of rare cell populations. *Nature communications* **6**, 6033 (2015).
35. J. A. Dahl *et al.*, Broad histone H3K4me3 domains in mouse oocytes modulate maternal-to-zygotic transition. *Nature* **537**, 548-552 (2016).
36. X. Liu *et al.*, Distinct features of H3K4me3 and H3K27me3 chromatin domains in pre-implantation embryos. *Nature* **537**, 558-562 (2016).
37. B. Zhang *et al.*, Allelic reprogramming of the histone modification H3K4me3 in early mammalian development. *Nature* **537**, 553-557 (2016).
38. Y. H. Jung *et al.*, Chromatin States in Mouse Sperm Correlate with Embryonic and Adult Regulatory Landscapes. *Cell Rep* **18**, 1366-1382 (2017).
39. Y. Ke *et al.*, 3D Chromatin Structures of Mature Gametes and Structural Reprogramming during Mammalian Embryogenesis. *Cell* **170**, 367-381 e320 (2017).
40. E. P. Consortium *et al.*, Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**, 799-816 (2007).
41. T. Clouaire, S. Webb, A. Bird, Cfp1 is required for gene expression-dependent H3K4 trimethylation and H3K9 acetylation in embryonic stem cells. *Genome biology* **15**, 451 (2014).
42. T. Clouaire *et al.*, Cfp1 integrates both CpG content and gene activity for accurate H3K4me3 deposition in embryonic stem cells. *Genes & development* **26**, 1714-1728 (2012).
43. T. L. Lenstra *et al.*, The specificity and topology of chromatin interaction pathways in yeast. *Molecular cell* **42**, 536-549 (2011).
44. R. Andersson, A. Sandelin, C. G. Danko, A unified architecture of transcriptional regulatory elements. *Trends Genet* **31**, 426-433 (2015).
45. C. C. Sze *et al.*, Histone H3K4 methylation-dependent and -independent functions of Set1A/COMPASS in embryonic stem cell self-renewal and differentiation. *Genes & development* **31**, 1732-1737 (2017).
46. H. H. Ng, F. Robert, R. A. Young, K. Struhl, Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. *Molecular cell* **11**, 709-719 (2003).
47. V. Borde *et al.*, Histone H3 lysine 4 trimethylation marks meiotic recombination initiation sites. *The EMBO journal* **28**, 99-111 (2009).
48. Y. Ding, M. Fromm, Z. Avramova, Multiple exposures to drought 'train' transcriptional responses in Arabidopsis. *Nature communications* **3**, ncomms1732 (2012).
49. T. Muramoto, I. Müller, G. Thomas, A. Melvin, J. R. Chubb, Methylation of H3K4 is required for inheritance of active transcriptional states. *Current Biology* **20**, 397-406 (2010).
50. S. Kageyama *et al.*, Alterations in epigenetic modifications during oocyte growth in mice. *Reproduction* **133**, 85-94 (2007).

51. N. A. Kearns *et al.*, Functional annotation of native enhancers with a Cas9-histone demethylase fusion. *Nat Methods* **12**, 401-403 (2015).
52. D. Cano-Rodriguez *et al.*, Writing of H3K4Me3 overcomes epigenetic silencing in a sustained but context-dependent manner. *Nature communications* **7**, 12284 (2016).
53. J. B. Black *et al.*, Targeted Epigenetic Remodeling of Endogenous Loci by CRISPR/Cas9-Based Transcriptional Activators Directly Converts Fibroblasts to Neuronal Cells. *Cell Stem Cell* **19**, 406-414 (2016).
54. I. B. Hilton *et al.*, Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nat Biotechnol* **33**, 510-517 (2015).
55. T. S. Klann *et al.*, CRISPR-Cas9 epigenome editing enables high-throughput screening for functional regulatory elements in the human genome. *Nat Biotechnol* **35**, 561-568 (2017).
56. M. Sauvageau *et al.*, Multiple knockout mouse models reveal lincRNAs are required for life and brain development. *Elife* **2**, e01749 (2013).
57. E. Sendler *et al.*, Stability, delivery and functions of human sperm RNAs at fertilization. *Nucleic acids research* **41**, 4104-4117 (2013).
58. U. Brykczynska *et al.*, Repressive and active histone methylation mark distinct promoters in human and mouse spermatozoa. *Nature structural & molecular biology* **17**, 679-687 (2010).
59. S. S. Hammoud *et al.*, Distinctive chromatin in human sperm packages genes for embryo development. *Nature* **460**, 473-478 (2009).
60. J. Garcia-Fernandez, The genesis and evolution of homeobox gene clusters. *Nat Rev Genet* **6**, 881-892 (2005).
61. M. Kalisz, M. Winzi, H. C. Bisgaard, P. Serup, EVEN-SKIPPED HOMEOBOX 1 controls human ES cell differentiation by directly repressing GOOSECOID expression. *Developmental biology* **362**, 94-103 (2012).
62. X. Zhang *et al.*, Systematic identification and characterization of long non-coding RNAs in mouse mature sperm. *PLoS One* **12**, e0173402 (2017).
63. Q. Chen *et al.*, Sperm tsRNAs contribute to intergenerational inheritance of an acquired metabolic disorder. *Science* **351**, 397-400 (2016).
64. U. Sharma *et al.*, Biogenesis and function of tRNA fragments during sperm maturation and fertilization in mammals. *Science* **351**, 391-396 (2016).
65. B. R. Carone *et al.*, High-resolution mapping of chromatin packaging in mouse embryonic stem cells and sperm. *Developmental cell* **30**, 11-22 (2014).
66. M. Hisano *et al.*, Genome-wide chromatin analysis in mature mouse and human spermatozoa. *Nature protocols* **8**, 2449-2470 (2013).
67. R. V. Chereji *et al.*, Genome-wide profiling of nucleosome sensitivity and chromatin accessibility in *Drosophila melanogaster*. *Nucleic acids research* **44**, 1036-1051 (2016).
68. R. V. Chereji, J. Ocampo, D. J. Clark, MNase-Sensitive Complexes in Yeast: Nucleosomes and Non-histone Barriers. *Molecular cell* **65**, 565-577 e563 (2017).
69. S. Kubik, M. J. Bruzzone, B. Albert, D. Shore, A Reply to "MNase-Sensitive Complexes in Yeast: Nucleosomes and Non-histone Barriers," by Chereji et al. *Molecular cell* **65**, 578-580 (2017).

70. S. Kubik, M. J. Bruzzone, D. Shore, Establishing nucleosome architecture and stability at promoters: Roles of pioneer transcription factors and the RSC chromatin remodeler. *Bioessays* **39**, (2017).
71. Y. Xi, J. Yao, R. Chen, W. Li, X. He, Nucleosome fragility reveals novel functional states of chromatin and poises genes for activation. *Genome research* **21**, 718-724 (2011).
72. B. Knight *et al.*, Two distinct promoter architectures centered on dynamic nucleosomes control ribosomal protein gene transcription. *Genes & development* **28**, 1695-1709 (2014).
73. G. Locke, D. Tolkunov, Z. Moqtaderi, K. Struhl, A. V. Morozov, High-throughput sequencing reveals a simple model of nucleosome energetics. *Proceedings of the National Academy of Sciences* **107**, 20998-21003 (2010).
74. S. G. Landt *et al.*, ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome research* **22**, 1813-1831 (2012).
75. A. T. Lun, G. K. Smyth, csaw: a Bioconductor package for differential binding analysis of ChIP-seq data using sliding windows. *Nucleic acids research* **44**, e45 (2016).
76. C. S. Ross-Innes *et al.*, Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature* **481**, 389-393 (2012).
77. M. D. Robinson, A. Oshlack, A scaling normalization method for differential expression analysis of RNA-seq data. *Genome biology* **11**, R25 (2010).
78. K. V. Ballman, D. E. Grill, A. L. Oberg, T. M. Therneau, Faster cyclic loess: normalizing RNA arrays via linear models. *Bioinformatics* **20**, 2778-2786 (2004).
79. A. T. Grzybowski, Z. Chen, A. J. Ruthenburg, Calibrating ChIP-Seq with Nucleosomal Internal Standards to Measure Histone Modification Density Genome Wide. *Molecular cell* **58**, 886-899 (2015).
80. D. A. Orlando *et al.*, Quantitative ChIP-Seq normalization reveals global modulation of the epigenome. *Cell Rep* **9**, 1163-1170 (2014).
81. S. Perez-Cerezales *et al.*, Elimination of methylation marks at lysines 4 and 9 of histone 3 (H3K4 and H3K9) of spermatozoa alters offspring phenotype. *Reprod Fertil Dev* **29**, 740-746 (2017).
82. S. C. Elgin, G. Reuter, Position-effect variegation, heterochromatin formation, and gene silencing in *Drosophila*. *Cold Spring Harbor perspectives in biology* **5**, a017780 (2013).
83. T. J. Schulz *et al.*, Variable expression of Cre recombinase transgenes precludes reliable prediction of tissue-specific gene disruption by tail-biopsy genotyping. *PLoS One* **2**, e1013 (2007).
84. T. Furuchi, K. Masuko, Y. Nishimune, M. Obinata, Y. Matsui, Inhibition of testicular germ cell apoptosis and differentiation in mice misexpressing Bcl-2 in spermatogonia. *Development* **122**, 1703-1709 (1996).
85. G. D. Johnson, P. Mackie, M. Jodar, S. Moskovtsev, S. A. Krawetz, Chromatin and extracellular vesicle associated sperm RNAs. *Nucleic acids research* **43**, 6847-6859 (2015).
86. G. D. Johnson *et al.*, Cleavage of rRNA ensures translational cessation in sperm at fertilization. *Mol Hum Reprod* **17**, 721-726 (2011).
87. W. Yan, Potential roles of noncoding RNAs in environmental epigenetic transgenerational inheritance. *Mol Cell Endocrinol* **398**, 24-30 (2014).

88. R. Goodrich, G. Johnson, S. A. Krawetz, The preparation of human spermatozoal RNA for clinical analysis. *Archives of andrology* **53**, 161-167 (2007).
89. X. Adiconis *et al.*, Comparative analysis of RNA sequencing methods for degraded or low-input samples. *Nat Methods* **10**, 623-629 (2013).
90. S. Mao, E. Sendler, R. J. Goodrich, R. Hauser, S. A. Krawetz, A comparison of sperm RNA-seq methods. *Systems biology in reproductive medicine* **60**, 308-315 (2014).

END