University of Massachusetts Medical School

eScholarship@UMMS

GSBS Dissertations and Theses   Graduate School of Biomedical Sciences

Paternal Effects on Metabolism in Mammals: A Dissertation

Jeremy M. Shea
University of Massachusetts Medical School

Let us know how access to this document benefits you.

Follow this and additional works at: https://escholarship.umassmed.edu/gsbs_diss

Part of the Biology Commons, Cellular and Molecular Physiology Commons, and the Genomics Commons

PATERNAL EFFECTS ON METABOLISM IN MAMMALS

A Dissertation Presented

By

Jeremy Michael Shea

Submitted to the Faculty of the

University of Massachusetts Graduate School of Biomedical Sciences, Worcester

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

MARCH 19, 2015

INTERDISCIPLINARY GRADUATE PROGRAM

A Dissertation Presented By

JEREMY MICHAEL SHEA

The signatures of the Dissertation Defense Committee signify completion and
approval as to style and content of the Dissertation

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

Thomas Fazzio Ph.D., Member of Committee

Jeanne Lawrence Ph.D., Member of Committee

Bernardo Lemos Ph.D., Member of Committee

Yong-Xu Wang Ph.D., Member of Committee

The signature of the Chair of the Committee signifies that the written dissertation
meets the requirements of the Dissertation Committee

Job Dekker Ph.D., Chair of Committee

The signature of the Dean of the Graduate School of Biomedical Sciences
signifies that the student has met all graduation requirements of the school.

Anthony Carruthers, Ph.D., Dean of the Graduate School of Biomedical Sciences

Program in Basic Biomedical Sciences

March 19, 2015

Dedication

This work is dedicated to all those who have supported me through this journey. First and foremost, my parents instilled in me a keen interest for the understanding how the world works. My brothers provided me with a caring competitiveness that drives me in any affair. I have made many amazing friends over the years that have challenged my ways of thinking, which has led insights that would not be possible without them. To Oliver Rando and members of the Rando laboratory, you are the smartest and most driven people I have met, and I aspire to be like all of you.

Acknowledgements

Abstract

The following work demonstrates that paternal diet controls medically important metabolic phenotypes in offspring. We observe transmission of dietary information to the zygote via sperm, and this information evades reprogramming that typically occurs after fertilization. Cytosine methylation is implicated as a major contributor to meiotic epigenetic inheritance in several transgenerational phenomena. Our extensive characterization of the sperm methylome reveals that diet does not significantly affect methylation patterns. However, we find that extensive epivariability in the sperm epigenome makes important contributions to offspring variation. Importantly, coordinate cytosine methylation and copy number changes over the ribosomal DNA locus contributes to variation in offspring metabolism. Thus, rDNA variability acts independently of post-adolescent paternal diet to influence offspring metabolism. Therefore, at least two mechanisms exist for epigenetically controlling offspring metabolism: stochastic epivariation and diet acting by an unknown mechanism to further modulate metabolism. This work argues that an offspring's phenotype can no longer be viewed solely as the result of genetic interactions with the developmental environment - the additional influences of paternal environment and inherited epigenetic variability must also be considered. These findings reveal novel contributions to metabolism that could revolutionize how we think about the risk factors for human health and disease.

Table of Contents

List of Tables[a]

---

[a] Since the majority of tables derived from genomics analysis are extremely large, they have mostly been excluded from the printed version of this thesis. Tables from Chapter I can be found at http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3039484/. Tables from other chapters are

List of Figures

List of Symbols, Abbreviations or Nomenclature

Acronym	Term
5mC	5-Methylcytosine
CGI	CpG Island
CNV	Copy Number Variation
DMR	Differentially Methylated Region
ICM	Inner Cell Mass
IVF	In vitro fertilization
RRBS	Reduced Representation Bisulfite Sequencing
rDNA	Ribosomal DNA
TEI	Transgenerational Epigenetic Inheritance
TF	Transcription Factor
WGBS	Whole Genome Bisulfite Sequencing

Preface

Chapter I is reprinted from the following co-authored work:

Paternally induced transgenerational environmental reprogramming of metabolic gene expression in mammals

Benjamin R. Carone*, Lucas Fauquier*, Naomi Habib*, Jeremy M. Shea*, Caroline E. Hart, Ruowang Li, Christoph Bock, Chengjian Li, Hongcang Gu, Phillip D. Zamore, Alexander Meissner, Zhiping Weng, Hans A. Hofmann, Nir Friedman, and Oliver J. Rando

*Contributions are listed at the beginning of Chapter I in its own preface.

Chapter II is a manuscript in preparation with myself as the first author:

Genetic and epigenetic variation in murine sperm contributes to offspring metabolism

Jeremy M. Shea, Benjamin R. Carone, Hennady P. Shulha, Michael Ziller, Alper Kucukural, Marcus Vallaster, Hongcang Gu, Andrew R. Tapper, Paul D. Gardner, Alexander Meissner, Manuel Garber, and Oliver J. Rando

*Contributions are listed at the beginning of Chapter II in its own preface.

Appendix I is preliminary data with a figure produced by myself that is included in the manuscript:

Paternal diet alters tRNA fragment levels throughout the male reproductive tract in mammals

Upasna Sharma, Colin C. Conine, Jeremy M. Shea, Benjamin R. Carone, Clemence Belleannee, Xin Li, Xin Y. Bing, Lucas Fauquier, Weifeng Gu, Philip D. Zamore, Robert Sullivan, Craig C. Mello, Manuel Garber, and Oliver J. Rando

INTRODUCTION

During development in multicellular organisms, transcriptional networks drive cell fate decisions to generate adult organisms with numerous cell types carrying out specific functions. Cell state inheritance is a classic case of epigenetic inheritance in which cells with the same genome nonetheless exhibit differences in heritable phenotypes. Moreover, genetic analysis of complex diseases often uncovers "missing heritability" (Manolio et al., 2009) in which all known genetic contributors to a given phenotype only explain a small fraction of heritability of that phenotype. Accumulating evidence suggests that ancestral epigenetic information influences offspring phenotypes, possibly accounting for this missing heritability (Danchin et al., 2011). Therefore, mature phenotypes arise from genetic interactions with the environment during development - as well as the previously ignored inherited epigenetic information. This revelation leads to important questions about the extent to which ancestral environments influence inherited phenotypes.

Epigenetic inheritance

Epigenetics plays an integral role in the development of multicellular organisms. In metazoa, hundreds of cell types result from the same genome and cellular identity is dictated by the specific subset of genes that are activated or repressed. Epigenetics provides a framework for cells to maintain their identity

through division by establishing a memory of genomic activity. For instance, although liver and kidney cells share the same genome, when a liver cell divides, it produces two liver cells and no kidney cells. The liver-specific gene expression program is recapitulated perfectly in both daughter cells. Thus, the liver phenotype is an example of epigenetic information, and genetic studies in multiple organisms show that epigenetic inheritance is essential for maintenance of cellular identity. Several overlapping mechanisms regulate such epigenetic information, including transcription factors, DNA methylation, and chromatin modifications. In addition, recent work has revealed higher order epigenetic regulation in the form of nuclear architecture that has been proposed to help coordinate gene expression. The integration of these epigenetic mechanisms ensures that cellular identity is established and maintained throughout development.

The essential role for epigenetic regulators in cell state inheritance is revealed by genetic analysis. For instance, classic genetic studies of development in *Drosophila melanogaster* identified the competing Trithorax group (TrxG) and Polycomb group (PcG) mutants, which are required for maintenance of active or repressed gene expression, respectively (Schuettengruber et al., 2007). PcG and TrxG genes maintain expression of important development regulators after the initiating signal is removed, and mutations in these genes result in malformations caused by the inability to propagate cellular identity after initiating differentiation.

PcG maintains repression by methylating Histone 3 at Lysine 27 (H3K27), while TrxG maintains gene activity by methylating Histone 3 at Lysine 4 (H3K4). These chromatin marks establish transcriptional memory at genes important for cell state identity.

In mammals, interfering with DNA methylation in fibroblast cultures causes them to differentiate into cardiomyocytes - the same transition induced by cardiomyocyte-specific transcription factors (TFs)(Ieda et al., 2010; Qian et al., 2012). In this case, an epigenetic information carrier provides a barrier against changing cellular identity. The generation of induced-pluripotent stem cells (iPS) from differentiated cells provides another example of epigenetic influence over cell fate. Inhibiting either DNA methylation or histone deacetylation (a chromatin modification) during generation of iPS cells increases the efficiency of reprogramming (Mikkelsen et al., 2008). Thus, DNA methylation and histone deacetylation provide barriers to cellular reprogramming and "lock in" specific cellular identities once they are established.

Aberrant epigenetic lesions highlight the roles of epigenetics in protecting and informing cellular identity, which can also be further appreciated from the extensive epigenetic deregulation in cancers (Hansen et al., 2011). Cancer arises from tissues through multistep processes resulting in loss of their programmed identity in their tissue of origin. Many epigenetic processes are

disturbed in cancer, including the deregulation of DNA methylation and PcG activity. Interestingly, a number of cancer therapeutics target epigenetic processes, including the DNA methylation inhibitor 5-azacytidine. Treatment of leukemias with 5-azacytidine leads to senescence of cancer cells, and reduction of their proliferative potential (Jasielec et al., 2014). Thus, cancer represents the loss of programmed cell identity, and is the reason behind therapeutics that target epigenetic processes.

In addition to the vital roles that epigenetics plays in development, epigenetic mechanisms are also "plastic", meaning that they respond to the environment. While the sequence of the genome remains constant throughout development, plasticity enables cells to respond to extracellular signals and environmental stresses without compromising cellular identity. For instance, chromatin packaging becomes altered in response to environmental conditions, as gene activity is generally well-correlated with a number of histone modifications (2012). Not only does chromatin state change in response to environmental cues, chromatin regulators are required for correct transcriptional responses to the environment (see for example (Weiner et al., 2012)). This plasticity ensures that cell types retain their cellular identity yet are flexible enough to respond to different environmental conditions.

Maintenance of cellular identity and plasticity are two complementary functions of epigenetics that operate during the lifetime of an organism – this is often referred to as mitotic epigenetic inheritance. An important question is whether epigenetic processes operate across generations. The ability of ancestral environments to influence phenotypes has been hotly debated for over a century, with the prevailing view for much of the 20th century being that organisms do not pass on environmental information to their offspring through the gametes. The germ plasm theory of August Weismann made a clear distinction between somatic cells and the germ line, which he thought was protected from somatic insults (Weismann et al., 1891). The development of the Modern Evolutionary Synthesis during the first half of the twentieth century left little room for epigenetic inheritance between generations, as this theory stated that phenotypes relied solely on inherited genotypes and developmental variations - offspring were not influenced by ancestral environments. These paradigms seemed to be confirmed by discoveries that extensive epigenetic reprogramming occurred after fertilization in mammals. However, in the past 10-20 years, clear examples of intergenerational epigenetic inheritance – the "inheritance of acquired characters" – have been described, necessitating a reevaluation of this dogma.

Epivariation

Some of the earliest confirmed cases of transgenerational epigenetic phenotypes were discovered in plants. One of the earliest and most famous examples of

transgenerational epigenetic inheritance is that of paramutation at the *r* locus in maize. The *r* locus encodes a transcription factor involved in pigment production, and Brink and co-workers discovered that genetically-identical maize could exhibit two distinct, heritable, color phenotypes – dark and light (Brink, 1956). Curiously, when a plant carrying a susceptible "paramutable" allele (*r*) was mated with a plant carrying the mutator allele (stippled), all offspring of this cross were found to have converted from *r* to *r'* (a dark pigment to a light pigment) even though the locus was genetically identical. The mutator allele thus alters the regulatory state of the susceptible allele in *trans*, heritably influencing the production of pigment for hundreds generations even though all alleles were genetically identical. The distinctive phenotype of the *r* locus facilitated its investigation, and epigenetic information regulators including small RNA pathways, cytosine methylation, and chromatin regulators have all been implicated in paramutation (Heard and Martienssen, 2014). More cases of epigenetic inheritance of pigmentation phenotypes have subsequently been found in plants (Chandler and Stam, 2004), revealing novel complexity in the hereditary information - epigenetic inheritance across generations increases phenotypic variation.

The plethora of phenomena in plants and absence in metazoa begged the question: "Are transgenerational phenomena restricted to the plant kingdom?" The conservation of transgenerational epigenetic inheritance (TEI) between

kingdoms was confirmed by studies of the agouti viable yellow locus (*Avy*) locus in mice (Morgan et al., 1999). The *Avy* locus results from the insertion of an intracisternal A particle (IAP) retrotransposon near the agouti gene, and genetically-identical *Avy* mice exhibit a range of coat colors from yellow to brown. Yellow *Avy* mice give birth to a higher fraction of yellow than brown offspring, and vice versa for brown mice, indicating that coat color phenotype in *Avy* mice is epigenetically heritable (albeit with incomplete penetrance). Thus, both the *Avy* and the *r* locus exhibit "metastable epialleles" – genes whose activity is ancestrally determined, but can vary between genetically identical organisms. The *r* locus in maize and the *Avy* allele in mice elucidate that vastly different heritable phenotypes result from the same genotype. Since these early discoveries, transgenerational phenomena have been found in most standard model organisms, including *C. elegans* (Fire et al., 1998; Greer et al., 2011), *D. melanogaster* (Cavalli and Paro, 1998), and *M. musculus* (Morgan et al., 1999), contributing to greater phenotypic complexity than would be expected if genotype were the sole conveyer of hereditary material.

Imprinting and "programmed" epigenetic information

While metastable epialleles demonstrate incomplete epigenetic inheritance between generations, in other cases epigenetic information can be stringently programmed. The most well known cases of programmed epigenetic marks are found in the study of imprinting, which is the allele-specific expression of genes

depending on the parent of origin (Reik and Walter, 2001). Seminal work into imprinting showed that the epigenetic carrier of transgenerational information is carried in *cis* on either the maternal or paternal allele of a given gene, with the epigenetic state of each allele being reset during gametogenesis in the next generation. The classic studies describing imprinting showed that the pronuclei transplanted between zygotes to produce gynogenetic (two maternal genomes) or androgenetic (two paternal genomes) embryos were unequal (Barton et al., 1984; Surani et al., 1984). The embryonic lethal gynogenetic and androgenetic embryos had two distinct phenotypes. Gynogenetic embryos had smaller placentas, while androgenetic embryos were smaller, but with normal placentas. These experiments revealed that the maternal and paternal epigenetic contributions to the embryo are unequal. Since these experiments only transferred the pronuclei between zygotes, they argue that the inherited epigenetic information causing these phenotypes is contained in the chromatin fraction.

Subsequent work on imprinting discovered the nature of this gametic inequality. The basis for the difference between maternal and paternal genomes is that epigenetic packaging over a number of genes does not reset after fertilization. Genetic lesions at the imprinted *Igf2r* cause embryonic lethality when maternally inherited, but offspring are viable when the mutation is paternally inherited (Barlow et al., 1991). These disparate phenotypes arise because the maternal

copy is the only one expressed during development as a result of inherited epigenetic information. The silent paternal copy of *Igf2r* is inherited and maintained in a silent state throughout development, so loss of the maternal copy results in virtual loss of all *Igf2r*. Parent-of-origin effects caused by imprinting are also present in human populations. The best-characterized case of imprinting in human populations is that of Prader-Willi and Angelman syndromes (Horsthemke and Wagstaff, 2008). These two syndromes are characterized by vastly different phenotypes, yet were found to be caused by the same exact deletion (chr15q11-13), with the constellation of symptoms exhibited by a patient lacking this chromosomal region dependent only on whether the deletion was maternally or paternally inherited.

Since the discovery of imprinting in the 1980's, approximately a hundred imprinted genes have been confirmed in both mice and humans. A number of these imprints are conserved between species, although a subset is imprinted in a species-specific manner. Imprinting has been proposed to arise from "genetic conflict" - the idea that in polygamous species, a father promotes the survival of his offspring, while a mother conserves resources to ensure her ability to mate with other males and have more offspring in the future (Moore and Haig, 1991). The lack of conservation for several imprinted genes indicates that because of changes in the nature of genetic conflict between species, programmed intergenerational information can be lost over evolutionary time.

Mechanisms of epigenetic inheritance

Genetic analyses of a number of epigenetic inheritance paradigms have identified several major pathways involved in epigenetic regulation. Classic epigenetics paradigms discovered by genetic analysis include cell state inheritance in flies, RNAi in worms, paramutation in maize, mating locus silencing in fission yeast, control of flowering time in Arabidopsis, imprinting in mammals, and many others. Genetic dissection of these and many other phenomena have identified chromatin packaging, small RNAs, and cytosine methylation as key regulators of epigenetic information. Importantly, it should be noted that these distinct mechanisms converge to coordinately regulate genomic processes. For instance, DNA methylation, heterochromatin, and the piRNA pathways all converge to silence transposable elements.

Below, I will focus on the extensively studied DNA methylation (Bourc'his et al., 2001; Mathieu et al., 2007), as analysis of this epigenetic mark formed the bulk of my thesis research. It is important to point out, however, than cytosine methylation is absent in many well-studied models such as *C. elegans* and *S. pombe*, which are nonetheless perfectly capable of epigenetic inheritance. Ongoing studies in the Rando laboratory focus on alternative epigenetic

information carriers, but these are beyond the scope of this project and so will be minimally discussed.

I will discuss the functions of cytosine methylation, followed by the mechanisms responsible for its genomic patterning. Subsequently, the roles it is known to play in intergenerational phenomena will be reviewed.

DNA Methylation

In addition to the four bases that comprise the majority of DNA, a fifth base – 5-methylcytosine – is present in diverse organisms including fungal species, mammals, plants, and many others. In mammals, DNA methylation regulates diverse processes including transcription, retrotransposon silencing, X-inactivation, and genomic stability (Goll and Bestor, 2005). The vast majority of cytosine methylation occurs in the context of the CpG dinucleotide, although a small amount of non-CpG methylation has been described in embryonic stem cells (Ziller et al., 2011). Cytosine methylation alters the topology of the major groove of DNA without affecting its charge, and has a variety of regulatory consequences including disrupting the binding of certain transcription factors, or recruiting heterochromatin complexes that contain methyl-CpG-binding domains (MBDs). The dynamic genomic patterning of DNA methylation along with its modulation of various DNA-associated activities explains its pleiotropic effects. Although DNA methylation occurs in many species, I will discuss topics relevant

to mammalian cytosine methylation. Many of the functions and mechanisms of DNA methylation are similar between species, and I will refer the reader to the review by Zemach and Zilberman for more in depth analysis of topics relevant to cytosine methylation evolution (Zemach and Zilberman, 2010).

Cytosine methylation regulates both differentiation and cell state identity, with cellular context determining the physiological role of cytosine methylation. Disruption of cytosine methylation in embryonic stem cells (ESC) blocks their differentiation to somatic lineages by preventing cytosine methylation-mediated silencing of pluripotent-specific factors (such as *Oct4*) (Feldman et al., 2006). Interestingly, disruption of cytosine methylation in ESCs doesn't affect their ability to propagate in any way (Tsumura et al., 2006). Thus, in the case of ESCs, cytosine methylation does not affect cellular identity, while it is necessary but not sufficient for their differentiation into other cell types. On the other hand, the ability of the DNA methylation inhibitor 5-azacytidine to differentiate murine embryonic fibroblasts (MEFs) into cardiomyocytes suggests that identity of differentiated cell types can be disrupted in the absence of cytosine methylation (although 5-azacytidine has targets other than DNA methylation) (Qian et al., 2012). Intriguingly, greater reduction of DNA methylation in MEFs leads to apoptosis (Jackson-Grusby et al., 2001). These opposite effects on cell lineage commitment and differentiation in ESCs and MEFs show that the roles of DNA methylation are contextual and cell type specific. These results demonstrate that

DNA methylation protects cellular identity in somatic cells, but that pluripotent cells rely less on DNA methylation and may need to remain in a plastic state for future lineage commitment. The phenotypes observed after inhibiting cytosine methylation in cell culture models hint at its crucial role in mammalian development.

Functions of DNA Methylation

As mentioned above, the developmental role of cytosine methylation derives in part from its ability to regulate gene expression by interfering with transcription. Many transcription factors (TFs) are sensitive to cytosine methylation, and exhibit reduced binding affinity for DNA sequences carrying a methylated CpG (Hu et al., 2013). Inhibition of TF binding by cytosine methylation can influence cellular phenotypes; as for example DNA binding by E2F TFs is modulated by methylation, resulting in altered proliferative phenotypes (Campanero et al., 2000). As sequence-specific DNA binding proteins have a wide range of regulatory effects on the genome, from transcriptional activation and repression through organization of chromatin domains, cytosine methylation can affect any of these processes. As a key example, CTCF is a non-activating genomic insulator that organizes the genome into distinct modules, and its DNA binding activity is disrupted by cytosine methylation in several instances. Cytosine methylation regulated binding of CTCF mediates enhancer-promoter choice at

13

the imprinted *H19/Igf2* locus (Hark et al., 2000) and trinucleotide-repeat instability at the spinocerebellar ataxia type 7 (*Sca7*) locus (Libby et al., 2008).

While cytosine methylation interrupts binding for several transcriptional regulators, much of its effect on transcription resides in recruitment of repressive methyl binding domain (MBD) containing proteins. Several MBD containing proteins reside in large heterochromatin complexes that reinforce genomic silencing (Klose and Bird, 2006). For instance, cytosine methylation recruits the MBD-containing MeCP2, which forms a complex with histone deacetylases (HDACs) and co-repressors, such as Sin3a (Nan et al., 1998). Furthermore, MBD-containing proteins recruit chromatin-remodeling complexes (such as NuRD) that stabilize repression (Baubec et al., 2013; Le Guezennec et al., 2006). In addition to MBD domains, some repressive zinc finger proteins (ZFPs) specifically recognize motifs containing methylated cytosines (Prokhortchouk et al., 2001). The modulation of TF binding and the recruitment of heterochromatin complexes by cytosine methylation leads to gene silencing, as well as non-canonical roles in maintenance of genomic architecture and genome stability.

Along with its role in transcriptional regulation of genes, cytosine methylation represses repetitive element (retroelement) activity. Repetitive elements comprise up to half of mammalian genomes, and are capable of destabilizing their host genome (Rowe and Trono, 2011). Redundant and highly specific

pathways, including cytosine methylation, coordinately repress transcription and mobilization of retroelements. In a fashion similar to transcriptional repression, cytosine methylation recruits silencing complexes, such as those formed by MeCP2 (Muotri et al., 2010), to inhibit retroelement activity. These multiple layers of repression are necessary for robust repetitive element silencing, as loss of cytosine methylation leads to activation of these elements and genome instability. The sensitivity of MEFs to the loss of cytosine methylation partially derives from activation of several classes of retroelements (Jackson-Grusby et al., 2001). Interestingly, perturbations of cytosine methylation machinery during gametic development lead to massive activation of retroelements that results in genomic instability, and ultimately apoptosis that prevents reproduction (Kato et al., 2007). These results show that epigenetic silencing of retroelements is an essential function of cytosine methylation.

Cytosine methylation is also integral to heritable control of gene dosage. The allele-specific expression of imprinted genes and the allele-specific silencing of the X-chromosome in females rely on cytosine methylation. In the case of imprinted genes, this epigenetic heritability extends from the parental generation to control gene dosage. The critical role of cytosine methylation in imprinting is observed in biallelic expression of imprinted genes at many stages of development in cytosine methylation mutants (Li et al., 1993). Unlike imprinted genes, random X-inactivation occurs after differentiation of the inner cell mass

(ICM), and the silenced copy is stably maintained throughout numerous cell divisions and development (Lee and Bartolomei, 2013). Disruption of cytosine methylation leads to activation of genes from the silenced X chromosome in differentiated cells (Mohandas et al., 1981). Interestingly, imprinted X-inactivation in marsupials occurs without cytosine methylation, and reactivation of the silenced X chromosome is a common phenomenon in marsupials, arguing that cytosine methylation contributes to long term silencing (Migeon et al., 1989). These examples of allelic regulation of transcription by cytosine methylation emphasize the role it plays in durable regulation of genetic activity.

Numerous other functions of cytosine methylation are beginning to be revealed, such as its role in the regulation of splicing. Increased cytosine methylation over exons regulates binding of CTCF, which alters the elongation kinetics of RNA polymerase II (Shukla et al., 2011). This change in elongation rate controls exon usage resulting in the expression of different isoforms of genes. Therefore, cytosine methylation not only affects gene expression at the level of transcription initiation, but is also a regulator of mRNA isoform usage.

Against the backdrop of *trans*-regulation of transcription by TFs and other DNA binding proteins, cytosine methylation adds a layer of *cis*-regulation that extends the complexity of genomic regulation. This *cis*-regulation modulates transcriptional activity, repetitive element silencing, gene dosage, as well as

other emerging activities. The intricate nature of this regulation makes cytosine methylation essential to mammalian survival.

Establishing and Maintaining Cytosine Methylation

Cytosine methylation in mammals is deposited by two *de novo* methyltransferases (DNMT3a and DNMT3b), which mediate the establishment of cytosine methylation, and maintained by DNMT1, which preferentially methylates the hemimethylated CpGs formed after replication of a symmetrically-methylated CpG. Recruitment of the *de novo* methyltransferases to DNA differs from that of the maintenance methyltransferase, each being recruited to DNA by different interaction partners. Active and passive processes also erase the patterns of cytosine methylation established and maintained by the methyltransferases during development. Therefore, genomic cytosine methylation patterns reflect the dynamic interaction between the antagonistic processes methylating and demethylating the genome.

After replication of a symmetrically methylated CpG, the newly synthesized daughter strand lacks DNA methylation while the complementary parental strand carries a methyl-cytosine on the opposite side of the CpG. This hemimethylated CpG is the preferred substrate for DNMT1, whose recruitment to hemi-methylated regions also requires the cooperation of PCNA (Proliferating Cell Nuclear Antigen) and UHRF1 (Ubiquitin-like containing PHD and RING finger

domains 1) (Sharif et al., 2007). PCNA ensures that DNMT1 is recruited to replicating DNA (Chuang et al., 1997), while UHRF1 targets it to hemi-methylated DNA. UHRF1 interacts with H3K9me2/3 (a heterochromatin mark) and hemimethylated DNA throughout replication and mitosis to precisely target DNMT1 to pre-existing cytosine methylation (Liu et al., 2013; Rothbart et al., 2012). UHRF1 recruits DNMT1 by ubiquitylating H3K23, a transient histone modification that is bound by DNMT1 during replication (Nishiyama et al., 2013). This elegant, multi-layered system ensures that cytosine methylation is faithfully reproduced during cell division.

In contrast to the maintenance of pre-existing methylation by DNMT1, DNMT3A and DNMT3B are responsible for establishing *de novo* cytosine methylation. These *de novo* methyltransferases associate with the catalytically inactive DNMT3L to form functional methylation complexes (Ooi et al., 2007). Since these complexes have little sequence specificity, recruitment depends on interaction with other DNA-associated factors. A variety of heterochromatin factors recruit *de novo* cytosine methylation, including Suv39H, G9a, and Setdb1-containing complexes (Epsztejn-Litman et al., 2008; Lehnertz et al., 2003; Li et al., 2006). It has also been found that the repressive histone modifications themselves recruit DNMTs, such as recruitment of DNMT3A by H4R3me2s (which is mediated by PRMT5) (Zhao et al., 2009). These modes of recruitment further accentuate the crosstalk between mechanisms responsible for silencing. Interestingly, the flavor

of heterochromatin determines the recruitment of cytosine methylation to a region. In contrast to genomic regions with H3K9me or H4R3me2s, regions with H3K27me (established by the Polycomb repressive complex (PRC)) are commonly devoid of cytosine methylation (Brinkman et al., 2012). Cytosine methylation encroaches into these regions when PRC function is compromised, a common feature of some cancers (Gal-Yam et al., 2008). PRC forms a competing complex with DNMT3L to prevent cytosine methylation at PRC bound regions, thus explaining the encroachment of cytosine methylation when PRC is disrupted (Neri et al., 2013). The ability of cytosine methylation to distinguish heterochromatic states is interesting because H3K27me is usually more labile, often being present in bivalent domains with the activating H3K4me modification, while H3K9me3-marked territories transmit long term silencing.

Interplay Between DNA Methylation and Chromatin

The interplay between chromatin and cytosine methylation explains the majority of the genomic methylation patterns. Reconstitution experiments in yeast demonstrate the role for the histone tail in establishing DNA methylation, as yeast lacking the H3 tail fail to establish cytosine methylation (Hu et al., 2009). Therefore, regions devoid of nucleosomes, commonly found in active promoters and enhancers, lack cytosine methylation (Chodavarapu et al., 2010). Nucleosome-depleted regions form through intrinsic nucleosome-deterring DNA sequences, as well as nucleosome displacement by a number of DNA interacting

factors, including TFs and chromatin remodelers (Feldmann et al., 2013; Kaplan et al., 2009). Therefore, many active regulatory regions are devoid of DNA methylation, since transcription factor binding commonly displaces nucleosomes.

The interplay between cytosine methylation and chromatin extends beyond the recruitment of cytosine methylation by heterochromatic processes. Several mechanisms reinforce the active state of regulatory regions by antagonizing heterochromatin formation. Multiple SET-containing histone methylases catalyze the methylation of H3K4 at transcriptionally active regions. These methylases are recruited to their targets by CxxC domain, which binds unmethylated CpGs (Clouaire et al., 2012). In turn, H3K4me (all forms) interrupts binding of the *de novo* cytosine methyltransferases with the H3 tail (Ooi et al., 2007). Therefore, the H3K4me present at transcribed elements antagonizes aberrant DNA methylation, reaffirming the active transcriptional status. This maintenance of unmethylated states ensures that TFs bind their cognate sites to promote transcriptional competency. Hence, numerous antagonistic chromatin processes dynamically regulate cytosine methylation.

The targeting of cytosine methylation through heterochromatin, along with the lack of a coherent copying mechanism for H3K9me has led to the idea that heterochromatin is established by histone modifications, and cytosine methylation locks in this heterochromatic state for extended silencing. A recent

paper targeted HP1 to a single allele of the highly expressed *Oct4* in ESCs to determine the kinetics of heterochromatin establishment and maintenance (Hathaway et al., 2012). HP1 recruitment led to H3K9me3 spreading over time, and the later establishment of DNA methylation. Full repression occurred before DNA methylation was observed, indicating that DNA methylation does not establish heterochromatin. However, reactivation of the locus occurred in clones with low DNA methylation, whereas clones with high methylation maintained repression over time. Inhibition of DNA methylation by 5-azacytidine treatment also reversed the repression of the locus. These results indicate that heterochromatin can be established independently of DNA methylation, but that epigenetic memory is stabilized through DNA methylation. This same mechanism appears to be used for the silencing kinetics during X-inactivation (Lee and Bartolomei, 2013). Organisms lacking DNA methylation still form heterochromatin; however, the absence of DNA methylation leads to greater plasticity - and reactivation of silenced genes.

DNA Methylation in Development

Mammals utilize cytosine methylation to dynamically regulate their genome throughout development (Smith and Meissner, 2013). The DNA methylome is reset after fertilization, reaching global levels of ~20% in the ICM in comparison to average methylation in somatic cells of ~70%. From this global minimum, DNA methylation is reestablished in cell-type specific contexts and gives rise to

characteristic methylation profiles (Ziller et al., 2013). These cell type-specific patterns are formed during development by *de novo* establishment and are stable for numerous cell divisions. Over the past several years, research has revealed that DNA methylation is dynamic over many regulatory elements during development, with active demethylation adding a layer to this epigenetic regulation.

Two major phases of methylome reprogramming take place in mammals. The first reprogramming occurs immediately after fertilization, when the gametic methylomes are erased. This erasure primes the epigenome for subsequent development, as well as preventing inheritance of ancestral insults. After reaching minimal levels in the ICM, the *de novo* methylation machinery establishes cell type-specific methylation patterns. The second major phase of reprogramming erases the somatic methylome in a subset of cells destined to become germ cells. Germ cells then establish sex-specific methylation patterns essential for reproduction. These two major phases of reprogramming were used in arguments against the possibility of transgenerational inheritance, since loss of epigenetic information at these stages should prevent inheritance of ancestral memory. However, we now know that imprinting and epialleles utilize memory stored in cytosine methylation to transmit intergenerational and transgenerational

information, respectively [b]. Therefore, the mechanisms of establishing and maintaining specific methylation patterns are important for development, as well as inheritance of intergenerational epigenetic information.

The dynamic nature of DNA methylation patterns throughout development requires reversing cytosine methylation on occasion (Wu and Zhang, 2014). Although DNA methylation is a chemical mark that stores epigenetic memory, certain situations signal its erasure. For example, establishment of the pluripotent state and development of germ cells require activation of genes previously silenced by cytosine methylation. In many instances, cytosine methylation represses regulatory genes early in development only for them to be demethylated at a later stage (Borgel et al., 2010). Several mechanisms mediate this reversal of cytosine methylation. Transcriptional down-regulation of the DNA methyltransferases triggers the loss of methylation in pre-implantation embryos and primordial germ cells (Messerschmidt et al., 2014; Seisenberger et al., 2012). Along with transcriptional control, the oocyte-specific isoform of DNMT1 (DNMT1o) localizes to the plasma membrane during early pre-implantation development - spatially restricting its access to DNA (Hirasawa et al., 2008). This passive demethylation relies on the replicative dilution of cytosine methylation to reset the methylome.

---

[b] Transgenerational refers to epigenetic information that survives both reprogramming stages (i.e.-epialleles), while intergenerational is the survival of information after fertilization that becomes reset in the germ line (i.e.-imprinted alleles).

Whereas passive demethylation requires replication, active demethylation occurs independently of cell division. Rapid demethylation occurs asymmetrically on the paternal pronuclei in the mammalian zygote (Santos et al., 2002), indicating that active demethylation takes place during this time of massive reprogramming, as well as at more limited regions throughout development. Active demethylation requires enzymatic removal of methylated cytosine. The TET enzymes catalyze oxidation of 5-methylcytosine, which is then deaminated by AID or Apobec to form 5-hydroxymethyluridine (5hmU) (Bhutani et al., 2010). TDG (Thymine DNA glycosylase) recognizes 5hmU in a 5hmU:G mismatch to initiate the base-excision repair (BER) machinery (Cortellino et al., 2011). The BER pathway ultimately replaces the oxidized cytosine with an unmodified cytosine, thus reversing methylation through an indirect process. Alternatively, TET mediated oxidation of cytosine reduces DNMT1 activity, leading to passive replication coupled dilution of hydroxymethylation (Hashimoto et al., 2012). Consequently, oxidation of 5-methylcytosine by TETs leads to demethylation by both base-excision repair and passive dilution of DNA methylation. Therefore, several processes mediate the erasure of cytosine methylation that is essential for reprogramming during developmental transitions.

While redundant mechanisms reset the methylome to a totipotent state after fertilization, imprints survive this reprogramming by actively protecting of cytosine

methylation. TET3 initiates active demethylation of a majority of the genome after fertilization, but *PGC7/Stella* protects selected regions from this process (Nakamura et al., 2012). In another example of crosstalk between cytosine methylation and chromatin, H3K9me2 recruits PGC7 to inhibit the enzymatic activity of TET3 (Bian and Yu, 2014). Although sperm have few nucleosomes when compared to somatic cells (~5%), a population of their residual nucleosomes has H3K9me2, including at imprinted genes. Although this mechanism protects imprints and other select loci from active demethylation, a noncanonical mechanism maintains cytosine methylation that is present in H3K9me2 occupied regions during early development. Site-specific recruitment of DNMTs to methylated regions by ZFP57 accomplishes this task (Quenneville et al., 2011). ZFP57 (a KRAB-containing zinc finger protein (ZFP)) recruits KAP1, Setdb1, and DNMTs to a methylated DNA motif to maintain cytosine methylation. Therefore, inhibition of active demethylation and site-specific recruitment of heterochromatin maintain cytosine methylation over imprints through early development. Methylation maintenance during development later transitions to canonical replication-coupled propagation mechanisms.

Epialleles represent the other class of well-studied transgenerational phenomena in mice that are not reprogrammed after fertilization. Intracisternal A Particle (IAP) retroelements regulate the activity of both the A^{vy} and *Axin*^{fused} epialleles in mice (Morgan et al., 1999; Rakyan et al., 2003). The A^{vy} epiallele contains an IAP

element that creates an alternative transcription start site, while the IAP in the *Axin^{fused}* epiallele drives expression of a truncated form of *Axin*. Metastable cytosine methylation controls the IAP elements of these epialleles. The actively inherited allele is unmethylated, while the repressed allele is highly methylated. Maintenance of DNA methylation over these epialleles through early development also relies on recruitment by ZFPs - as loss of KAP1 leads to demethylation of these elements in ESCs (Maksakova et al., 2013). Unlike imprinted genes that become reset in the germ lineage, IAP elements maintain cytosine methylation throughout this period (Seisenberger et al., 2012) - providing a possible mechanism for true transgenerational inheritance.

Plasticity of Inherited Epigenetic Information

The plasticity of inherited epigenomes is important for offspring phenotypic variation. If parents can inform their offspring of environmental conditions, these offspring could be better suited to the prevailing environment - a stress response that connects generations. As laid out above, mechanisms transmit information between generations, and these mechanisms rely on factors that are altered between individuals (ie-A^{vy}) to produce a spectrum of phenotypes. Metastable epialleles reveal that transgenerational plasticity may be prevalent - but what environments induce transgenerational responses? If environmental variables influence offspring phenotypes, this would be important in determining

phenotypic variability, which was an idea championed by Jean Baptiste Lamarck (Jablonka and Lamb, 1995).

The discovery of environmentally induced transgenerational phenotypes has traditionally been elusive, but numerous genuine examples have accumulated through the years. An early, but still investigated case relates to the inherited effects of heat shock in *Drosophila melanogaster*. Heat shock induces veinless phenotypes in flies, and more extreme heat stress results in heritable veinless phenotypes that persist for several generations (Waddington, 1953). Recently, it was shown that heritable responses to heat shock are mediated by ATF-2, which establishes and maintains heterochromatin (Seong et al., 2011). Heat stress signals through the Mekk1-p38 pathways to destabilize ATF-2, leading to a heritable release of silencing. Therefore, environmental disruption of heterochromatin by heat shock can be inherited in *D. melanogaster*.

Numerous organisms respond to pervasive temperature variations, as temperature also mediates transgenerational responses in plants and worms. Many plants undergo a process of vernalization, whereby seeds enter a dormant state in cold weather that primes them for subsequent warming. This process involves cytosine methylation of FLC, and disruption of methylation leads to early flowering in subsequent wild type generations (Finnegan et al., 1998). Heat stress also affects heritable transcriptional gene silencing in plants, another

example of heat disrupting inheritance of heterochromatin (Zhong et al., 2013). Thus, like in flies, heat shock disrupts heritable heterochromatin in plants. On the other hand, *C. elegans* show fertility defects in response to heat shock, and inhibiting the RNAi pathway in *C.elegans* exacerbates these transgenerational fertility defects (Conine et al., 2013). The difference between utilization of silencing mechanisms between all of these species leads to the specific responses of each organism to temperature shifts and the dynamics of those shifts.

Examples from *D. melanogaster*, *A. thaliana*, and *C. elegans* point to mechanisms that actively prevent transgenerational inheritance of stressed states. However, with enough stress these barriers are eventually overridden, and environmentally induced epigenetic information seeps into the next generation. Greater plasticity in subsequent generations is a result of the disruption of these mechanisms. For instance, plants lacking maintenance cytosine methylation have extreme phenotypic variation, a further indication that heterochromatic processes prevent intergenerational epigenetic inheritance (Mathieu et al., 2007).

Some organisms, especially in the plant kingdom, have vastly different reproductive strategies than mammalian systems, which make it easy for them to inherit ancestral epigenetic information. Mammalian reprogramming during

gametogenesis and post-fertilization prevent most environmental information from crossing between generations, leading to repression of intergenerational epigenetic plasticity. However, this barrier to epigenetic inheritance is not complete, as epialleles demonstrate that a subset of the mammalian genome exhibits transgenerational plasticity. This plasticity results from epigenetic mechanisms responding to different signals at these loci.

One of these signals is nutrition, as the A^{vy} and $Axin^{fused}$ epialleles in mice are responsive to dietary supplementation in parents. For instance, long-term methyl donor supplementation (such as folate) leads to greater cytosine methylation of the A^{vy} and $Axin^{fused}$ loci, lowering their heritable expression levels (Waterland and Jirtle, 2003a). Methyl donors feed into one carbon metabolism, the pathway responsible for cytosine and protein methylation by S-adenosylmethionine (Kaelin and McKnight, 2013), and common dietary supplements (such as folate, the B vitamins, or methionine) alter the balance of one carbon metabolism. This affects cytosine methylation at labile regions like epialleles, making these regions environmentally responsive. Since epialleles aren't reprogrammed during gametogenesis or early development, these cytosine methylation changes provide a means to transmit transgenerational information. Additionally, although massive reprogramming after fertilization erases most DNA methylation, studies have indicated that up to 20% of methylation is heritable across generations (McRae et al., 2014). These refractory regions, enriched for certain classes of

repetitive elements (such as IAPs) (Smith et al., 2012), might be carriers of inherited epigenetic information.

Epialleles present an opportunity to understand the mechanisms that are environmentally responsive in mammals. Interestingly, the genetic background of a mouse determines whether an epiallele can be paternally inherited, a strong indication that genetic differences within a species influence the prevalence of transgenerational inheritance (Rakyan et al., 2003). Furthermore, an ongoing mutation screen in the Whitelaw laboratory is searching for genetic modifiers of transgenerational epigenetic inheritance, already producing 20 modifiers of epigenetic inheritance (Daxinger et al., 2013). The common thread among many of these factors is their involvement in the formation of heterochromatin, including the DNA methylation machinery (DNMT1, UHRF1), the H3K9 methylation machinery (Setdb1, KAP1), and chromatin remodelers (Smarca5). As much crosstalk exists between these factors, coordination of their activity could influence epigenetic inheritance. Also, the activity of several of these factors has been demonstrated to be responsive to behavioral interventions, including prescribed exercise regimes that alter global cytosine methylation patterns (Ronn et al., 2013). Relatedly, several common dietary supplements have considerable effects on cytosine methylation, including ascorbic acid (Blaschke et al., 2013) and folic acid (Waterland and Jirtle, 2003a). These studies emphasize that heterochromatin is the major carrier of intergenerational epigenetic information,

and factors controlling heterochromatin are environmentally responsive. However, research needs to be undertaken to determine the responsiveness of individual factors to additional environmental variations.

Similar to heat shock in several organisms, mammals share some transgenerational environmental responses with other species. For instance, radiation leads to a heritable increase in mutation rates in both plants and animals (Molinier et al., 2006) (Barber et al., 2002). This response alters the genetic material, which disqualifies it from being strictly epigenetic; but the regulation of the pathway appears to be epigenetic, as the response is lost over several generations. This shared response is interesting because it indicates that epigenetic mechanisms communicate with the genetics of an organism, possibly increasing genetic variation in response to severe stresses.

The transgenerational responses to heat shock, dietary supplementation, and environmental radiation demonstrate that a wide array of signals can elicit transgenerational responses. The set of ancestral environments that affect phenotypes is an important question. Some relatively new environmental variations may have arisen through industrialization, and the production of environmental chemicals. A number of industrially produced chemicals elicit cellular responses through known signaling mechanisms. For instance, certain chemicals act as endocrine disruptors, altering hormone responses in mammals

by blocking or activating specific receptors. One of these endocrine disrupters, vinclozolin (an agronomically prevalent fungicide), leads to paternally inherited spermatogenic deficits in rats (Anway et al., 2005). This response lasts for up to three generations, making it especially deleterious. These deficits are also associated with differences in DNA methylation in the sperm of these animals, demonstrating that the sperm methylome is responsive to environmental toxicants. Thereby, the disruption of cellular processes by industrial chemicals can have profound impacts for generations to come, and adds to the environmental conditions that influence offspring phenotypes.

Many environmental variables elicit transgenerational responses in species-specific manners. What makes an environmental stress likely to influence subsequent generations? The intensity and duration appears to play central roles in this determination, as harsher treatments carry more of a transgenerational effect (ie- increased heat shock or radiation). Additionally, there appears to be periods of greater susceptibility in the lifetime of organisms as well. Vinclozolin shows larger effects when given to pregnant females earlier in pregnancy, when germ cells are reprogrammed, indicating that the mammalian germline might be especially susceptible to altered *in utero* conditions. These findings in model systems may prove pertinent to the etiology of human health and disease. As evidenced by epidemiological studies, the conservation of intergenerational

epigenetic inheritance indicates that a portion of phenotypic variation in humans might be caused by ancestral conditions.

These epidemiological studies show that transgenerational epigenetic plasticity plays an important, yet understated role in human health and disease. One of these studies focused on the parish of Overkhalix, Sweden; which kept detailed records on regional harvests, as well as mortality schedules. The Overkhalix cohort reveals that nutritional conditions of grandparents influence the health of the same sex grandchild (Pembrey et al., 2006). The Overkhalix cohort also shows that the time-period of the insult strongly affects the insult - going so far as to indicate that the same insult can have opposite effects on inherited phenotypes depending on when the ancestral generation encountered it. Thus, complex dynamics in the parental generation regulate the susceptibility of offspring to transgenerational phenotypes. The Overkhalix cohort and studies of nutritional supplementation of epiallelic mice demonstrate that nutritional status affects transgenerational phenomena. In the future, it will be necessary to determine what specific nutrients elicit responses, as well as whether organisms transmit information about general nutritional status.

All the evidence listed above evidence strongly suggests that chromatin-based processes control transgenerational inheritance, with the regulation of heterochromatin being the strongest indicator of transgenerational phenomena.

Heterochromatin machinery differs widely between species; a number of species with strong environmentally induced transgenerational phenotypes lack DNA methylation (or have minimal levels), and inhibition of DNA methylation in plants increases the prevalence of transgenerational variation. Therefore, it appears that DNA methylation promotes inherited phenotypic robustness, since species with genomic hypermethylation show reduced responsiveness to environmental variation.

Although the field of transgenerational epigenetics is still in its infancy, we see that transgenerational epigenetic mechanisms strongly influence offspring phenotypes. Accumulating evidence suggests that most multicellular sexually reproducing organisms exhibit transgenerational epigenetic inheritance, and that a subset of these states is environmentally responsive. On one hand, hardwired intergenerational epigenetic phenomena, such as imprinting, prove necessary for organismal survival. On the other hand, the ability of the transgenerational "epiphenotype" to respond to environments provides added phenotypic variability, possibly conferring fitness benefits on offspring. Characterization of the inducing environments with organismal responses will go a long way in determining the extent of intergenerational epigenetics, along with providing the framework to elucidate the mechanisms responsible for the inheritance of epigenetic information. Finally, a better understanding of intergenerational epigenetics will help solve the mysteries of complex disease.

CHAPTER I:

PREFACE

Chapter I is reprinted from the following co-authored work:

Paternally induced transgenerational environmental reprogramming of metabolic gene expression in mammals

Benjamin R. Carone*, Lucas Fauquier*, Naomi Habib*, Jeremy X. Shea*, Caroline E. Hart, Ruowang Li, Christoph Bock, Chengjian Li, Hongcang Gu, Phillip D. Zamore, Alexander X. Meissner, Zhiping Weng, Hans A. Hofmann, Nir Friedman, and Oliver J. Rando
Cell 2010 Dec 23; 143(7): 1084-96. doi: 10.1016/j.cell.2010.12.008.

Contributions:

Benjamin Carone, Jeremy Shea, and Oliver Rando performed gene expression microarray and qPCR experiments (Fig.1.1, S1.1, S1.2, and S1.3). Naomi Habib performed comparisons with public datasets (Fig.1.2 and 1.3B). Jeremy Shea processed lipomics mass spec samples (Fig.1.3C). Lucas Fauqier (along with Chengjian Li and Phillip Zamore) is responsible for miRNA measurements in Fig.1.4. Jeremy Shea performed MeDIP and MeDIP-seq (Fig.S1.4, S1.5, and 1.7). Jeremy Shea performed bisulfite sequencing (Fig.S1.4 and 1.6). Christoph Bock and Hongcang Gu are responsible for RRBS (Fig.1.5). Sperm preparations made and analyzed by Lucas Fauquier and Jeremy Shea (FigS1.6). Sperm chromatin and RNA analysis done by Benjamin Carone and Lucas Fauquier, respectively (FigS1.7). Bioinformatic analysis performed by Naomi Habib, Ruowang Li, Nir Friedman, Zhiping Weng, and Oliver Rando. Oliver Rando wrote the manuscript with editorial input from Phillip Zamore, Nir Friedman, and Naomi Habib; as well as Benjamin Carone, Lucas Fauquier, and Jeremy Shea.

CHAPTER I:

Paternally-induced transgenerational environmental reprogramming of metabolic gene expression in mammals

Abstract

Epigenetic information can be inherited through the mammalian germline, and represents a plausible transgenerational carrier of environmental information. To test whether transgenerational inheritance of environmental information occurs in mammals, we carried out an expression profiling screen for genes in mice that responded to paternal diet. Offspring of males fed a low protein diet exhibited elevated hepatic expression of many genes involved in lipid and cholesterol biosynthesis, and decreased levels of cholesterol esters, relative to the offspring of males fed a control diet. Epigenomic profiling of offspring livers revealed numerous modest (~20%) changes in cytosine methylation depending on paternal diet, including reproducible changes in methylation over a likely enhancer for the key lipid regulator PPARα. These results, in conjunction with recent human epidemiological data, indicate that parental diet can affect cholesterol and lipid metabolism in offspring, and define a model system to study environmental reprogramming of the heritable epigenome.

Introduction

The past few decades have seen an important expansion of our understanding of inheritance, as a wide variety of epigenetically-inherited traits have been described (Jablonka and Lamb, 1995; Rando and Verstrepen, 2007). One implication of epigenetic inheritance systems is that they provide a potential mechanism by which parents could transfer information to their offspring about the environment they experienced. In other words, mechanisms exist that could allow organisms to "inform" their progeny about prevailing environmental conditions. Under certain historical circumstances – for example, repeated exposure over evolutionary time to a moderately toxic environment that persists for tens of generations – such non-Mendelian information transfer could be adaptive (reviewed in (Jablonka and Lamb, 1995; Rando and Verstrepen, 2007)). Whether or not organisms can inherit characters induced by ancestral environments has far-reaching implications, and this type of inheritance has come to be called "Lamarckian" inheritance after the early evolutionary theorist J.B. Lamarck, although it is worth noting that both Darwin and Lamarck believed in the inheritance of acquired characters.

Despite these theoretical considerations, at present there is scant evidence for transgenerational effects of the environment, particularly in mammals. The majority of examples of transgenerational environmental effects described have been maternal effects (see (Harris and Seckl, 2011; Whitelaw and Whitelaw,

2008; Youngson and Whitelaw, 2008) for review), including in utero passage of photoperiod information in various rodents (Horton, 2005), cultural inheritance of stress reactivity and maternal grooming behavior in rats (Meaney et al., 2007; Weaver et al., 2004), and metabolic and psychiatric sequelae of fetal malnutrition in humans and rodents (Hales and Barker, 2001; Harris and Seckl, 2011; Symonds et al., 2009). However, maternal effects are difficult to separate from direct effects of in-utero environmental exposure on offspring.

A small number of studies have identified heritable epigenetic effects of environmental perturbations on offspring. Treatment of pregnant rat mothers with the endocrine disruptor vinclozolin results in decreased fertility and behavioral changes in several generations of offspring (Anway et al., 2005; Crews et al., 2007). In another study, withholding methyl donors from pregnant female mice resulted in decreased cytosine methylation across the agouti viable yellow *Avy* reporter locus (Waterland and Jirtle, 2003b), and the altered cytosine methylation profile persisted well beyond the first generation (Cropley et al., 2006).

While demonstration of multi-generational changes (eg- an F2 effect) is important when using maternal treatment protocols to rule out simple plastic responses of offspring to the in utero environment, paternal effects avoid this issue as fathers often contribute little more than sperm to offspring. A handful of paternal effects have been documented in the literature – pre-mating fasting of male mice has been reported to affect serum glucose levels in offspring (Anderson et al., 2006),

and chronic exposure of male rats to high fat diet affects pancreatic islet biology in offspring (Ng et al., 2010). Furthermore, epidemiological data from human populations links experience of famine in paternal grandfathers to obesity and cardiovascular disease two generations later (Kaati et al., 2002; Pembrey et al., 2006). These results motivate a deeper exploration of the mechanisms of pre-mating paternal diet on offspring phenotype.

It is therefore of great interest to determine what environmental conditions have transgenerational effects in mammals, and to characterize the mechanisms that mediate these effects. Here, we describe a genomic screen for transgenerational effects of paternal diet on gene expression in offspring in mice. Expression of hundreds of genes changes in the offspring of males fed a low protein diet, with coherent upregulation of lipid and cholesterol biosynthetic pathways. Epigenomic profiling in offspring livers identified changes in cytosine methylation at a putative enhancer for the key lipid transcription factor PPARα, and these changes correlated with the downregulation of this gene in offspring. Interestingly, we did not find effects of paternal diet on methylation of this locus in sperm, and overall sperm cytosine methylation patterns were largely conserved under various dietary regimes. These results establish an inbred, genetically tractable model system for the study of transgenerational effects of diet, and may have implications for the epidemiology of several major human diseases.

Experimental paradigm

Male mice were fed control or low protein diet (11% rather than 20% protein, with the remaining mass made up with sucrose) from weaning until sexual maturity. Note that while the relevant dietary change in this experiment could be protein content, sucrose content, fat/protein ratio, etc., for simplicity we refer to the diet as low protein throughout the text. Mice on either diet were then mated to females reared on control diet (Figure 1.1A, S1.1A). Fathers were removed after one or two days of mating, limiting their influence on their progeny to the mating itself. All mothers were maintained on control diet throughout the course of the experiment. After birth, the offspring were reared with their mothers until three weeks old, at which point their livers were harvested for RNA isolation. DNA microarrays were used to profile global gene expression differences in the livers of the offspring from the two types of crosses (Table S1.1).

A screen for genes regulated by paternal diet

Testing for differences between 26 matched pairs of mice from the two F1 groups, we found a significant overabundance of differentially-expressed genes, relative to the null hypothesis that the parental treatment does not affect offspring (1,595 genes at false discovery rate – FDR – of 0.001, Figure S1.1B–C). We also identified a more robust (t-test with null hypothesis of mean change 0.2, FDR of

A

Sex: ♂ ♀ ♂ ♀

Diet: Control Control Low protein Control

B

Offspring sex: ☐ Female ■ Male

C

Sqle/Vtn, fold change relative to wt females

64
32
16
8
4
2
1
0.5
0.25

p = 0.006   *p* = 0.005

Offspring sex: ♀ ♂

Paternal diet: C P C P

4-fold down in LP/C   4-fold up in LP/C
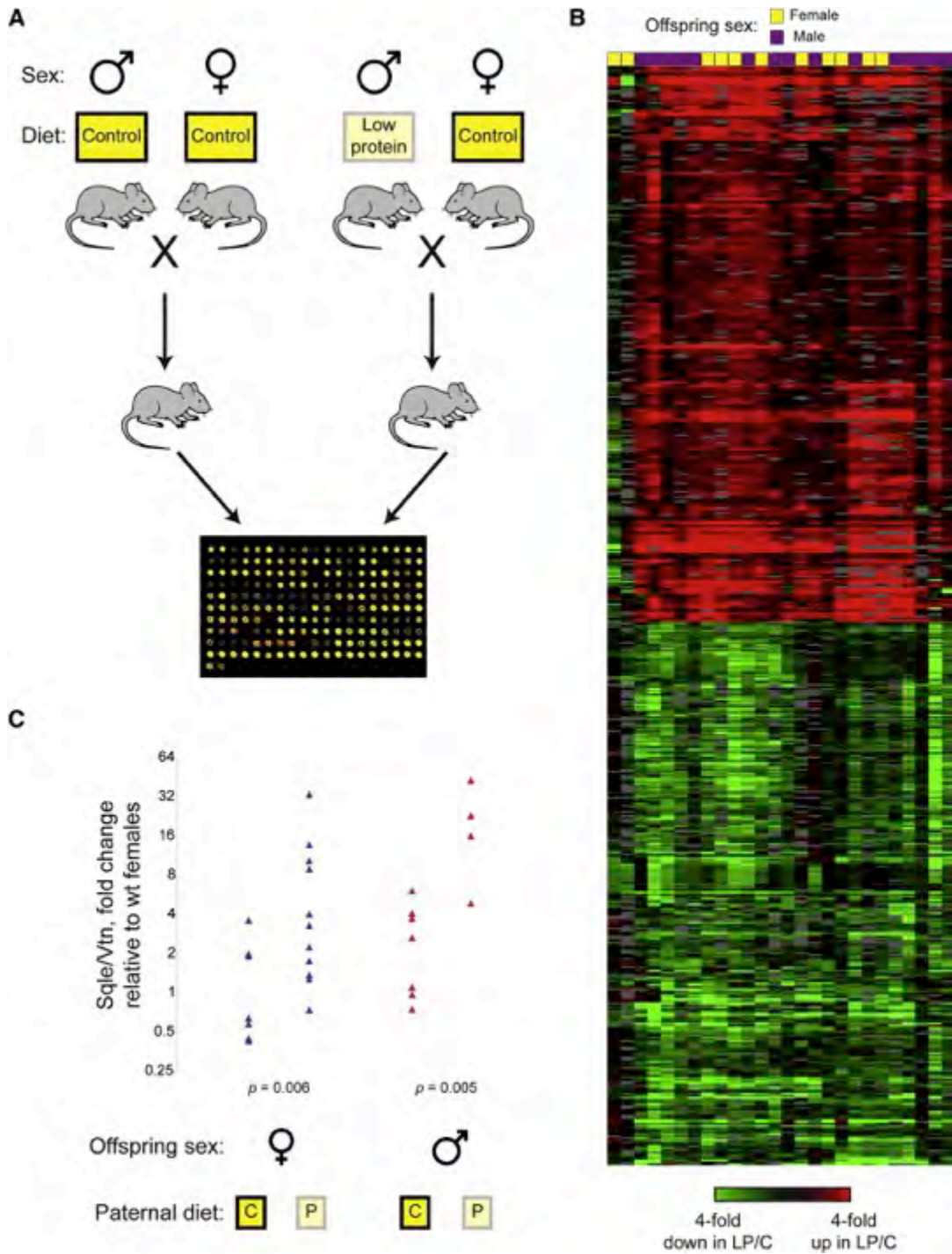
Figure 1.1

A Screen for Genes Regulated by Paternal Diet

(A) Experimental design. Male mice were fed control or low (11%) protein diet from weaning until sexual maturity, then were mated to females that were raised on control diet. Males were removed after 1 or 2 days of mating. Livers were harvested from offspring at 3 weeks, and RNA was prepared, labeled, and hybridized to oligonucleotide microarrays.

(B) Overview of microarray data, comparing offspring of sibling males fed different diets—red boxes indicate higher RNA levels in low-protein than control offspring, green indicates higher expression in controls. Boxes at the top indicate comparisons between two male (purple) or two female (yellow) offspring. Each column shows results from a comparison of a pair of offspring. Only genes passing the stringent threshold for significant change (Figure S1B) are shown. Data are clustered by experiment (columns) and by genes (rows).

(C) Validation of microarray data. Quantitative RT-PCR was used to determine levels of Squalene epoxidase (*Sqle*) relative to the control gene Vitronectin (*Vtn*), which showed no change in the microarray dataset. Animals are grouped by paternal diet and by sex, and data are expressed as ΔCT between *Sqle* and *Vtn*, normalized relative to the average of control females.
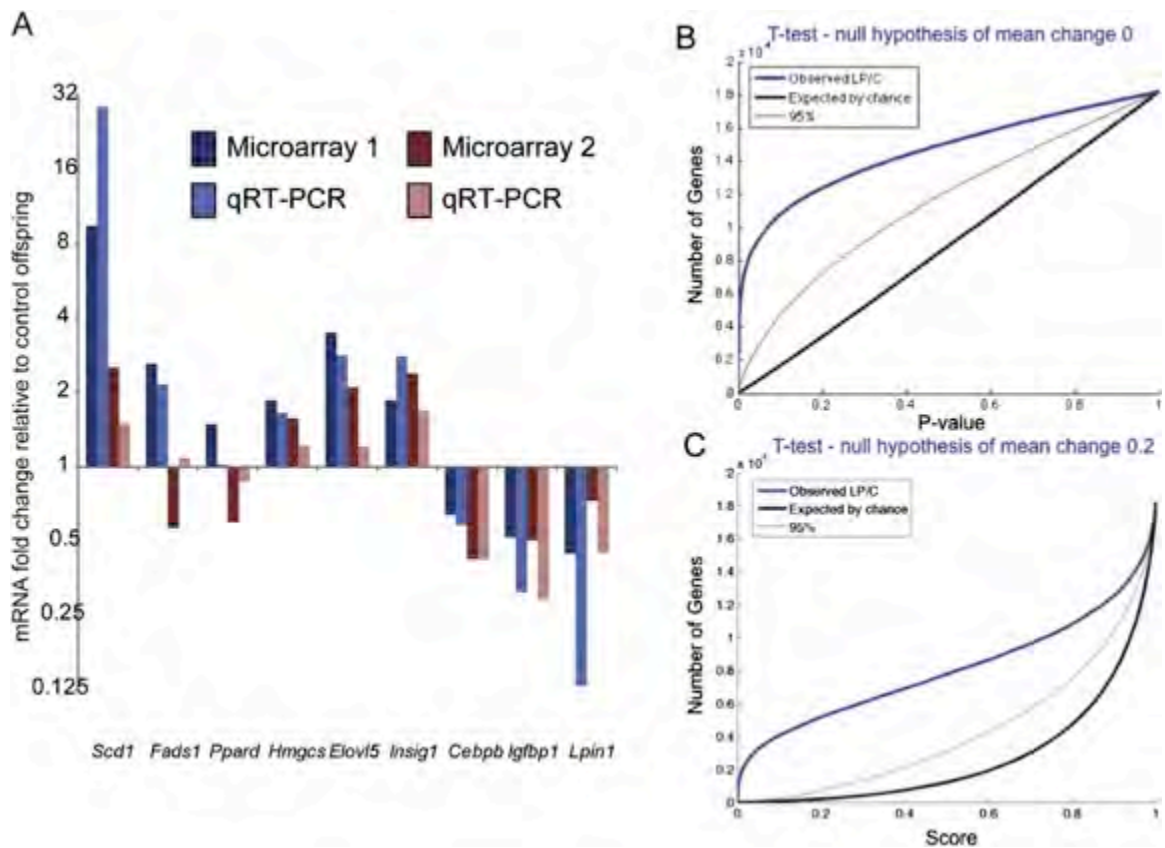
Figure S1.1

Validation and Identification of Differentially Expressed Genes, Related to Figure 1

(A) Microarray data and q-RT-PCR results are shown for the indicated genes, for two offspring comparisons.

(B and C) Evaluating the statistical significance of the number of genes that are differentially expressed between offspring of low-protein diet fathers and control diet fathers. Blue line, the number of differentially expressed genes that separate the two sets of offspring (y axis) that were scored a given p value (x axis) in a t test; black line, the number of genes expected by chance with that p value from 1000 simulations with random reshuffling of subject labels; light gray or red line, the range of numbers of differentially expressed genes in the 95th percentile of 1000 random simulations. Overabundance of differentially expressed genes is observed when using both tests: (B) combination of two one-tailed t tests; (C) combination of two one-tailed t tests using a null hypothesis with mean change of 0.2. In this case the random reshuffling of the data corresponds to a null hypothesis with mean 0 rather than 0.2 and thus is an upper bound on the number expectance by chance.

0.01) group of 445 genes whose expression strongly depended on the diet consumed by their fathers (Figure 1.1B). In our analysis we focus on this more robust group of genes, however, all the phenomena described below are true for the larger group as well. These gene expression changes were observed in 13 (7 low protein, 6 control) litters in experiments spanning several years, carried out in three different animal facilities (Figure S1.2A–B). In principle, random factors should be distributed equally between our two groups given the numbers of offspring examined, but we directly address a number of potential artifacts nonetheless, including changes in cell populations, circadian cycle, litter size, order of sacrifice, and cage location (Figure S1.2, see Experimental procedures).

We confirmed our results by q-RT-PCR (Figures 1.1C, S1.1A). Squalene epoxidase (*Sqle*), which catalyzes the first oxygenation step in sterol biosynthesis, exhibited a ~3-fold increase in the low protein cohort in our microarray data, and q-RT-PCR showed a similar average expression difference across over 25 animals, gathered in crosses carried out several years apart (Figure 1.1C). The differences we observe occur in both male and female progeny (Figures 1.1C, S1.2C), though these dietary history-dependent differences are superimposed on a baseline of differential expression between the sexes.

A

| Control | Low protein |
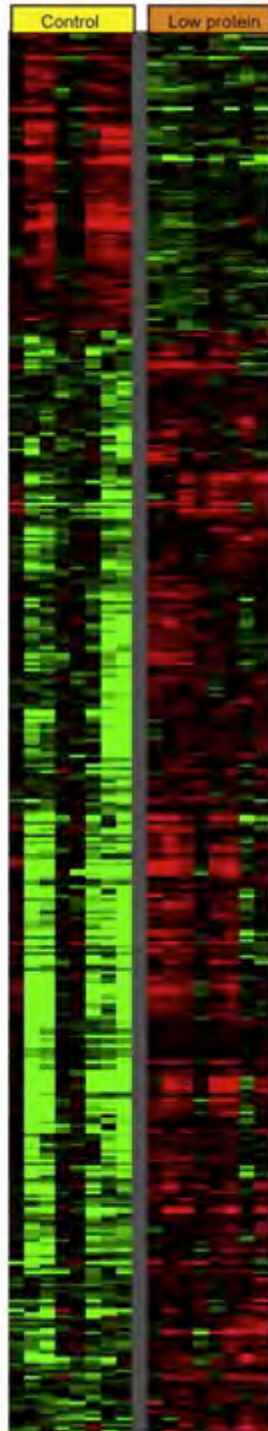
uncoupling protein 2, mitochondrial
ribose 5-phosphate isomerase A
phospholipid transfer protein
cyclin D1
adiponutrin
fibroblast growth factor receptor-like 1
riboflavin kinase
mitogen-activated protein kinase 6
transaldolase 1
acyl-CoA synthetase long-chain family member 3
lipin 2
fatty acid binding protein 5, epidermal
hydroxysteroid (17-beta) dehydrogenase 12
fatty acid binding protein 2, intestinal
2,3-bisphosphoglycerate mutase
cytochrome P450, family 2, subfamily r, polypeptide 1
ribulose-5-phosphate-3-epimerase
mevalonate (diphospho) decarboxylase
citrate synthase
acetyl-Coenzyme A carboxylase
phosphogluconate dehydrogenase
glucose-6-phosphate dehydrogenase X-linked
pyruvate dehydrogenase kinase, isoenzyme 1
aquaporin 8
lanosterol synthase
hydroxysteroid (17-beta) dehydrogenase 7
sterol-C5-desaturase
3-hydroxy-3-methylglutaryl-Coenzyme A synthase 1
acetyl-Coenzyme A acetyltransferase 2
acetyl-Coenzyme A synthetase 2 (ADP forming)
retinol dehydrogenase 11
phosphomevalonate kinase
NAD(P) dependent steroid dehydrogenase-like
7-dehydrocholesterol reductase
farnesyl diphosphate synthetase
isopentenyl-diphosphate delta isomerase
squalene epoxidase
fatty acid desaturase 2
fatty acid desaturase 1
diaphorase 1 (NADH)
acyl-CoA synthetase long-chain family member 5
ELOVL family member 5
nerve growth factor, alpha
pyruvate kinase liver and red blood cell
ELOVL family member 6
ATP citrate lyase
hydroxysteroid (17-beta) dehydrogenase 7
isopentenyl-diphosphate delta isomerase
low density lipoprotein receptor-related protein associated protein 1
galactose-4-epimerase, UDP
transketolase
citrate synthase
fatty acid synthase
malic enzyme, supernatant
glycerol phosphate dehydrogenase 2, mitochondrial
stearoyl-Coenzyme A desaturase 1
NAD(P)H dehydrogenase, quinone 2
3-hydroxy-3-methylglutaryl-Coenzyme A reductase
mal, T-cell differentiation protein 2
fibroblast growth factor 21
major urinary protein 3
glucokinase
peroxisomal membrane protein 4
insulin degrading enzyme
coxsackievirus and adenovirus receptor
NIMA (never in mitosis gene a)-related expressed kinase 6
solute carrier family 39 (zinc transporter), member 4
acyl-CoA synthetase long-chain family member 4
glucose phosphate isomerase 1
glutamyl aminopeptidase
stearoyl-Coenzyme A desaturase 2
MAP kinase-interacting serine/threonine kinase 2

B

| Facility 2 | Facility 3 |

C

| Males | Females |

D

Average gene expression,
low protein vs. control

Phase of peak expression

45
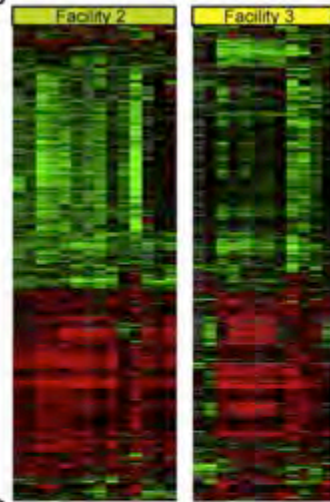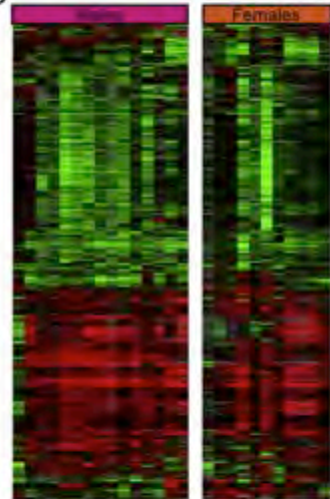
Figure S1.2.

Gene Expression Differences Repeat in Multiple Animal Facilities and Are Observed in Both Genders, Related to Figure 1.1

(A) Data from 16 animals (8 control, 8 low-protein offspring) from an early iteration of the experiment presented in Figure 1A. Data are from Affymetrix microarrays, with each column representing single-channel intensity data for one animal. Each gene is normalized to the median across all 16 experiments (i.e., zero-centered). These animals were raised in Facility 1, distinct from the animals presented in the rest of this paper.

(B and C) Data from Figure 1.1B, segregated by animal facility (B) or offspring gender (C).

(D) Gene expression differences do not reflect changes in circadian rhythms. Periodically expressed genes from (Miller et al., 2007) are ordered according to time of peak expression during the circadian cycle, and average change in gene expression in our dataset is plotted alongside.

Upregulation of proliferation and lipid biosynthesis genes in low protein offspring

To help define the physiological differences between our cohorts, we calculated enrichments of various Gene Ontology (GO) processes in the differentially expressed genes. Genes upregulated in our treatment group's offspring were enriched for a number of categories of genes involved in fat and cholesterol biosynthesis, including lipid biosynthesis ($p < 9 \times 10^{-26}$), steroid biosynthesis ($p < 3 \times 10^{-19}$), cholesterol biosynthesis ($p < 2 \times 10^{-12}$), and oxidation-reduction ($p < 4 \times 10^{-10}$). Another major group of upregulated genes are annotated to be involved in S phase, such as DNA replication ($p < 2 \times 10^{-9}$) and related annotations. Downregulated genes were enriched for GO annotations such as sequence specific DNA binding ($p < 6 \times 10^{-6}$) and ligand-dependent nuclear receptor activity ($p < 6 \times 10^{-5}$), although the number of genes matching these annotations was small (14 and 5, respectively).

The increase in S phase genes likely indicates a hyperproliferative state, while the metabolic expression differences suggest that lipid metabolism is altered in these animals. To explore the mechanisms responsible for these altered gene expression programs, we asked whether the observed gene expression differences might reflect altered regulation of a small number of pathways. We checked for significant overlaps of the gene expression profile observed in our low protein offspring with a compendium of 120 publicly available murine liver

gene expression datasets (Experimental Procedures). Our low protein offspring gene expression profile significantly ($p < .05$ after Bonferroni correction) overlapped gene expression changes from 28 published profiles (Figure 1.2, Table S2), including gene expression profiles associated with perturbation of transcription factors that regulate cholesterol and lipid metabolism (SREBP (Horton et al., 2003), KLF15 (Gray et al., 2007), PPARα (Rakhshandehroo et al., 2007), and ZFP90 (Yang et al., 2009)). Our gene expression dataset also significantly matched hepatic gene expression in a variety of mice with mutations affecting growth hormone (GH) and insulin-like growth factor 1 (IGF-1) levels (Boylston et al., 2004; Madsen et al., 2004; Tsuchiya et al., 2004). Hierarchical clustering according to the enriched public profiles revealed two types of prominent gene functions in our data: DNA replication ($p < 6 \times 10^{-14}$) and lipid or cholesterol biosynthesis ($p < 2 \times 10^{-27}$) (Figure 1.2). The partial overlap observed with each of many different transcription factor and growth factor profiles suggests that the altered gene expression profile observed in low protein offspring is likely related to reprogramming of multiple distinct pathways

Multiple pathways are affected by paternal diet

To assess whether the reprogrammed state in offspring reproduces the paternal response to low protein diet, we measured global gene expression changes in the livers of pairs of animals weaned to control or low protein diet as in Figure 1.1A. Genes that change in offspring are not the same as the genes induced in

48
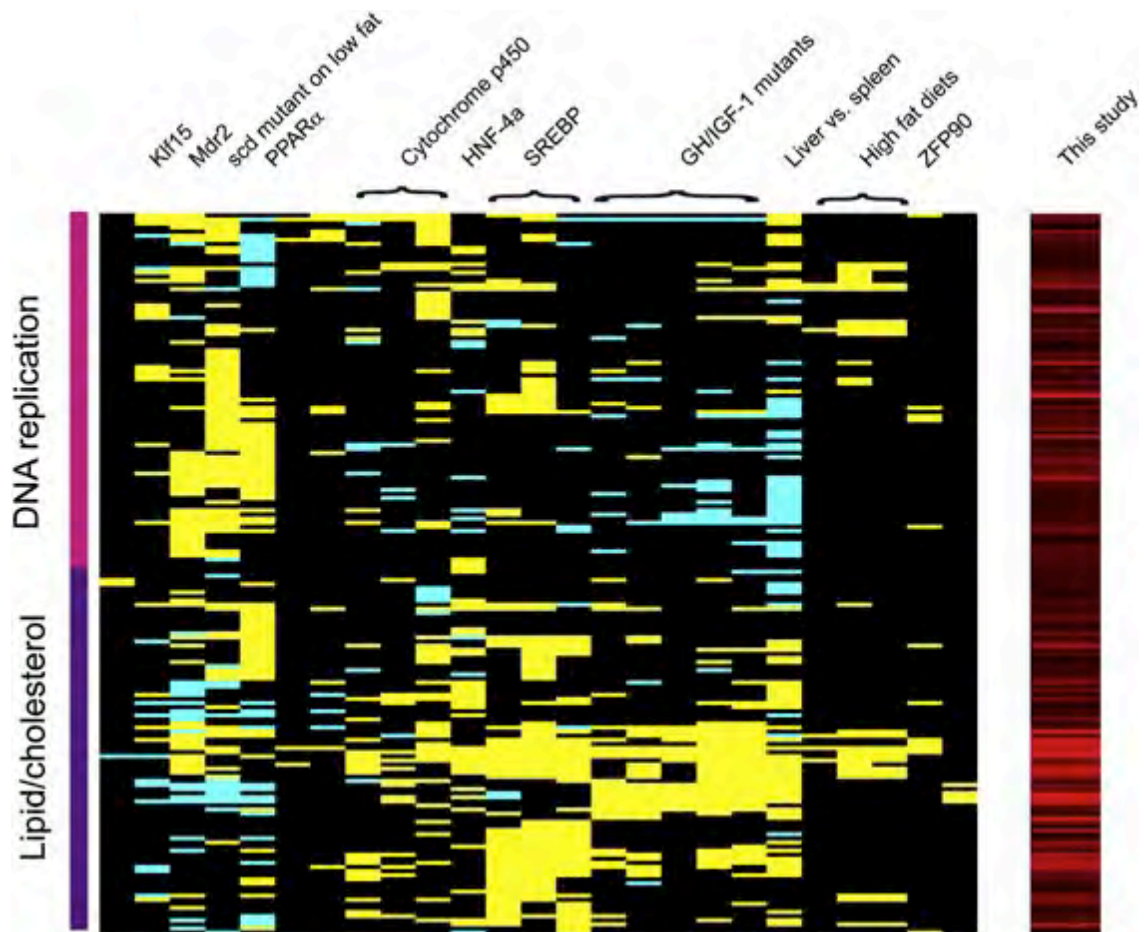
Figure 1.2

Multiple Pathways Are Affected by Paternal Diet

Comparison of upregulated gene expression profile with a compendium of public datasets of hepatic gene expression. A clustering of our upregulated genes according to their notation in the 28 significant (p < 0.00025) overlapping signatures from an assembled compendium of 120 publicly available murine liver signatures under various conditions and genetic perturbations (GEO; (Horton et al., 2003) and (Yang et al., 2009)). For each significant profile, the majority of overlapping genes are shown as yellow, whereas genes with opposite regulation (i.e., down rather than up in the dataset in question) are blue. The genes divide into two distinct clusters, one enriched in DNA replication and the other in various categories of fat and cholesterol biosynthesis. See also Table S1.2 and Figure S1.3.

the parental generation by these protocols (Figure S1.3). Instead, males fed the low protein diet upregulate immune response and apoptosis-related genes, and downregulate genes involved in carboxylic acid metabolism (analysis not shown).

Transgenerational effects on lipid metabolism

We further focused on cholesterol biosynthesis genes. Coherent upregulation of genes involved in cholesterol metabolism is observed in the offspring of low protein fathers (Figure 1.3A). Figure 1.3B shows a more detailed comparison between our upregulated dataset and published data (Horton et al., 2003) for genes activated by a major transcriptional regulator of cholesterol metabolism, SREBP. Many of the genes upregulated in low protein offspring have previously been shown to be upregulated by overexpression of SREBP-1a or SREBP-2 or downregulated by loss of the SREBP-activating gene, *Scap*.

Altered cholesterol metabolism in the low protein cohort

To explore the correspondence between hepatic gene expression and physiology, we measured lipid levels in three pairs of control and treatment livers to determine whether increased levels of lipid biosynthesis genes resulted from changes in lipid levels (Figure 1.3C, Experimental Procedures). Livers in the cohort with low protein diet fathers were depleted of cholesterol and cholesterol

A

Gene expression in LP vs
C-treated animals ("dads")

Gene expression in LP vs C offspring

B

Log2(LP/C) dads

Log2 (LP/C) offspring
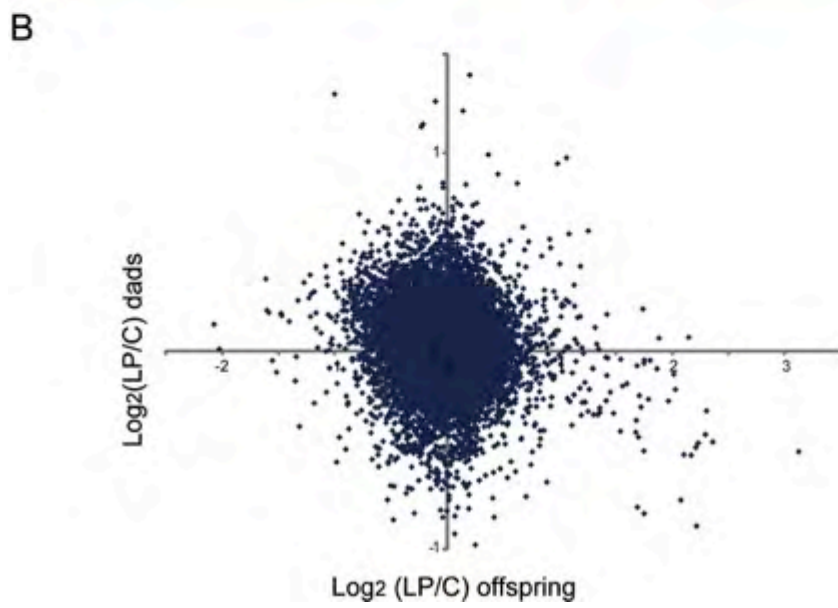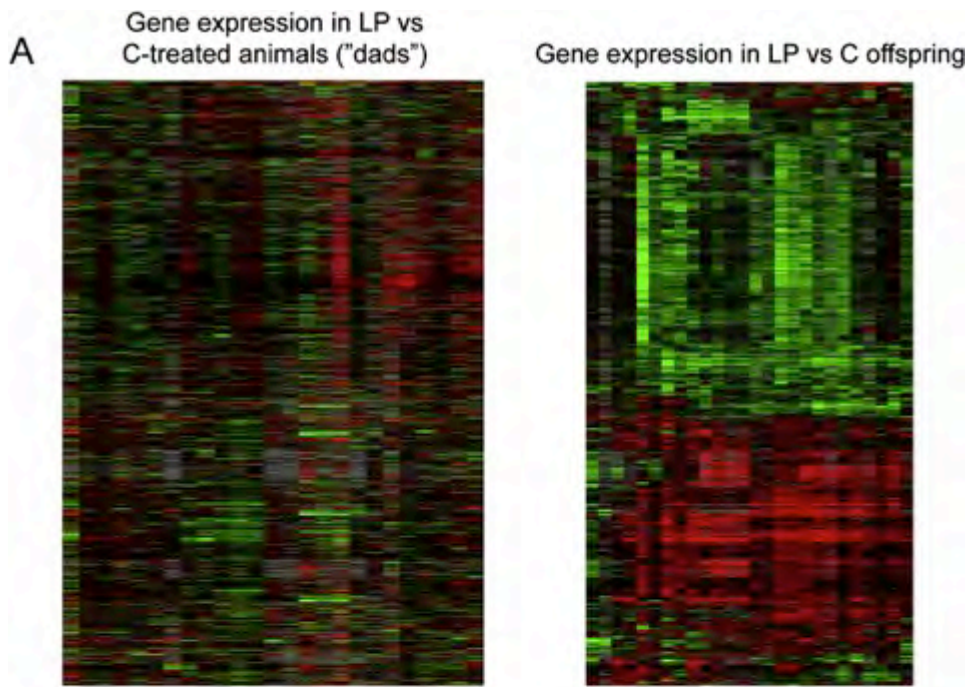
Figure S1.3.

Analysis of Paternal Response to Low-Protein Diet, Related to Figure 1.2

(A) Males were fed control or low-protein diet from weaning until sexual maturity, then were sacrificed and livers were harvested for gene expression profiling as in Figure 1.1. Here, genes are ordered as in Figure S1.2B (right panel)—left panel shows gene expression differences as low-protein/control. Gene expression differences in offspring do not reflect the paternal response to the dietary regimes (note that these males were not fathers of the offspring analyzed in Figure 1.1B, but were treated equivalently).

(B) Scatterplot of average gene expression in offspring (x axis) versus in males treated with LP or C diet (y axis). Only genes were chosen with fewer than 30% missing spots in each experiment (26 arrays each). R = −0.129.

A

3-hydroxy-3-methylglutaryl-CoA (HMG-CoA)
↓ Hmgcr
Mevalonate
↓ Mvk
Mevalonate-5P
↓ Pmvk
Mevalonate-5PP
↓ Mvd
Isopentenyl-PP
↓ Idi1
Geranyl-PP
↓ Fdps
Farnesyl-PP
↓ Fdft1
Squalene
↓ Sqle
(S)-Squalene-2,3-epoxide
↓ Lss
Lanosterol
↓ Cyp51
4,4-Dimethyl-cholesta-8,14,24-trienol
↓ Tm7sf2
14-Demethyl-lanosterol
↓ Sc4mol
4-Methylzymosterol-carboxylate
↓ Nsdhl
3-Keto-4-methyl-zymosterol
↓ Hsd17b7
4-Methylzymosterol
↓ Dhcr24
Lathosterol
↓ Sc5d
7-Dehydrocholesterol
↓ Dhcr7
Cholesterol

No change in LP/C — 4-fold up in LP/C

B

SREBP-1a   SREBP-2   scap-/-

C

Fold change low protein relative to control

$p = 0.01$ $p = 0.04$ $p = 0.05$

Lipid class: CE, PE, FC, TAG, PC, CL, PS, FA, LYPC, DAG

53

Figure 1.3.

Altered Cholesterol Metabolism in the Low-Protein Cohort

(A) Cholesterol biosynthesis. Genes annotated as cholesterol biosynthesis genes are shown, with colors indicating average difference in expression in low-protein versus control comparisons.

(B) Many genes upregulated in the low-protein cohort are SREBP targets. Upregulated cluster from Figure 1B is shown, along with data from Horton et al. (2003). Genes scored as up in both replicates from Horton et al. (2003) are shown as yellow, genes scored as down are blue. Columns show data from transgenic mice overexpressing SREBP-1a or SREBP-2 or from Scap knockout mice.

(C) Cholesterol levels are decreased in livers of low-protein offspring. Data from lipidomic profiling of liver tissue from three control and three low-protein animals are shown as mean ± standard deviation. Red line indicates no change. p values were calculated using a paired t test on log-transformed lipid abundance data. Cholesterol esters, CE; phosphatidylethanolamine, PE; free cholesterol, FC; triacylglycerol, TAG; phopshatidylcholine, PC; cardiolipin, CL; phosphatidylserine, PS ; free fatty acid, FA; lysophosphatidylcholine, LYPC; and diacylglycerol, DAG.

See also Table S1.3.

(whose levels were reduced more than two-fold). Additional differences were found in specific lipid classes, such as substantial increases in relative levels of saturated cardiolipins, saturated free fatty acids, and saturated and monounsaturated triacylglycerides in low protein offspring (Table S1.3). Together, these results demonstrate that paternal diet affects metabolites of key biomedical importance in offspring.

MicroRNAs in offspring

Small (19–35) RNAs such as microRNAs (miRNAs) have recently been implicated in epigenetic inheritance in mice (Wagner et al., 2008). To determine whether altered small RNA populations might drive our reprogramming effect, we characterized the small (19–35 bp) RNA population from control and low protein offspring livers by high throughput sequencing (Ghildiyal et al., 2008), and mapped reads to known microRNAs (Table S1.4). A number of miRNAs changed expression in the offspring from low protein diet fathers (Figure 1.4). Changes were often subtle in magnitude (~50%), but were reproduced in four control vs. low protein comparisons (paired t-test), and given the number of sequencing reads obtained for these RNAs this magnitude of difference is well outside of counting error (Table S1.4). Offspring of low protein cohort upregulated *miR-21*, *let-7*, *miR-199*, and *miR-98*, and downregulated *miR-210*. Many of these upregulated miRNAs are associated with proliferation in liver, with *miR-21* and *miR-199* both associated with hepatocellular carcinoma (Jiang et al., 2008), while
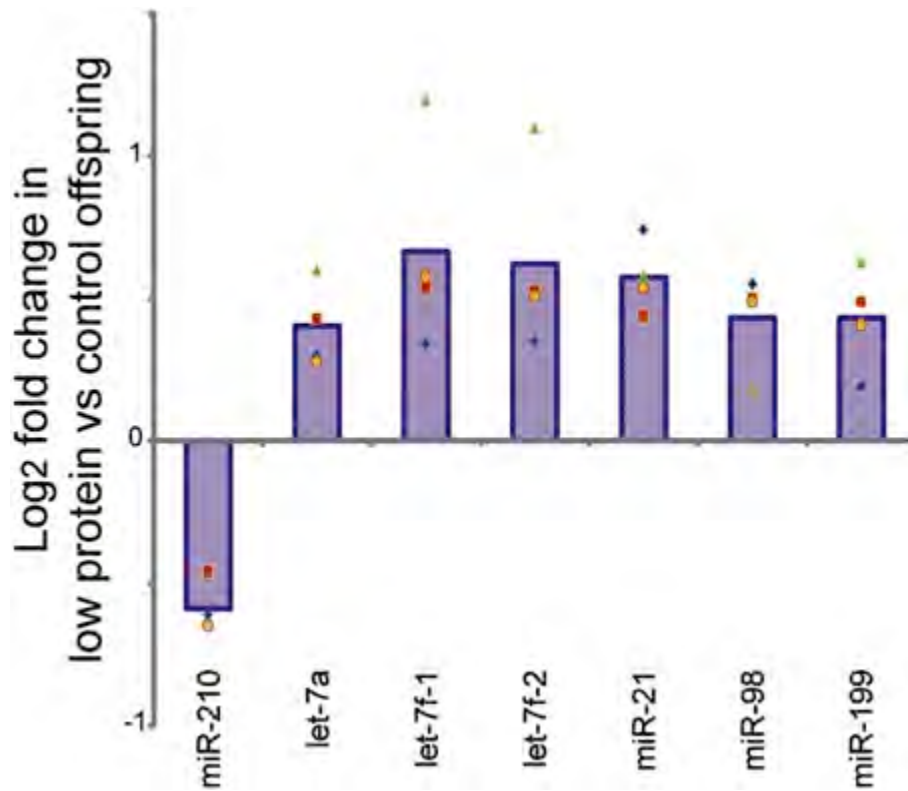
Figure 1.4

Proliferation-Related MicroRNAs Respond to Paternal Diet

Small (<35 nt) RNAs from the livers of eight offspring (four control, four low-protein) were isolated and subjected to high-throughput sequencing. MicroRNAs that exhibited consistent changes in all four pairs of animals are shown, with average change shown as a bar and individual comparisons shown as points. See also Table S1.4.

let-7 is well-known as a tumor suppressor (Jerome et al., 2007). The increase in growth-associated miRNAs is consistent with the hyperproliferative gene expression profile observed in the offspring of low protein diet fathers.

Proliferation-related microRNAs respond to paternal diet

We found no statistically-significant overlap ($p > 0.05$) between the predicted targets of the miRNAs here and the gene expression changes we observe, though the subtle (~50%) changes in miRNA abundance we observe might be expected to have little effect on mRNA – even when specific miRNAs are artificially introduced in cells, downregulation of target mRNAs is less than 2-fold for the majority of predicted targets (Hendrickson et al., 2008). Our results therefore suggest that miRNAs are likely to be additional targets of the reprogramming pathway, yet are likely not the direct upstream regulators of the entire response (but see (Wagner et al., 2008)).

Cytosine methylation in offspring

How are offspring reprogrammed by paternal diet? Cytosine methylation is a widespread DNA modification that is environmentally responsive, and carries at least some heritable information between generations (Bartolomei et al., 1993; Cropley et al., 2006; Holliday, 1987; Rakyan et al., 2003; Waterland and Jirtle, 2003b). As imprinted loci are often involved in growth control (Moore and Haig, 1991), we first asked whether a subset of candidate imprinted loci exhibited

altered cytosine methylation in low protein offspring (Figure S1.4A). As these loci did not exhibit significant changes in methylation, we therefore turned to genome-scale mapping studies to search for differentially methylated loci between control and low protein offspring.

We performed reduced representation bisulfite sequencing (RRBS, (Meissner et al., 2008)) to characterize cytosine methylation at single nucleotide resolution across ~1% of the mouse genome (Table S1.5). RRBS was performed for livers from a pair of control and low protein offspring, and fraction of methylated CpGs was calculated for a variety of features such as promoters, enhancers, and other nongenic CpG islands. In general, we found that cytosine methylation was well correlated between control and low protein offspring (Figures 1.5A, B). However, we did observe widespread modest (~10–20%) changes in CpG methylation between the two samples (red and green dots in Figures 1.5A, B), consistent with many observations indicating that environmental changes tend to have small quantitative effects on cytosine methylation (Blewitt et al., 2006; Heijmans et al., 2008; Ng et al., 2010; Weaver et al., 2004). Importantly, changes in promoter methylation did not globally correlate with changes in gene expression in offspring, indicating that the gene expression program in offspring is unlikely to be epigenetically specified at each individual gene (Figure 1.5C). Of course, widespread gene expression differences can be caused by changes to a small number of upstream regulators, and a number of differentially methylated regions are associated with cholesterol or lipid-related genes (Table S1.5).
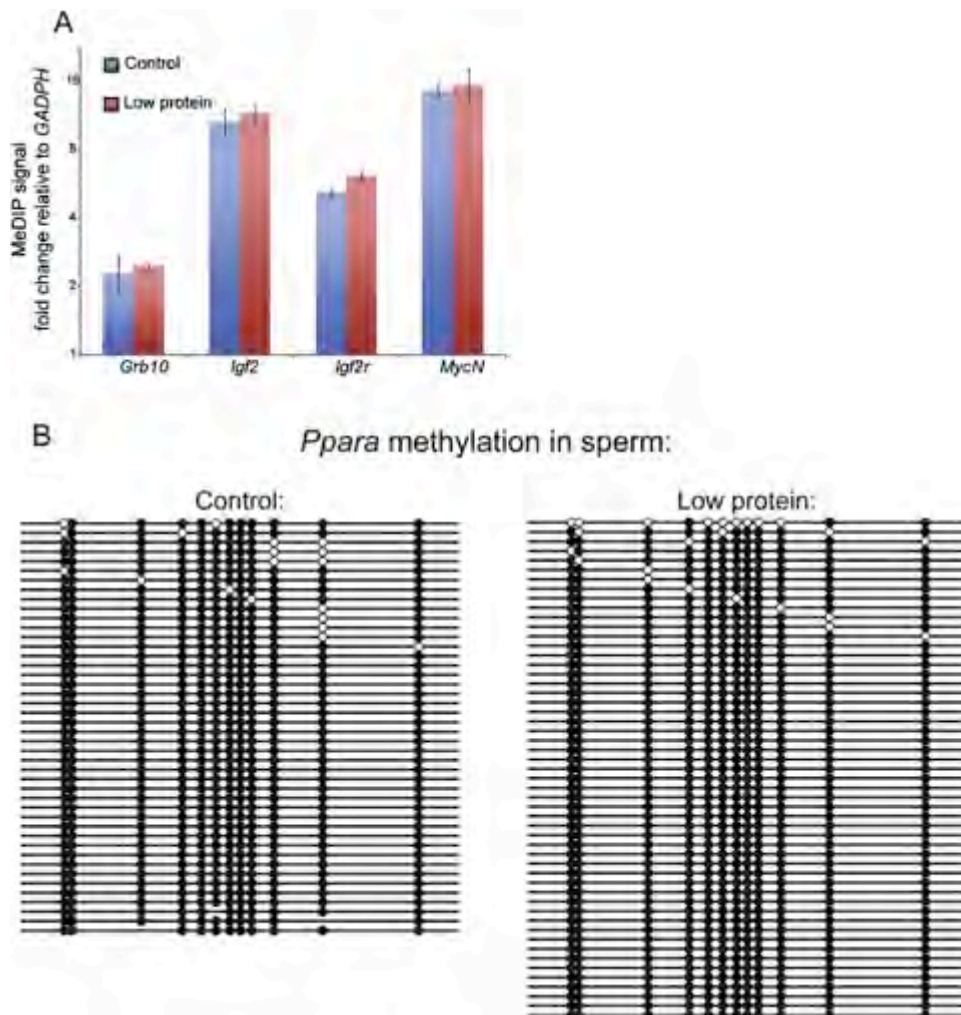
Figure S1.4

Cytosine Methylation in Offspring, Related to Figure 1.5, Figure 1.6, and Figure 1.7

(A) Growth-related imprinted genes are similarly methylated in control and low-protein offspring. Cytosine methylation was measured by methyl-cytosine immunoprecipitation (MeDIP) of genomic DNA from control and low-protein offspring livers, followed by q-PCR. Fold methylation relative to a control locus (*Gapdh*) is indicated.

(B) *Ppara* enhancer methylation in sperm is not responsive to diet. Individual bisulfite sequencing clones are shown for sperm isolated from males consuming control diet or low-protein diet. CpGs are shown as in Figure 1.6. Data from at least three animals are pooled here.

A Promoters

Methylation in Low Protein offspring (y-axis)
Methylation in Control offspring (x-axis)
Apof

B Non-genic CpG islands

Methylation in Low Protein offspring (y-axis)
Methylation in Control offspring (x-axis)
Wnt7b/Ppara
Map3k1
Ncor2
Smad2
Pdgfb

C

Δ promoter methylation (RRBS) (y-axis)
Δ mRNA (log2) (x-axis)
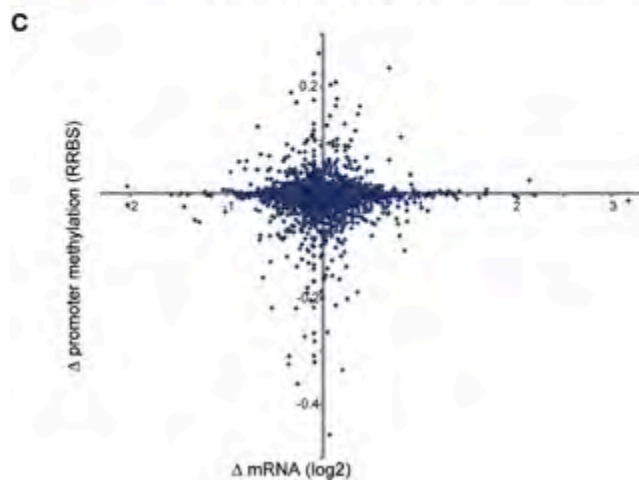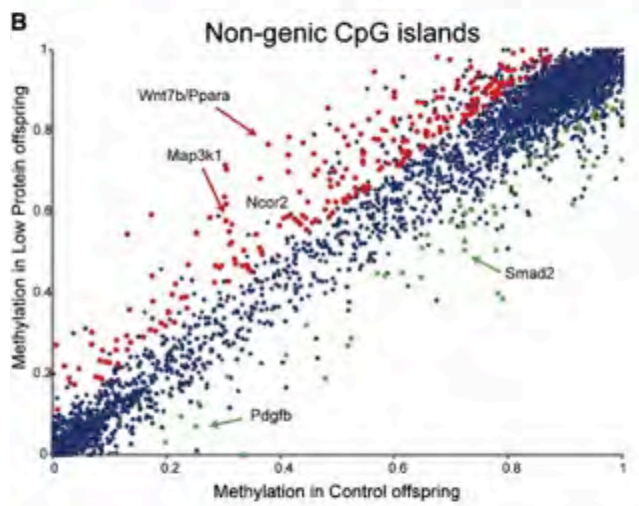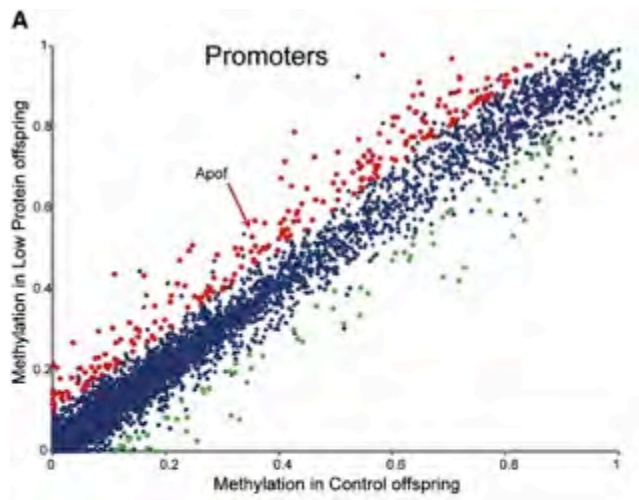
Figure 1.5.

Transgenerational Effects of Paternal Diet on Hepatic Cytosine Methylation

(A) Genomic DNA from control and low-protein offspring livers was subjected to reduced representation bisulfite sequencing (RRBS). For all annotated promoters, average fraction of CpGs that were methylated is shown for the control sample (x axis) compared to the low-protein sample (y axis). Red and green dots indicate promoters with significant ($p < 0.05$) methylation changes of over 10%.

(B) As in (A), for nongenic CpG islands.

(C) Promoter cytosine methylation changes are uncorrelated with gene expression changes. For each promoter, the average change in cytosine methylation is compared to the change in mRNA abundance from Figure 1.1B.

See also Table S1.5 and Figure S1.4.

Transgenerational effects of paternal diet on hepatic cytosine methylation

Most interestingly, we found a substantial (~30%) increase in methylation at an intergenic CpG island ~50 kb upstream of *Ppara* (Figure 1.6A). This locus is likely an enhancer for *Ppara*, as it is associated with the enhancer chromatin mark H3K4me1 (Heintzman et al., 2007) in murine liver (F. Yue and B. Ren, personal communication). *Ppara* is downregulated in the majority (but not all) of offspring livers (Table S1.1, Figure 1.6B), and the overall gene expression profile in our offspring livers significantly matches the gene expression changes observed in *Ppara*−/− knockout mice (Figure 1.2), suggesting that epigenetic regulation of this single locus could drive a substantial fraction of the observed gene expression changes in offspring. Indeed, variance of *Ppara* mRNA levels alone can be used to explain ~13.7% of the variance in the entire gene expression dataset (although this of course does not determine causality).

Effects of paternal diet on methylation of a putative *Ppara* enhancer

We therefore assayed the methylation status of this locus by bisulfite sequencing in an additional 17 offspring livers (8 control and 9 low protein), finding average differences of up to 8% methylation between low protein and control livers at several CpGs in this locus (Figure 1.6C). Importantly, these pooled data underestimate the potential role of this locus in reprogramming as they include animals exhibiting a range of changes in *Ppara* gene expression – individual

63
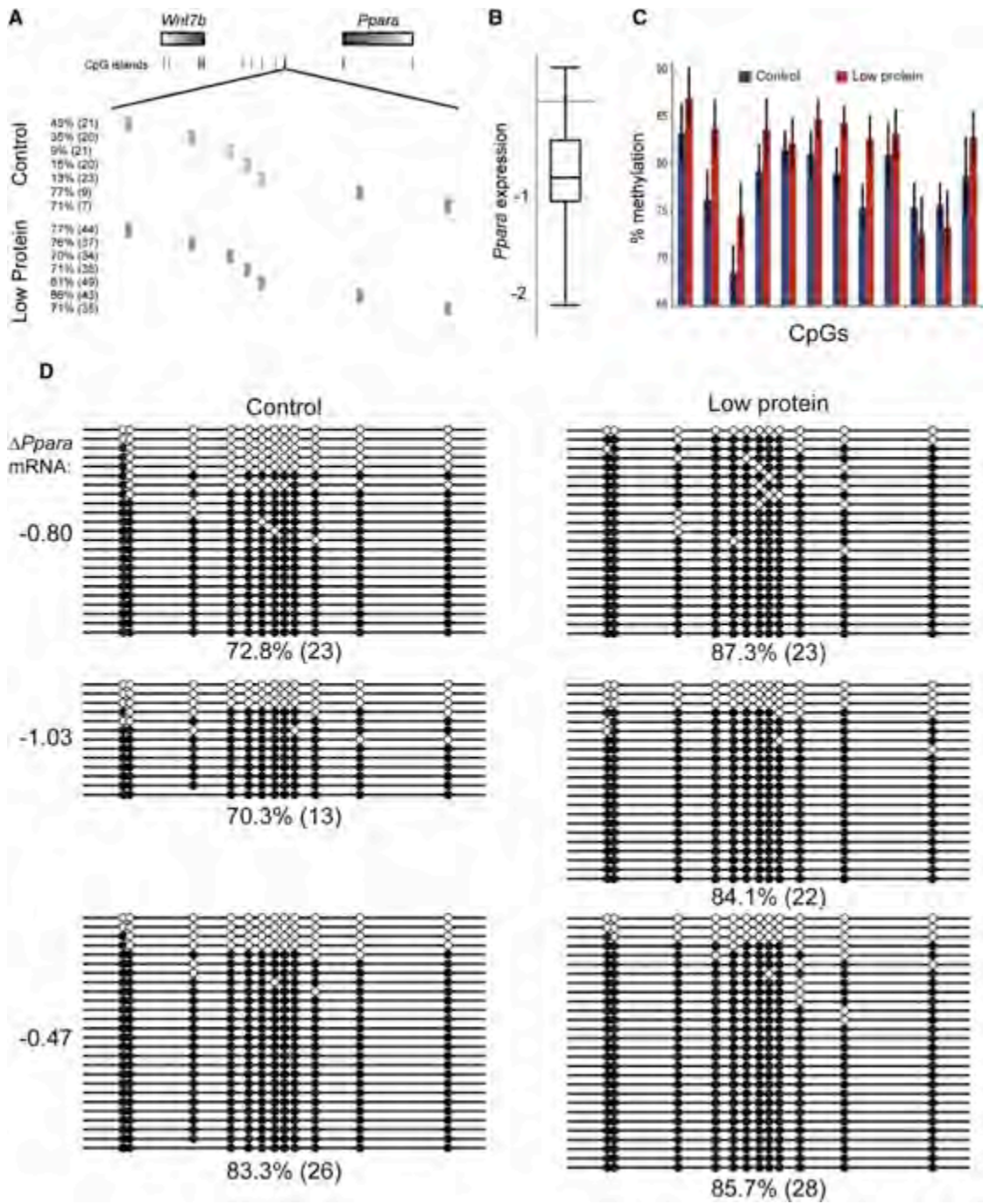
Figure 1.6

Effects of Paternal Diet on Methylation of a Putative *Ppara* Enhancer

(A) Differential methylation of a putative *Ppara* enhancer. Top panel shows a schematic of chromosome 15: 85,360,000–85,640,000. Zoomed in region represents chr15: 85,514,715-85,514,920. RRBS data for one control and one low-protein offspring pair are shown below, with assayed CpGs represented as boxes colored to indicate % of clones methylated. Numbers to the left indicate % methylation, with number of sequence reads covering the CpG in parentheses.

(B) *Ppara* is downregulated in most low-protein offspring livers. Box plot shows mean, quartiles, and highest and lowest values from Table S1.1.

(C) Putative enhancer methylation correlates with *Ppara* downregulation. DNA from eight control and nine low-protein pairs of offspring livers was bisulfite treated, and at least 13 clones were analyzed for each animal. Percent methylation at each of the 12 CpGs in this region plotted on the y axis; data are shown as mean ± standard error of the mean (SEM).

(D) Individual bisulfite clones are shown for three control and three low-protein offspring. White circles indicate unmethylated CpGs, black circles indicate methylated CpGs. Microarray data for change in *Ppara* RNA levels between the paired animals are shown to the left, in log2. Values under each bisulfite grouping indicate overall % methylation, with number of clones analyzed in parentheses.

animal pairs with large differences in *Ppara* mRNA levels exhibit differences of up to 30% at various cytosines across this locus. Figure 1.6D shows individual bisulfite clones for three pairs of animals with varying extents of *Ppara* downregulation (not all animals used for methylation analysis were analyzed by microarray). Taken together, these results identify a differentially methylated locus that is a strong candidate to be one of the upstream controllers of the hepatic gene expression response.

Cytosine methylation, RNA, and chromatin in sperm

The link between paternal diet and offspring methylation patterns lead us to consider the hypothesis that paternal diet affects cytosine methylation patterns in sperm. We therefore isolated highly pure (>99%) sperm from the caudal epididymis of males consuming control or low protein diet. We assayed the *Ppara* enhancer for methylation by bisulfite sequencing, but found no significant changes between males consuming control or low protein diet (Figure S1.4B). These results indicate either that cytosine methylation in sperm is not the relevant paternally-transmitted dietary information at this locus (but changes at some point during development – (Blewitt et al., 2006)), or that we captured animals whose offspring would not manifest significant changes in expression of the associated genes – as seen in Figures 1.1B or or 1.6B, *Ppara* downregulation is variably penetrant in low protein offspring.

To globally investigate effects of paternal diet on sperm cytosine methylation, we isolated sperm from four males – two consuming control diet, one consuming low protein diet, and one subjected to a caloric restriction regimen. We then surveyed cytosine methylation patterns across the entire genome via MeDIP-Seq (immunoprecipitation using antibodies against 5me-C followed by deep sequencing (Jacinto et al., 2008; Weber et al., 2005)) (Figures 1.7A, S1.5A, S1.6). Notably, global cytosine methylation profiles were highly correlated between any pair of samples, indicating that the sperm "epigenome" is largely unresponsive to these differences in diet (Figures 1.7B–D, S1.5B–E). Indeed, littermates on different diets (Figures 1.7B, C) were better correlated for promoter methylation than were the pair of control animals from different litters (Figure 1.7D). While these results do not rule out cytosine methylation in sperm as the relevant carrier of epigenetic information about paternal diet, the high correlation between samples, coupled with the absence of cytosine methylation changes at the *Ppara* enhancer in sperm, lead us to consider alternative epigenetic information carriers including RNA (Rassoulzadegan et al., 2006; Wagner et al., 2008) and chromatin (Arpanahi et al., 2009; Brykczynska et al., 2010; Hammoud et al., 2009; Ooi and Henikoff, 2007).

Modest effects of diet on the sperm epigenome

We analyzed RNA levels for three pairs of males and for two matched epididymis samples by Affymetrix microarray (Figures S1.6, S1.7A). Curiously, low protein
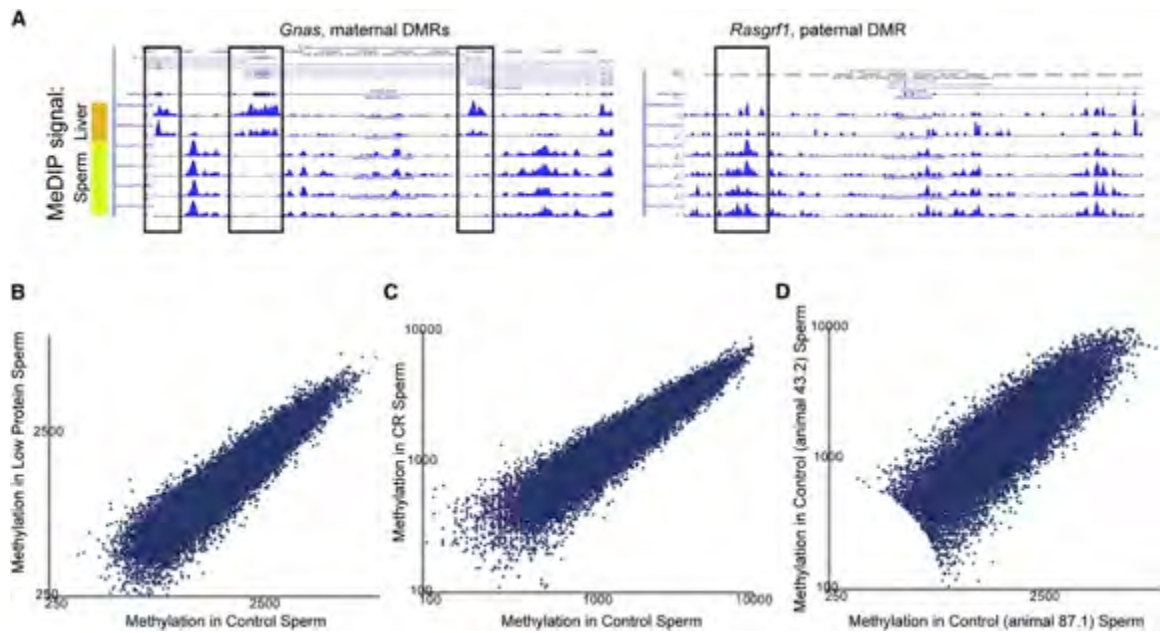
Figure 1.7

Modest Effects of Diet on the Sperm Epigenome

(A) MeDIP sequencing data are shown for two liver samples (top two tracks) and four sperm samples (bottom four) at a maternally methylated region (*Gnas*, left) and a paternally methylated region (*Rasgrf1*, right).

(B) Comparison of control and low-protein methylation. For each promoter, methylation levels were averaged for 8 kb surrounding the TSS, and values are scatterplotted for control sperm (x axis) versus low-protein sperm (y axis). x and y axes are plotted on logarithmic scales.

(C) As in (B), but for control versus caloric restriction.

(D) As in (B), but for the pair of control samples.

Similar results for (B)–(D) are found when focusing on the 1 kb surrounding the TSS (not shown). See Figure S1.7 for analyses of consistent RNA and chromatin differences between low-protein and control sperm.

A

B

Sperm MeDIP data aligned by TSS

Control (87.1)   Control (43.2)   Low Protein (87.2)   Caloric Restriction (43.1)

C

Sperm MeDIP data aligned by nongenic CGI

Control (87.1)   Control (43.2)   Low Protein (87.2)   Caloric Restriction (43.1)

D

E
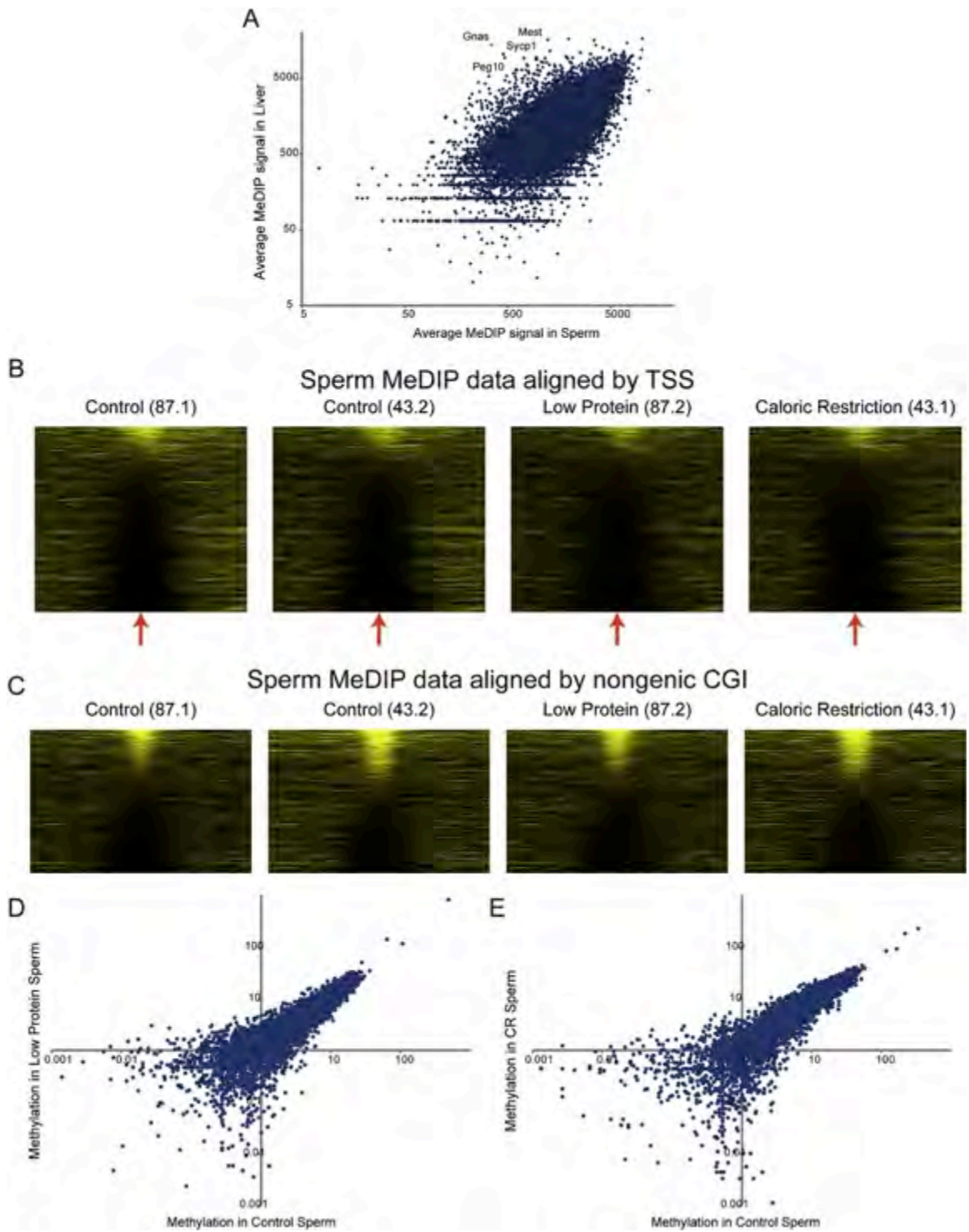
Figure S1.5

Modest Effects of Diet on the Sperm Epigenome, Related to Figure 1.7

(A) MeDIP data identify expected tissue-specific methylation patterns. Scatterplot of average methylation for 8 kb surrounding the TSS for ~22,000 annotated genes—average methylation in four sperm samples is shown on x axis, average methylation in two liver samples is shown on y axis (axes are on a log scale). Genes exhibiting high methylation in liver but not sperm include a number of maternal differentially methylated regions, and overall are enriched for genesets expressed at high levels in sperm (with GO annotations such as spermatogenesis). Conversely, genes exhibiting relatively high promoter methylation in sperm include a wide variety of developmental regulators (with GO annotations such as organ development), and many genes highly-expressed in the liver (with GO annotations such as lipid metabolic process).

(B) Sperm cytosine methylation is globally similar across the genome under three distinct dietary regimes. Methylation data was mapped to the mouse genome, and data for 22,000 named genes is aligned by transcription start site (red arrow). For all four samples, genes are ordered by the extent of methylation across the 1 kb surrounding the TSS in the leftmost control animal.

(C) As in (B), but for 6 kb surrounding ~5000 nongenic CpG islands. All data are sorted by the extent of methylation in the central 1 kb for the 87.1 control sample.

(D and E) Scatterplots of average MeDIP-seq signal over the central 1 kb of nongenic CpG islands for the indicated pairs of samples.
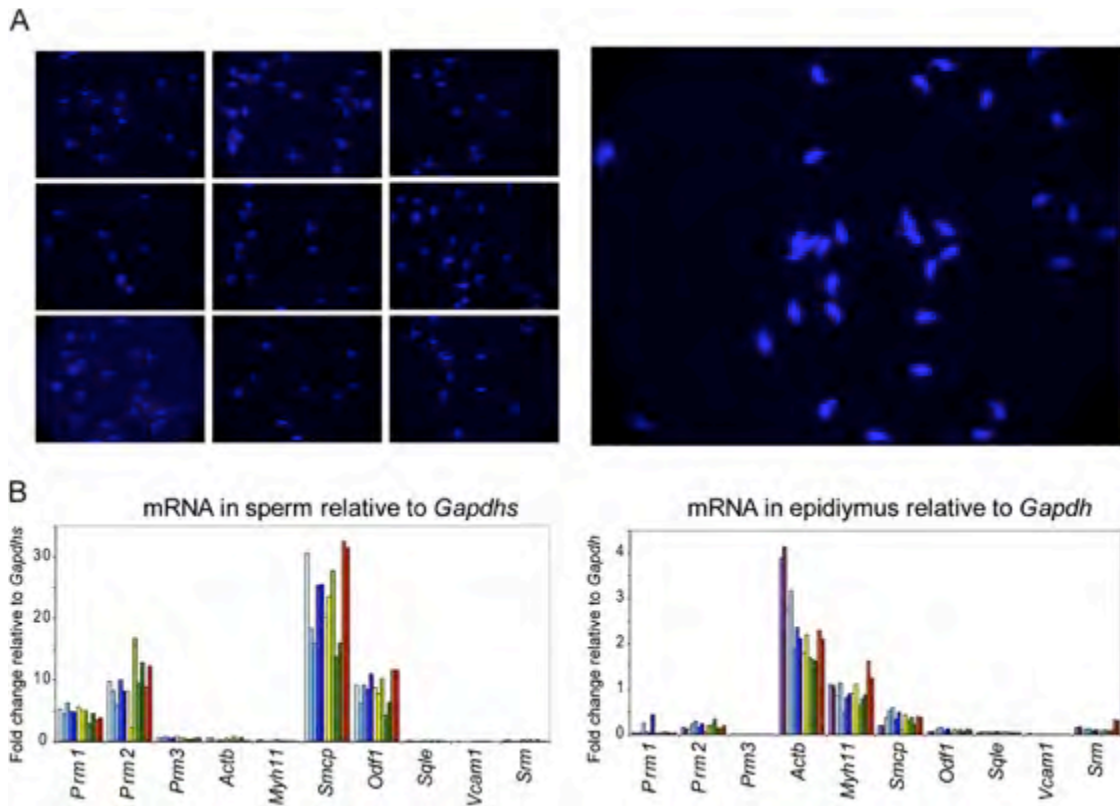
Figure S1.6

Characterization of Sperm Preparations, Related to Figure 1.7

(A) DAPI-stained images of a typical control sperm preparation. Ten fields are shown, with 100% of nuclei showing characteristic murine sperm morphology. All nuclei shown also strained positively with anti-Sycp3 (not shown).

(B) q-RT-PCR of sperm RNA samples show high levels of sperm-enriched genes such as *Prm1*, *Prm2*, *Smcp*, and *Odf1*, and low levels of epididymis-enriched genes such as *Actb* and *Myh11* (left panel shows sperm RNA samples, right panel shows epididymis RNA samples). Different bars within each set represent independent sperm samples.

A

Number of genes

Enriched GO categories:
translation, spermatogenesis,
metabolism, catabolism,
apoptosis, glycolysis

Sperm RNA abundance (arbitrary units)

B

Control sperm

LP/CR sperm

Epididymis

C

Log2(Low protein sperm/control sperm)

Log2(Sperm/epididymis)

Pparo

Smarcd3

D

LP/C (rep 1)

LP/C (rep 2)

CR/C

Ppard
Smarcd3
Igf1r
Impact
Mlf1
Fad2
Elmsr1
Ptrc3
Lss
Hoxb5
Bglap

E

Log2 low protein sperm/control sperm

Dnahc3 Kpna2 Pfkp Sucla2 Mdh1b Acsbg2 Mas1 Actb Gmcl1 Prm2 Ppard Smarcd3 Bglap2

Microarray 1 (after correction for epididymis)
Microarray 2 (after correction for epididymis)
q-RT-PCR (average of 7 pairs of animals)

F

Log2(Low protein sperm/control sperm)

Acsbg2 Actb Bglap2 Dnahc3 Mdh1b Pfkp Ppard Smarcd3 Sucla2

G

H3K27me3

Log2(control/low protein)

Eftud1 Maoa Mid1

H

pan-H3

Eftud1 Maoa

71

Figure S1.7

Effects of Diet on Sperm RNA Content and Chromatin, Related to Figure 1.7

(A) Sperm RNA populations exhibit expected enrichments. Histogram of average Affymetrix microarray probe intensities for all six sperm samples is shown, with abundant RNAs in sperm exhibiting expected GO enrichments.

(B) Sperm from animals consuming low-protein or caloric restriction diets exhibit relative depletion of epididymis-enriched genes, relative to sperm from animals on control diet. Data from 8 Affymetrix microarray analyses are shown. Log-transformed abundance data for each gene was row-normalized (i.e., the average value of each row is zero), and genes with fold change > 1.8 in at least two samples are shown. Thus, the upper half of the cluster shows genes that are relatively abundant in epididymis (red), relatively depleted in low-protein and caloric restriction sperm (green), and of intermediate abundance in control sperm (black/light green).

(C) Low-protein sperm are more "sperm-like" than are control sperm. Scatterplot of difference in RNA signal between sperm and epididymis (x axis) versus difference between sperm from one of the pairs of low-protein versus control animals (y axis). Red line shows LOWESS fit between sperm/epididymis and low-protein/control, and red and green dots show genes exhibiting a "corrected" low-protein/control enrichment above or below 1.8-fold.

(D) Cluster of corrected sperm RNA changes between two low-protein/control pairs and one caloric restriction/control pair. Genes depleted in low-protein sperm are enriched for GO annotations including lipid metabolism, regulation of transcription, and organ development.

(E) Validation of microarray results. q-RT-PCR was performed for the indicated genes, normalized against *Gapdh*, and low-protein/control ratios are shown (± SEM). Microarray values shown are LOWESS-corrected for possible epididymis contamination as in (C).
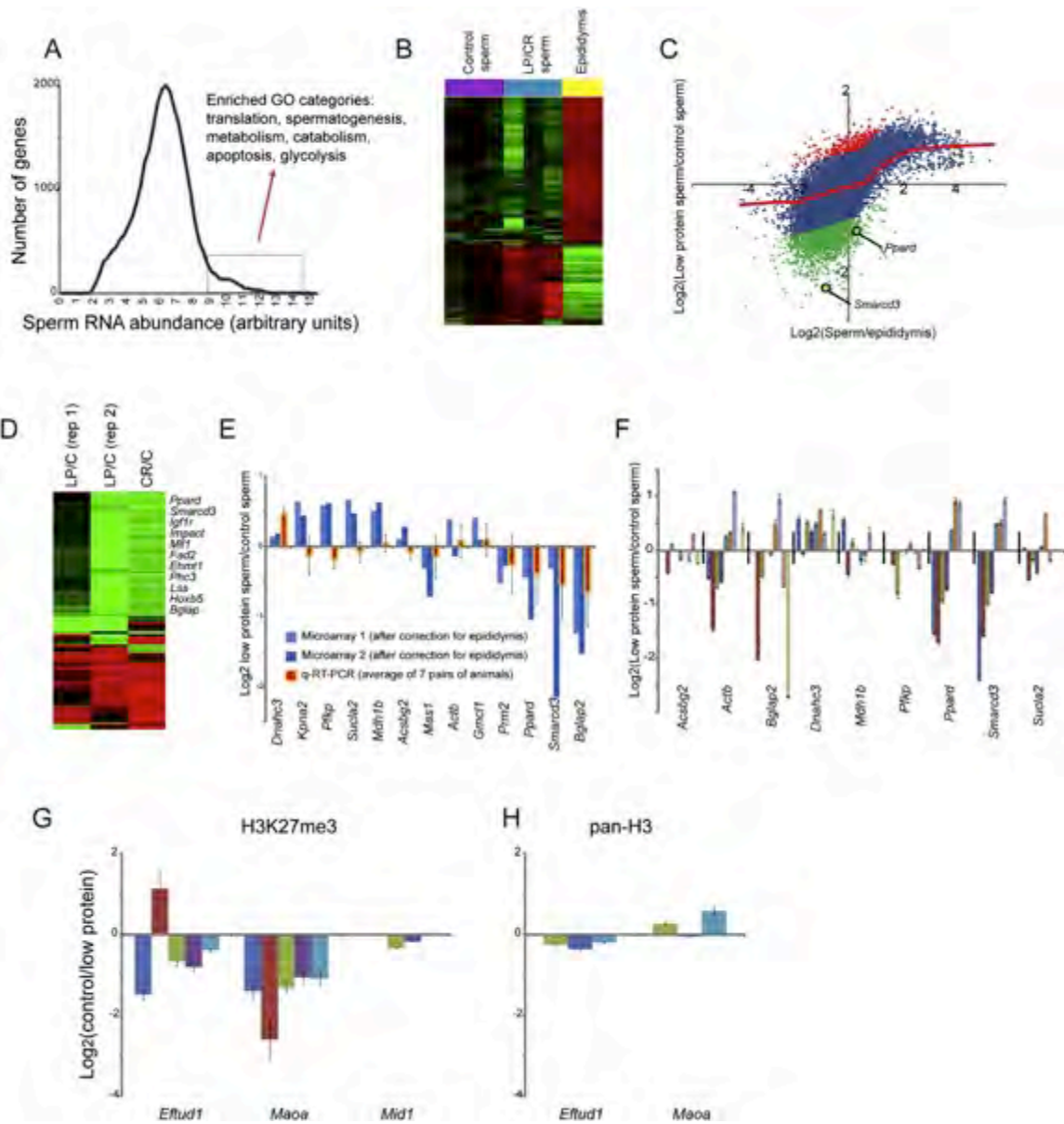
(F) Individual low-protein/control ratios for nine animal pairs (most genes only have data for seven or eight of the nine pairs due to failed PCRs) used for (E). Note that there is significant variability in RNA changes, with the only consistent change being increased *Dnahc3* levels in an additional 7/8 low-protein animals examined. *Smarcd3*, *Bglap*, and *Ppard* trend in the same direction as the microarray data, but variability results in insignificant (p > 0.05 by t test) changes.

(G) H3K27me3 decreases over the *Maoa* and *Eftud1* promoters in low-protein sperm. Data are shown as mean ± standard deviation. Q-PCR was carried out with primers to the indicated promoters, and normalized to *Kctd16*. Data for *Mid1* show that choice of normalization control is not the major driver for this result.

(H) H3 levels do not change significantly at *Maoa* and *Eftud1* promoters. As in (G), but with anti-pan H3 ChIP.

and caloric restriction samples consistently exhibited more "sperm-like" RNA populations (as opposed to epididymis RNA) than did control samples (Figure S1.7B–C). Whether this reflects systematic contamination issues or biological differences in sperm maturity or quality is presently unknown, although we note that we confirmed consistently-higher levels of the sperm-specific *Dnahc3* by q-RT-PCR in an additional 7/8 low protein sperm samples (Figure S1.7E). We note that control sperm samples were routinely >99.5% sperm as assayed by microscopy (Figure S1.6), but nonetheless we cannot completely rule out systematic contamination issues. With this possibility in mind, we identified genes were differentially-packaged in control vs. low protein sperm by correcting for potential epididymal contamination (Figures S1.7B–F). Interestingly, we observed downregulation of a number of transcription factors and chromatin regulators such as *Smarcd3* and *Pparδ*, although q-RT-PCR validation was not statistically significant due to high inter-animal variability (Figure S1.7F).

Although the downregulation of *Smarcd3* was not significantly confirmed by q-RT-PCR, this could reflect the variable penetrance of paternal diet on offspring described above. Given that heterozygous mutants in chromatin remodelers can affect offspring phenotype even when the mutant allele segregates away (Chong et al., 2007), we used an initial genome-wide mapping (not shown) of overall histone retention (pan-H3 ChIP) abundance and the key epigenetic histone modification H3K27me3 in sperm to identify targets for single locus analysis. We

73

observed a consistent decrease in H3K27me3 in low protein sperm at the promoter of *Maoa* (Monoamine oxidase) in 5/5 pairs of sperm samples, and a decrease in H3K27me3 at *Eftud1* in 4/5 paired samples (Figure S1.7G–H). These results demonstrate proof of principle that the sperm epigenome is regulated by dietary conditions, although the biological implications of these observations are not yet clear.

Discussion

Taken together, our results demonstrate that paternal diet affects lipid- and proliferation-related gene expression in the offspring of inbred mice, and that epigenetic information carriers in sperm respond to environmental conditions. These results have potential implications for human health, and raise numerous mechanistic questions, discussed below.

Paternal diet affects metabolism in offspring

Our results clearly identify a set of physiological pathways whose expression is sensitive to paternal diet. Specifically, we find that hepatic expression of genes involved in proliferation and cholesterol biosynthesis can be regulated by paternal diet, and these changes are reflected in levels of several lipid metabolites. Combined with data showing that offspring glucose levels are

affected by paternal fasting in mice (Anderson et al., 2006), these results demonstrate that paternal diet has wide-ranging effects on the metabolism of offspring in rodents. Interestingly, a very recent study from Ng et al (Ng et al., 2010) reported that chronic exposure of male rats to high fat diet was associated with pancreatic beta cell dysfunction in female offspring. It will naturally be of great interest in the near future to compare the transgenerational effects of high fat and low protein diets, although one clear difference is that in our system a transgenerational effect is observed in both sex offspring.

Whether the effects we observe on cholesterol metabolism prove advantageous in low protein conditions remains to be tested, but it will be important to investigate ecologically-relevant diets in order to speculate more firmly about adaptive significance of any observed transgenerational effects. For example, at present we cannot say with certainty what aspect of the low protein regimen is sensed by males – it is possible that offspring metabolism is affected by overall protein consumption, or high sucrose, or fat/protein ratio, or even levels of micronutrients, as our males consumed diets ad libitum and thus might have overconsumed the low protein diet.

The reprogrammed state: liver

What is the mechanistic basis for the reprogrammed gene expression state? Genome-scale analyses of cytosine methylation in offspring livers identified

several lipid-related genes that were differentially-methylated depending on paternal diet. Most notably, a putative enhancer for a major lipid regulator, *Ppara*, exhibited generally higher methylation in low protein offspring than in control offspring. Methylation at this locus was variable between animals, consistent with the partial penetrance of *Ppara* downregulation in our dataset. The overall gene expression profile observed in low protein offspring significantly overlaps gene expression changes observed in *Ppara−/−* mice (Rakhshandehroo et al., 2007), leading to the hypothesis that epigenetic *Ppara* downregulation via enhancer methylation is an upstream event that affects an entire downstream regulon in reprogrammed animals. Note that while the hepatic downregulation of *Ppara* suggests a liver-autonomous epigenetic change, we cannot rule out that hepatic gene expression changes result from global physiological changes resulting from downregulation of *Ppara* in some other tissue.

Interestingly, *Ppara* expression in liver is also regulated by maternal diet – offspring of female mice consuming a high fat diet exhibit altered hepatic *Ppara* expression, with increased expression at birth but decreased expression at weaning (Yamaguchi et al., 2010). Together with our data, these results suggest that *Ppara* is a key nexus that integrates ancestral dietary information to control offspring metabolism.

Mechanistic basis for transgenerational paternal effects

Paternal diet could potentially affect offspring phenotype via a number of different mechanisms. While we focus here on epigenetic inheritance systems, it is important to note that parental information can also be passed to offspring via social or cultural inheritance systems (Avital and Jablonka, 2000; Champagne and Meaney, 2001; Jablonka and Lamb, 1995; Meaney et al., 2007; Weaver et al., 2004). While such maternally-provided social inheritance is unlikely in our paternal effect system – males were typically only in females' cages for one day – it is known that in some animals females can judge mate quality and allocate resources accordingly (Pryke and Griffith, 2009), and that seminal fluid can influence female postcopulatory behavior in Drosophila (Fricke et al., 2008; Wolfner, 2002). These and other plausible transgenerational information carriers cannot be excluded at present – ongoing artificial insemination and in vitro fertilization experiments will determine whether sperm carry the relevant metabolic information in our system.

Here we focused on the hypothesis that paternal dietary information does indeed reside in sperm epigenetic information carriers. First, a subset of cytosine methylation patterns in sperm are known to be heritable (Chong et al., 2007; Cropley et al., 2006; Rakyan et al., 2003; Waterland and Jirtle, 2003b). Second, several reports suggest that RNA molecules packaged in sperm can affect offspring phenotype (Rassoulzadegan et al., 2006; Wagner et al., 2008). Third,

77

chromatin structure has been proposed to carry epigenetic information, as sperm are largely devoid of histone proteins but retain them at a subset of developmentally-important loci (Arpanahi et al., 2009; Brykczynska et al., 2010; Chong et al., 2007; Hammoud et al., 2009). Finally, it is conceivable that additional or novel epigenetic regulators (such as prions) are packaged into sperm, or that sperm quality is affected by diet, or that genetic changes are directed by the environment (although it is important to emphasize that inbred mouse strains were used in this study).

Here, we report whole genome characterization of cytosine methylation patterns and RNA content in sperm obtained from mice maintained on control, low protein, and caloric restriction diets. Globally, cytosine methylation patterns are similar in all three conditions, indicating that the sperm epigenome is largely unaffected by these diets. Nonetheless, changes in relatively few loci can have profound effects in the developing animal, and our data do not rule out the possibility of inheritance through sperm cytosine methylation, especially given that MeDIP is unlikely to identify ~10–20% differences in methylation at a small number of cytosines. Importantly, the putative enhancer of *Ppara* (Figure S1.4) was not differentially methylated in sperm. It will therefore be of great interest in the future to determine when during development the differential methylation observed in liver is established, and to identify the upstream events leading to differential methylation (Blewitt et al., 2006).

Interestingly, we did identify effects of diet on RNA content and chromatin packaging of sperm. For example, sperm from control animals were consistently depleted of the highly sperm-specific *Dnahc3* gene (Figure S1.7) relative to sperm from low protein animals. We cannot presently determine whether this represents reproducible differences in contamination, differences in sperm maturity, or something else. Finally, based on our observation that low protein sperm tended to be depleted of genes encoding a number of chromatin regulators, we have begun to search for dietary effects on sperm chromatin structure. Interestingly we found that the *Maoa* promoter was consistently depleted of the key Polycomb-related chromatin mark H3K27me3 (Figure S1.7G), demonstrating as a proof of concept that chromatin packaging of the sperm genome is responsive to the environment, and motivating genome-wide investigation into dietary effects on sperm chromatin. Given the common behavorial changes observed in other transgenerational inheritance paradigms, the possibility that H3K27me3 at *Maoa* affects offspring behavior (potentially via altered offspring responses to maternal stress – (Harris and Seckl, 2011)) will be of great future interest.

Relevance to human disease

These results are likely to be relevant for human disease, because not only is maternal starvation in humans correlated with obesity and diabetes in children (Lumey et al., 2007), but, remarkably, limited food in paternal grandfathers has

also been associated with changed risk of diabetes and cardiovascular disease in grandchildren (Kaati et al., 2002; Pembrey et al., 2006). Interestingly, in these studies ancestral access to food and disease risk was not associated with disease risk in the next generation, but was only associated with F2 disease risk. However, it is important to note that the transgenerational effects of food availability for paternal grandfathers depend on the exact period during childhood of exposure to rich or poor diets (Pembrey et al., 2006), whereas our experimental protocol involved continuous low protein diet from weaning until mating. Thus, future studies are required to define when and how paternal exposure to a low protein diet affects epigenetic programming of offspring metabolism.

Together, these results suggest rethinking basic practices in epidemiological studies of complex diseases such as diabetes, heart disease, or alcoholism. We believe that future environmental exposure histories will need to include parental exposure histories as well as the exposure histories of the patient, to disentangle induced epigenetic effects from the currently-sought genetic and environmental factors underlying complex diseases. Our observations provide an inbred mammalian model for transgenerational reprogramming of metabolic phenotype that will enable dissection of the exposure history necessary for reprogramming, genetic analysis of the machinery involved in reprogramming, and suggest a number of specific pathways likely to be the direct targets of epigenetic reprogramming.

Methods

Mouse husbandry

All animal care and use procedures were in accordance with guidelines of the Institutional Animal Care and Use Committee. C57/Bl6 mice were obtained from Jackson Labs and from Charles River Laboratories (for different iterations of this experiment). All experiments were performed with mice, which had been raised for at least two generations on control diet to attempt to minimize any transgenerational effects of transitioning to control diet from chow provided by animal provider. For all comparisons shown, male mice were weaned from mothers at 21 days of age, and sibling males were put into cages with low protein or control diet (moistened with water to allow mice to break the hard pellets). Females were weaned to control diet. Males were raised on diet until 9–12 weeks of age, at which point they were placed with females for one or two days. Control and low protein mating cages were always interspersed with one another. Note we always used virgin females to avoid confounding effects of the female's litter number, although this results in many lost litters as first litters were often consumed by their mothers. After one to two days, males were removed, and pregnant females were left alone with control diet and a shepherd shack until their litters were three weeks of age. At three weeks of age offspring were sacrificed by isoflurane and cervical dislocation, and median lobe of liver was rapidly dissected out and flash-frozen in liquid N2.

Diets

Diets were obtained from Bio-serv, and compositions are listed in Table S1.7. For most experiments only low protein diet was sterilized per standard protocol at Bio-serv. For later experiments, both diets were sterilized.

Table S1.7A
Murine Diet Composition

Ingredient	Control	Low Protein
Cornstarch	39.75% (w/w)	39.75%
Casein	20%	10.9%
Maltodextrin	13.2%	13.2%
Sucrose	10%	19.1%
Soybean Oil	7%	7%
Cellulose	5%	5%
Mineral Mix	3.5%	3.5%
Vitamin Mix	1%	1%
L-Cysteine	0.3%	0.3%
Choline Bitartrate	0.25%	0.25%
Tert-butyl hydroquinone	0.0014%	0.0014%

RNA extraction

Liver samples were ground with a liquid N2-cooled mortar and pestle. Total RNA for microarray analysis was extracted from liver powder using Trizol.

Microarray hybridization

30 μg of total RNA was labeled for 2 hours at 42 C with Superscript II reverse transcriptase using 4 μg of random hexamer and 4 μg of oligo dT. Cy3 and Cy5-labeled samples were hybridized to home-printed "MEEBO" microarrays. MEEBO information is at http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPL6352. Microarrays were hybridized at 65 C for 16 hours, washed as previously described (Diehn et al., 2002), and scanned using Axon Genepix 4000B microarray scanner.

Data availability

All microarray data and deep sequencing data used in this study have been deposited to GEO (http://www.ncbi.nlm.nih.gov/projects/geo/), accession # GSE25899. Tables are available at http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3039484/?report=classic#SD2

Comparison to public murine liver microarray data

We built a compendium of public microarray data consisting of 120 gene-expression profiles in the murine liver under various conditions and genetic perturbations. Signatures of differentially expressed genes were determined using a combination of two one-tailed t-test, with FDR correction of 0.1. Profiles significantly enriched with up or down regulated genes in low protein offspring were defined by a Hyper-Geometric p-value $<= 0.05$ after correction for multiple hypotheses (p<0.00025).

Lipid measurements

~50–100 mg of ground liver tissue from six animals (three paired sets) was sent to Lipomics for "Truemass" mass spectrometry characterization of 450 lipid levels (Table S4). Note that samples 73-1 and 76-1 come from PBS-perfused livers, while the other four samples were dissected without perfusion.

Small RNA cloning and sequencing

Total RNA was isolated from ground liver tissue using mirVana (Ambion). 18–35 nt small RNA was purified from 100 µg of total RNA, ligated to adaptors, amplified, gel-purified, and sequenced using a Solexa Genome Analyzer (Illumina) (Ghildiyal et al., 2008).

RRBS

Reduced representation bisulfite sequencing was carried out as previously described (Meissner et al., 2008). Data are available at http://thrifty-epigenome.computational-epigenetics.org

Sperm isolation

Caudal epididymis was dissected from sacrificed animals, punctured, and incubated for 30 minutes in M2 media (Sigma) at 37 C. Supernatant was removed, pelleted (3,000g for 5 minutes), and washed 2× with PBS, 1× in water, and incubated in Somatic Cell Lysis buffer. Sperm preparations were only used that were >99.5% pure as assessed by microscopy, and q-RT-PCR was also used to reject any sperm samples based on the ratio between epididymis-specific genes *Actb* or *Myh11* compared to sperm-specific genes *Smcp* or *Odf1* (Figure S1.6).

MeDIP

Methyl-DNA immunoprecipitation was carried out essentially as described (Weber et al., 2005; Weber et al., 2007). 4 µg of purified genomic DNA was fragmented to a mean size of 300bp using a Covaris machine, denatured, and immunoprecipitated with 5mC antibody (Eurogentec). ChIP material was Solexa sequenced, with ~21 million uniquely mappable reads per library.

Primers

gene	Fw primer	Rev primer	Assay:
Acsbg2	TGCTGAGCGATTCAGTGCTT	TGGCTGCCTTTCGACACATT	q-RT-PCR
Actb	ACTATTGGCAACGAGCGGTT	AGCACTGTGTTGGCATAGAGGT	q-RT-PCR
Bglap	ACCCTGCTTGTGACGAGCTATC	AGTTTGGCTTTAGGGCAGCACA	q-RT-PCR
Cebpb	CGCACCACGACTTCCTCTC	CGAGGCTCACGTAACCGTAGT	q-RT-PCR
Dnahc3	ACTTGGCAGGCTTGCAGAAACT	TCGCTGTATCCTCAGAGGTTTGGA	q-RT-PCR
Elovl5	TTCGATGCGTCACTCAGTACCT	TGGTCCCAGCCATACAATGAGT	q-RT-PCR
Fads1	TCTTTGGCACCTCGACATGGAA	AAGCAGTTAGGCTTGGCATGGT	q-RT-PCR
Gapdhs	TTGGCTGGCATCCTTGCTTACA	AGGGCAATTCCAGCCTTAGCAT	q-RT-PCR
Gmcl1l	AGTGGTTGACTTTGCTGCGA	TCGTCCAACTGACCACTTCCAA	q-RT-PCR
Hmgcs1	TCTTCAATGCCGTGAACTGGGT	TCCACCTGTAGGTCTGGCATTT	q-RT-PCR
Igfbp1	AACGCCATCAGCACCTATAGCA	TGTTGGGCTGCAGCTAATCTCT	q-RT-PCR
Insig1	GCCAGCGTTATGCGCTGTA	GGAACGATCAAATGTCCACCAC	q-RT-PCR
Kpna2	TGCAGGAGCACTTGCAGTCTTT	AGGCCGTGATTCACAACTTGCT	q-RT-PCR
Lpin1	AACCTGGAAATGCTCTGGCTGT	ACTCGCTGTGAATGGCCTGAAA	q-RT-PCR
Mas1	TATTTGGCTACAACACGGGCCT	AGACGAATGCTGACTGGTGCTT	q-RT-PCR
Mdh1b	ACTTCCAACATGACGACGGAGT	TGGCAATCCAGACCTGCAAA	q-RT-PCR
Myh11	TCCTTCCTGGGCATTCTGGATA	TGCAGCTTCTCGTTGGTGTAGT	q-RT-PCR
Odf1	ATCGCTCCGCAGTTTAGAGAGA	CAGGTTCAAAGCCGCACACATT	q-RT-PCR
Pfkp	TTTCAACCAGTGGCAGAGCTGA	ATCTTCATGATGGGCCGCAGTT	q-RT-PCR
Ppard	AGACAAACCCACGGTAAAGGCAGT	TCCTGTGGCTGTTCCATGACTGA	q-RT-PCR
Prm1	ACCTTTCTAGGATGCTGCCGT	TGTGGCGAGATGCTCTTGAAGT	q-RT-PCR
Prm2	CTACATAGGATCCACAAGAGGCGT	GCTTAGTGATGGTGCCTCCTACAT	q-RT-PCR
Prm3	TGGCCTGTGTGAGTCAAGACAA	TTCACCGGGATTTGCTCCTCTT	q-RT-PCR
Scd1	AAGTGAGGCGAGCAACTGACTA	ACCGTCTTCACCTTCTCTCGTT	q-RT-PCR
Slx	AGGCCAGCATTTGGCAAACA	AAACGCTGCCATTCCACTGAG	q-RT-PCR
Sly	TATCGCCTCAAGCAGAAGCA	TGCCTACTTCATGTCCCATGTC	q-RT-PCR
Smarcd3	ACTCTGATGGCAGCATTGCCT	TGGTTGTCTGGGCCATAAAGGT	q-RT-PCR
Smcp	AGAGCCCAAGGAAGAACTGTGT	ACAGCAAGGTGGTTTCTGTGGA	q-RT-PCR
Sqle	TCCTTGCATCAGCTCCGAAA	GAAAGCAACCCAACAGGACCG	q-RT-PCR
Srm	TGGTCCAGTGCGAGATTGATGA	TCATGAACTCAAAGCCATCGCC	q-RT-PCR
Sucla2	ACATTGAAGATGTGGCTGCCGA	TTCTGTGCAAGTGTCACGGCTT	q-RT-PCR
Sycp3	AAGGGCCAGGTTTCCTCAGAT	ATCAACCAAAGGTGGCTTCCCA	q-RT-PCR
Vcam1	CTTGGGAGCCTCAACGGTAC	CCACGTGGATACTTCGTTCC	q-RT-PCR
Igf2	TGAAGAGACCCTGAGAGGGAGTTT	TGTCATTCGAGCAGGTGCCTTACT	q-PCR
Igf2r	AGAGGATTCCGCAAAGGAAGGGTT	ATCTTCACCCTAGCGCTGAATCTC	q-PCR
MycN	TTGACGCTCCAGGATGTTGTGGTT	AAGATCAAGAGCGAGGCTTCTCCA	q-PCR
Grb10	TCGTTTAGGAGCTAGTTCGCTGTG	AACACGCGCCAACATACGTGTTAC	q-PCR
Gapdh	CTCTGCTCCTCCCTGTTCC	TCCCTAGACCCGTACAGTGC	q-PCR
Maoa	AAGTCCCATGCAGGCAGTAGTA	ATCACTGAGGCTCTGAGGGATGTT	q-PCR

Eftud1	TGCCTTCTTCCACTTTCTCCCA	ACATGTGGACAGAGAGACCCAGA	q-PCR
Kctd16	AAATTCCTAAACTGCTGGCAGGGC	CACACTCTAACAGAAAGAGGAAGGC	q-PCR
Mid1	ACAGTCATCCTTCTGCCTCA	GGGTGACTGTGAGTGATT	q-PCR
PPARa	TTTAGATTGGAGGAAGTTGAAGTAGAT	CCCTACCCTAAATCCAACATATATACA	Bisulfite

CHAPTER II

Preface

Chapter II is a manuscript in preparation:

Genetic and epigenetic variation in murine sperm contributes to offspring metabolism

Jeremy M. Shea, Benjamin R. Carone, Hennady P. Shulha, Michael Ziller, Alper Kucukural, Marcus Vallaster, Hongcang Gu, Andrew R. Tapper, Paul D. Gardner, Alexander Meissner, Manuel Garber, and Oliver J. Rando

Contributions:

Jeremy Shea and Benjamin Carone performed IVF experiments. Jeremy Shea performed gene expression analysis. Michael Ziller and Hongcang Gu performed RRBS and WGBS. Jeremy Shea performed bisulfite pyrosequencing and copy number analysis. Bioinformatic analysis performed by Hennady Shulha, Alper Kucukural, and Manuel Garber. Jeremy Shea and Oliver Rando designed the experiments and analyzed the data. Jeremy Shea wrote the manuscript with editorial input from Oliver Rando.

CHAPTER II

Genetic and epigenetic variation in murine sperm

contributes to offspring metabolism

Abstract

Paternal environmental conditions impact offspring metabolic phenotypes. Although many intergenerational inheritance paradigms exhibit altered metabolism, there is a paucity of mechanistic details about the carrier of this information. In trying to identify this carrier, we performed *in vitro* fertilization experiments that localized the information to sperm, a major step in elucidating the signaling cascade that informs offspring of paternal environment. Subsequently, we characterized dietary effects on sperm cytosine methylation by whole-genome bisulfite sequencing (WGBS) on pooled samples, as well as reduced-representation bisulfite sequencing (RRBS) on 61 additional individual sperm samples. These experiments revealed that "epivariation" made far greater contributions to shaping the sperm methylome than dietary effects. This epivariation was prevalent over tandem repeats and linked with copy number variation of ribosomal DNA (rDNA). Genetic and epigenetic variability of the rDNA locus in sperm correlated with offspring cholesterol metabolism, revealing potential contributions of epivariability to inherited phenotypic variation. These results demonstrate that sperm carry paternal dietary information utilizing a

mechanism independent of genetic variation and cytosine methylation variation, which also sets offspring metabolism.

Introduction

Examples of intergenerational epigenetic inheritance have been discovered in a diverse set of organisms, including plants, water fleas, worms, and mammals (Heard and Martienssen, 2014; Lim and Brunet, 2013; Rando, 2012). In numerous cases, changes to parental environments alter offspring phenotypes. Recently, and with relevance to human health, multiple groups discovered that paternal dietary changes in rodents influence offspring phenotypes (Anderson et al., 2006; Carone et al., 2010; Jimenez-Chillaron et al., 2009; Ng et al., 2010). These dietary-induced transgenerational phenotypes are metabolic in nature, with offspring lipid and cholesterol changes, insulin resistance, and cardiovascular deficiencies being induced by paternal diet. Additionally, epidemiological studies in human populations conclude that the grandparent's environmental conditions affect their grandchildren's likelihood to develop metabolic diseases (Pembrey et al., 2006). Therefore, parental exposures contribute to variations in offspring phenotypes, including ones with medical implications.

Previously, we discovered that fathers fed low protein diets sired offspring with altered hepatic metabolism (Carone et al., 2010). The expression of cholesterol and lipid metabolism genes increased in livers of offspring from low protein fathers. These offspring also had altered levels of cholesterol metabolites, uniting gene expression and metabolite changes. It's been known for many years that cholesterol dysfunction is associated with cardiovascular disease and diabetes (O'Donnell and Elosua, 2008), and another published report on mice recently found that older offspring from low protein fathers develop cardiovascular disease and diabetes (Watkins and Sinclair, 2014). Therefore, discovering the mechanisms responsible for transmitting this epigenetic information is crucial for human health and disease.

Therefore, we have undertaken a research program aimed at discovering the mechanisms behind the inheritance of paternal dietary information. This program includes the establishment of an *in vitro* fertilization (IVF) experimental paradigm to uncover the carrier of paternal dietary information. In addition to testing for the carrier of paternal dietary information, this experimental paradigm allows us to test direct relationships between changes in sperm, and changes in offspring phenotypes. Another aspect of this research program includes the characterization of transmissible epigenetic information, for which we have extensively characterized the sperm DNA methylome. This research increases our knowledge of the inheritance of environmentally induced phenotypes.

Results

The Carrier of Paternal Dietary Information

Even with the increasing number of environmentally induced transgenerational phenomena, a paucity of mechanistic details exist for the transmission of this information. Possible carriers of environmentally-induced information in males include sperm, seminal fluid, and other cryptic messengers (Rando, 2012). Determining the carrier of this information allows us to focus downstream investigations on changes inherent to that carrier. We performed murine IVF experiments to test the hypothesis that sperm carries dietary information to offspring (Figure S2.1). Since early embryo culture influences later phenotypes (Feuer et al., 2014), we controlled for potential biases caused by parental age, media effects, and mechanical manipulation to minimize variation from these confounding factors. In addition to our previous Low Protein (LP; 10% instead of 19% protein) treatment, we expanded our dietary repertoire to include High Fat (HF; 60% instead of 20% fat calories) and Caloric Restriction (CR; 60% of daily *ad libitum* consumption) interventions as well. Male mice were placed on one of the three interventions at weaning, while siblings were given control diet. These males were maintained on these diets until use in IVF experiments. After reaching sexual maturity, sperm from matched males was used to fertilize
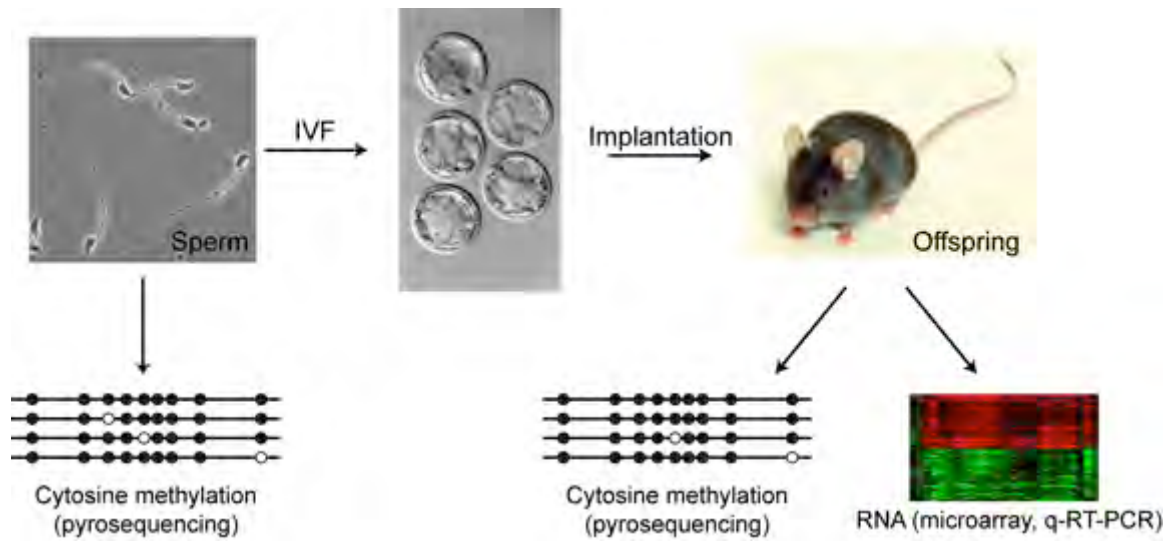
Figure S2.1
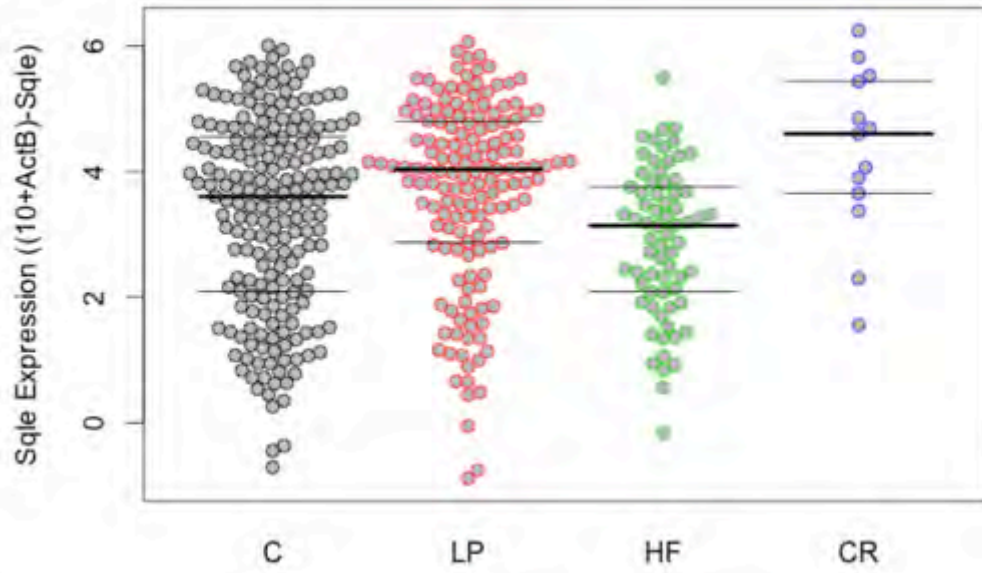
In vitro Fertilization (IVF) Paradigm

Sperm isolated from males is used to fertilize oocytes from superovulated females. Embryos develop *in vitro*, and are implanted into pseudopregnant females. Females give birth to the offspring, and the offspring are analyzed at weaning for gene expression or epigenomic differences. The epigenome of sperm is also analyzed.

oocytes derived from a single litter of females. We developed embryos *in vitro* until the 2-cell stage, at which point we transferred them to recipient females. Offspring derived from oviduct transfer of IVF embryos were analyzed for transcriptional changes at weaning, similar to our original investigations (Carone et al., 2010).

We analyzed hepatic gene expression from our IVF-derived offspring to determine if they have altered regulation of cholesterol and lipid biosynthetic pathways. Utilizing Squalene Epoxidase (*Sqle*) as a representative gene for the cholesterol pathway, as it controls flux in cholesterol synthesis (Gill et al., 2011), we found IVF-derived Low Protein offspring had increased expression of *Sqle* when compared to Control offspring in a large cohort of animals (Figure 2.1A). Performing microarray analysis of a subset of samples, we found that the cholesterol biosynthetic pathway was coherently increased in Low Protein derived offspring livers (Figure 2.1B). These results clearly indicate that sperm carries the epigenetic information induced by paternal Low Protein diet. However, the possibility remains that there may be additional carriers of epigenetic information since the effect we see on *Sqle* is reduced in IVF offspring when compared with natural matings; but this may be the result of genetic background, as we necessarily used FVB/NJ mice in IVF experiments, while our original findings derived from C57BL/6 mice.

A

Offspring Liver Sqle Expression

B

IVF Offspring   IVF Average   Natural Average

Acas2
Hmgcs1
Pmvk
Sc4mol
Dhcr7
Idi1
Cyp51
Lss
Nsdhl
Sqle
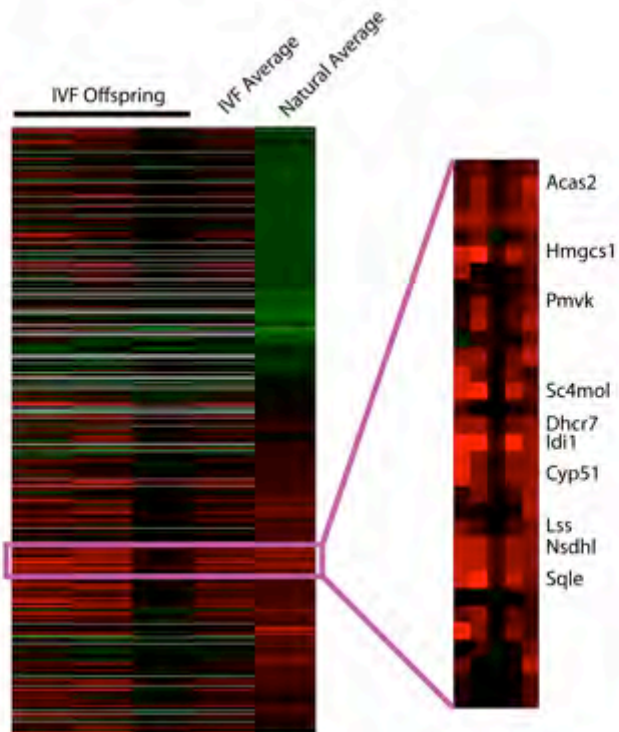
Figure 2.1

IVF Recapitulates Paternal Dietary Effects

A. qPCR analysis of offspring hepatic Sqle expression. Offspring from fathers on different diets are represented on the X-axis: C (Control), LP (Low Protein), HF (High Fat), and CR (Caloric Restriction). Each point represents the expression of Sqle from an individual offspring liver. Quartiles represented by thin black lines, while median is represented by thick black line. Expression was normalized to β-Actin. Nonparametric relative contrast effects shows significant differences (p<0.05) for C:LP (p=0.03), C:HF (p=0.006), and HF:LP (4×10^{-8}) comparisons. Similar significance values were found with T-Tests.

B. Affymetrix microarray analysis of IVF offspring livers. Rightmost column represents average gene expression changes from Carone *et al* (Carone et al., 2010). IVF average is the average of the three leftmost columns. Zoomed in region represents node containing several genes in the cholesterol biosynthesis pathway.

The additional paternal dietary interventions also caused metabolic gene expression changes in offspring (Figure 2.1A). Most interestingly, the direction and magnitude of the *Sqle* change mirrored the form and intensity of the dietary manipulation. Although the dietary interventions are not directly comparable to one another, the resulting differences in the father's weight point to High Fat diet representing a major increase in energy intake, while Caloric Restriction significantly decreases energy abundance (not shown). Low Protein diet causes a nominal decrease in paternal body weight. High Fat diet had the opposite effect on offspring *Sqle* expression as compared with changes caused by Low Protein diet, while Caloric Restriction further increased the average *Sqle* expression when compared to Low Protein diet (Figure 2.1A). These results demonstrate that fathers are likely sending information about energy abundance to offspring, but should be directly tested in the future (Solon-Biet et al., 2014).

Sperm DNA Methylomes

Since sperm carry epigenetic information about a father's diet, we could narrow down the mechanisms utilized to transfer this information. So, we set out to characterize the epigenomic changes in sperm induced by diet. For this purpose, we focused on sperm cytosine methylation, since relevant research has implicated it in control of epigenetic phenomena, such as imprinting and epialleles (Chong et al., 2007; Li et al., 1993). We characterized the sperm methylome by RRBS and WGBS, as complementary methods to understand the

97

population structures and dietary influences on sperm (Gifford et al., 2013; Meissner et al., 2008). We used RRBS to profile a subset of the genome at base pair resolution in numerous samples, while WGBS complemented this approach by characterizing the entire methylome for matched pools of sperm samples.

Previously, we had observed large amounts of variation between control animal methylomes using low resolution MeDIP-Seq (Carone et al., 2010), which we wanted to follow up on to determine the factors shaping sperm methylomes. We analyzed the sperm methylomes of 61 mice on various diets using RRBS to gain a better understanding of the population structures of sperm methylation. RRBS characterizes the methylome at base pair resolution for ~4% of CpGs across the mouse genome, with an enrichment of CpGs located in GC-rich regulatory elements. The RRBS dataset included sperm samples primarily from Control, Low Protein, and High Fat diets (Table S2.1). CpGs captured in our RRBS dataset had a bimodal distribution of cytosine methylation (Figure S2.2A) that was strongly anti-correlated with CpG density (Figure S2.2B,C). The relationship between CpG density and cytosine methylation was skewed at repetitive elements, as even CpG-rich retroelements remained highly methylated (Figure S2.2B,D). Other groups have reported these hallmarks of sperm methylomes (Molaro et al., 2011), signifying that we have generated high quality data for making demographic inferences.

A

Number of CpGs (X10⁵)

All CpGs (n=1,041,729)
Non-repeat (n=622,516)
Repeat (n=419,213)

Methylation level

B

Number of CpGs (X10⁴)

HYPOmethylated, NO repeat
HYPERmethylated, NO repeat
HYPOmethylated, REPEAT
HYPERmethylated, REPEAT

CpG count (per 200 bp)

C

CpG count, 200 bp window

Methylation level

D

CpG count, 200 bp window

Repeat element
Non-repeat

E

Number of CpGs (X10⁵)

All sperm samples (n=63)
Control (n=30)
Low Protein (n=16)
High Fat (n=6)
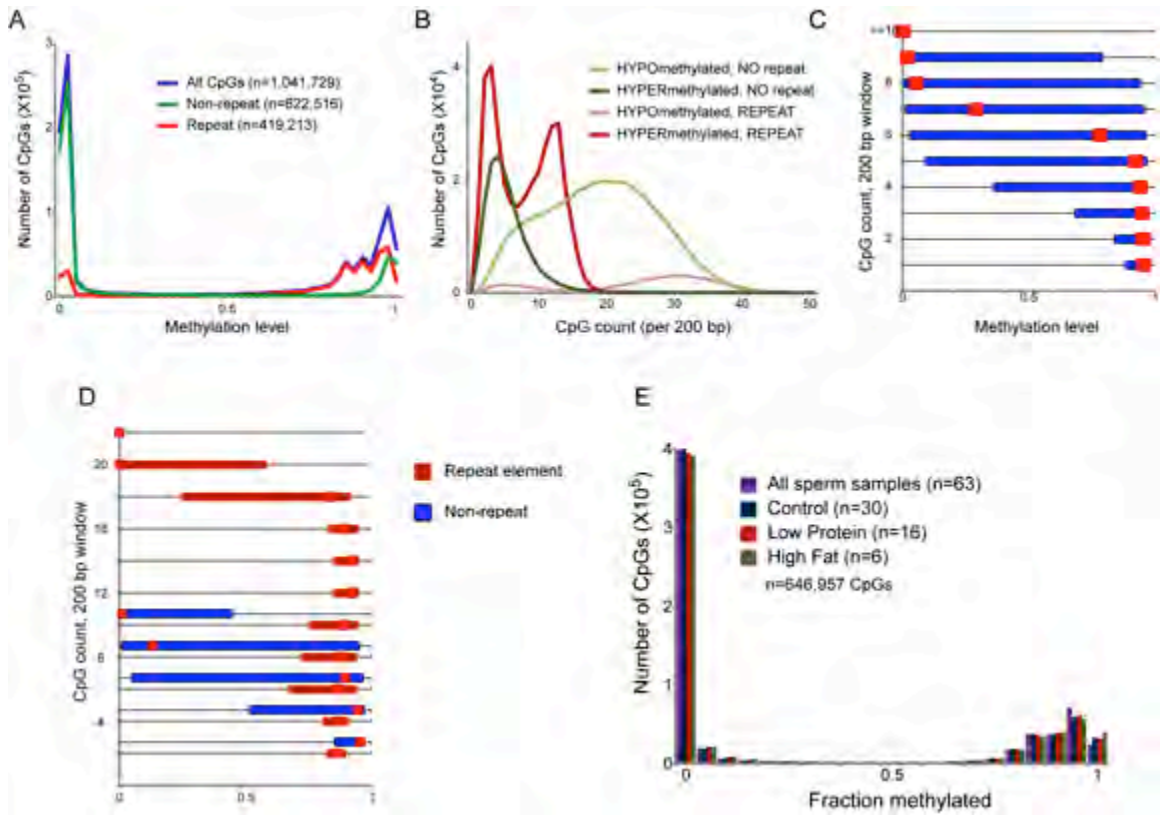n=646,957 CpGs

Fraction methylated

99

Figure S2.2

Overview of RRBS Dataset

A. Average methylation levels for 1,041,729 individual CpGs which were sequenced at least 10 times in at least one RRBS dataset. Y axis shows number of CpGs exhibiting a given methylation percentage, as indicated on the X axis. Methylation distributions are shown for all CpGs, and separately for CpGs located in annotated repeat elements and for non-repeat CpGs.

B. Relationship between CpG methylation and local CpG density. CpGs were classified as hypomethylated (<20% methylated) or hypermethylated (>80% methylated), and were separated into those located in repeats and those outside repeats. Number of CpGs in each class is shown (Y axis) for CpGs according to their local CpG density (# CpGs in the surrounding 200 bp).

C. Boxplots showing the range of methylation values for all CpGs located in various local CpG density regions. Red box shows median, blue bar shows quartiles, and whiskers show max and min.

D. As in (C), but with CpGs located within and outside of repeat elements plotted separately.

E. Methylation distributions are shown as in A. for CpGs averaged across all animals, or across groups of males consuming the indicated diets. Here, only CpGs that were sequenced at least 10 times in at least half of the RRBS datasets (646,957 CpGs) are shown.

Diet does not alter the Global Methylome

Sperm methylomes were very similar with all pairs having a correlation coefficient greater than 0.90, and an average coefficient of 0.967 (Figure S2.3A). Hierarchical clustering of the sperm methylomes did not group them by diet, which indicates that diet is not the primary driving force for global changes in the sperm methylome (Figure 2.2A). Notably, the correlations between pairs of Control or Low Protein animals were identical to those among Control and Low Protein animals (Figure 2.2B and S2.3B). These observations indicate that changing diets did not have a discernable effect on the global methylomes of these mice. Therefore, diet plays a secondary role for influencing the relationship of sperm methylomes.

Epivariability of the Sperm Methylome

Strikingly, pairwise comparisons of samples revealed that sibling sperm methylomes were the most related in our inbred mice (Figure 2.2B). We previously noticed this increased similarity between siblings when performing MeDIP-seq (Carone et al., 2010). These results indicate that distantly related inbred mice contain an inherent epivariability in their sperm methylomes. Also, these results suggest that the shared environment of siblings prior to separation largely determines sperm methylomes. The period prior to separation may also include the parental generation, as epiallelic traits controlled by cytosine

Figure S2.3

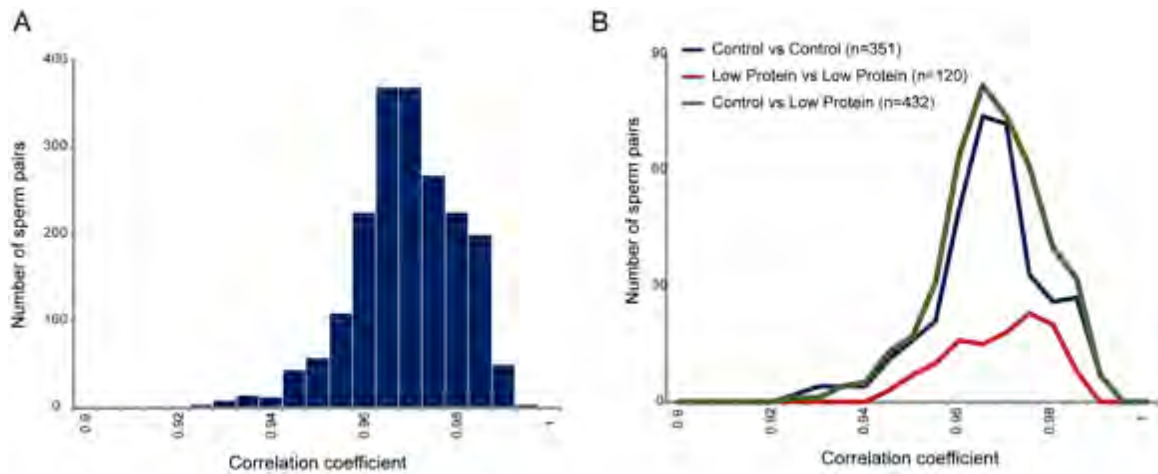Distributions of correlations between individual sperm samples

A. Correlation coefficients were calculated between all pairs of RRBS datasets, and distribution of correlation coefficients is plotted as a histogram.

B. As in (A), but with comparisons shown between pairs of sperm samples isolated from animals on Control diet, between animals on Low Protein diet, or between pairs of animals on Control vs. Low Protein.
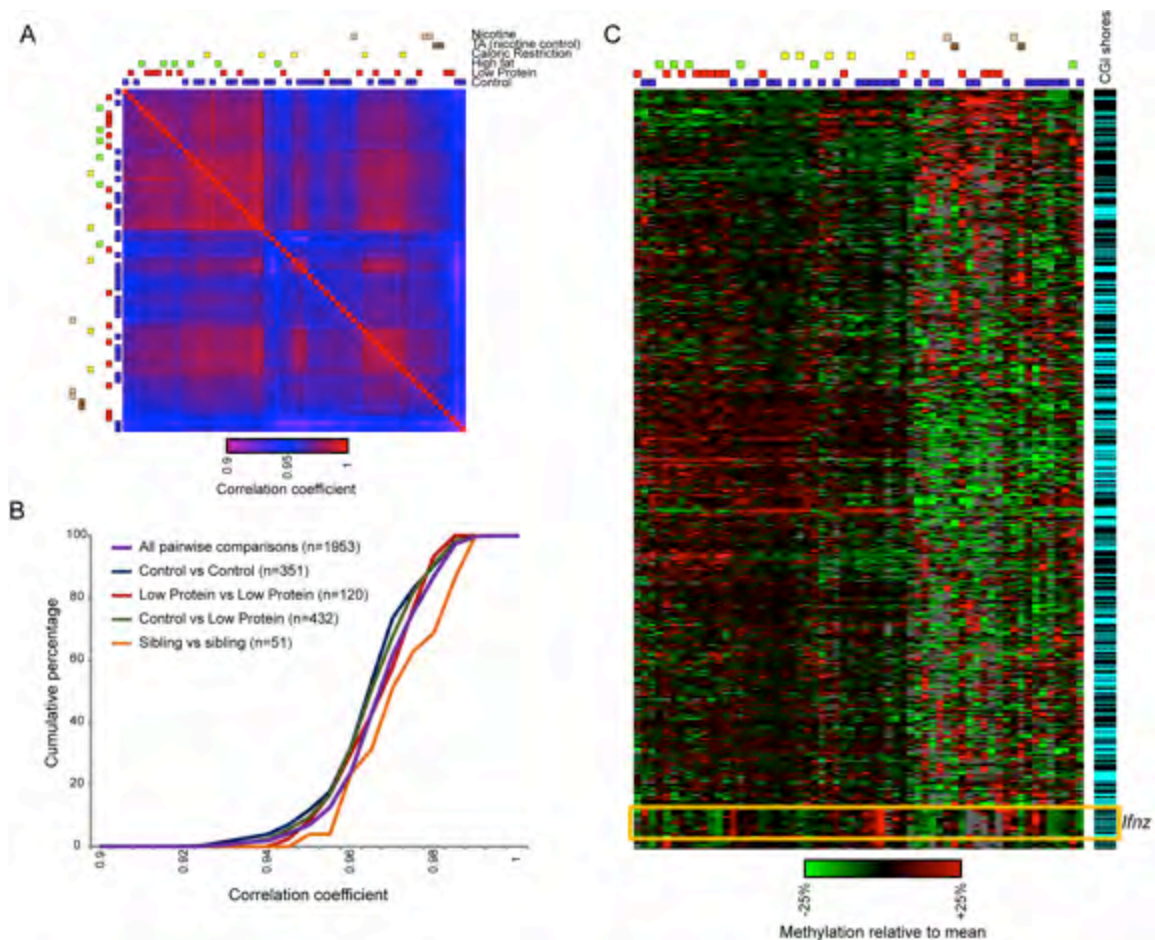
Figure 2.2

Epivariation among sperm methylomes

A. Correlations between 61 individual RRBS libraries. The environmental conditions for each male are indicated as colored boxes. Data include animals subject to Control, Low Protein, High Fat, Caloric Restriction, nicotine (administered 200 mg/mL free base in saccharine-sweetened drinking water), and the matched tartaric acid only control.

B. Cumulative distribution plots for inter-sample correlations, for the indicated animal pairs.

C. Heatmap showing 748 regulatory elements (promoters, distal CpG islands, and CpG island shores) as rows, with all CpGs within each element averaged. Heatmap shows zero-centered data, grouped both by animal and by regulatory element. Right panel shows that the majority of epivariable elements are CpG island shores.

methylation and changes in cytosine methylation caused by endocrine disrupters can persist for several generations (Anway et al., 2005; Morgan et al., 1999). The environmental factors that guide the demographics of the samples remain unknown at this point, since we found no correlation between methylome differences and the number of littermates, time of year, etc. These results demonstrate that diet-induced methylation changes take place on an "epivariable" methylome guided by shared histories, indicating that "epivariability" might accumulate over generations in a fashion similar to *Arabidopsis thaliana* (Schmitz et al., 2013). Therefore, many of the phenotypes captured in intergenerational experiments involving diet exist on top of an epivariable background - even for inbred mouse strains.

We sought to limit hypotheses for the mechanistic basis of this epivariability in mice by uncovering shared features of epivariable regions. We calculated the methylation levels for individual CpGs with high confidence, and used the 3,396 most variable CpGs to search for enriched characteristics (Figure S2.4A). We found an abundance of "epivariable" CpGs positioned distal to transcription start sites, at an average distance of ~1.5kb away from the transcription start site (TSS) (Figure S2.4B). Many of these distal epivariable CpGs were located in CpG island shores, which have been shown to harbor variably methylated CpGs between cells (Irizarry et al., 2009). When we averaged methylation scores for
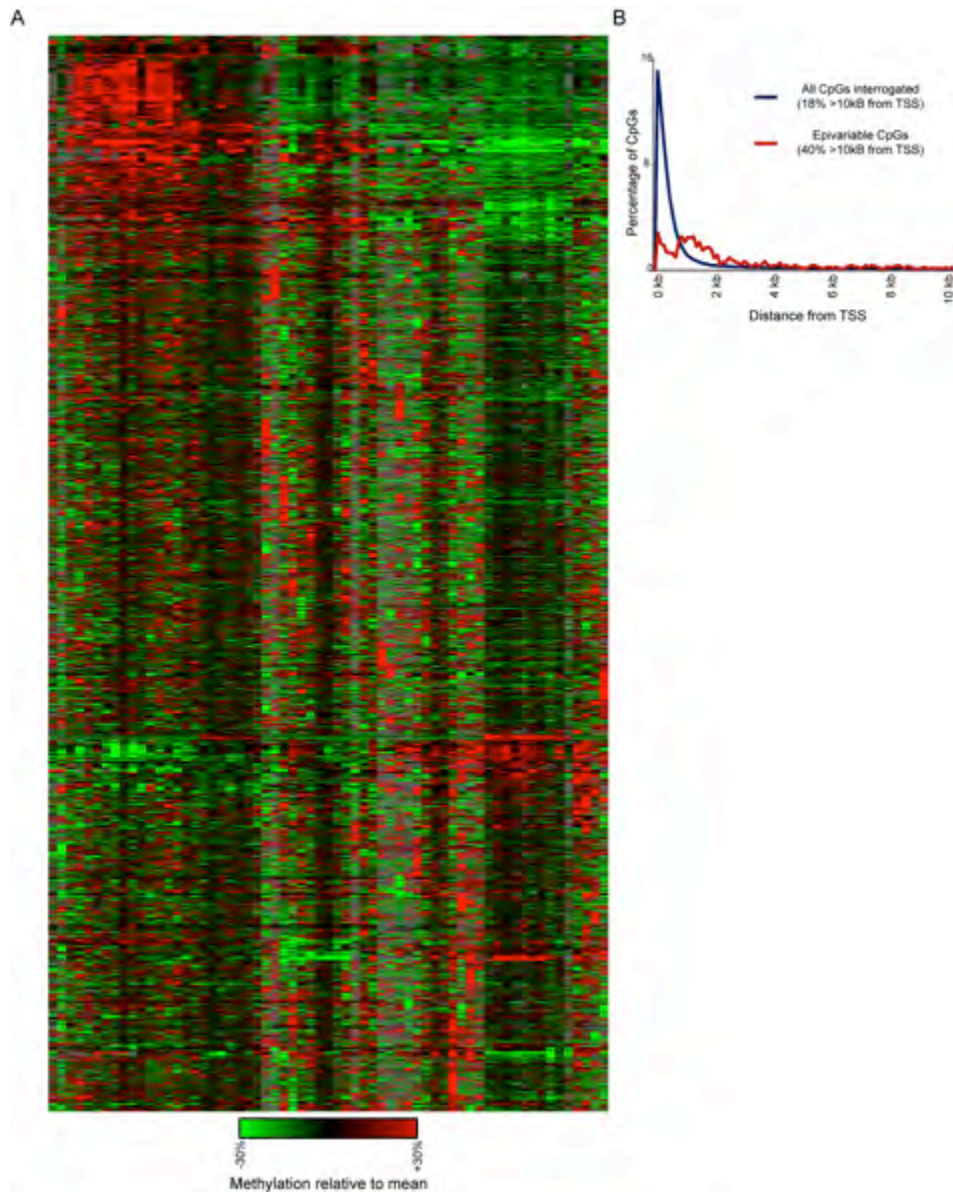
Figure S2.4

Epivariable CpGs

A. Heatmap of epivariation for 3,396 highly variable individual CpGs. CpGs were selected from non-repeat regions, and sequenced at least 10 times in 80% or more of sperm samples. They deviated from the mean methylation value of that locus by at least 20% in at least 6 individual sperm samples. Data for each CpG are normalized to mean methylation across all animals.

B. Distance from epivariable CpGs to the nearest TSS, with distribution of total RRBS dataset shown for background.

various regulatory elements (promoters, CpG islands, and CpG island shores), we recapitulated the enrichment of epivariable regions in CpG island shores (Figure 2.2C). Thus, the labile regions of the sperm methylome are located at the boundaries of regulatory elements, which may point to differential usage of these loci between animals resulting in phenotypic variability. This is, to our knowledge, the first study to determine the population epivariability of identical cell types between inbred animals.

Search for Diet-Induced Differences in Sperm Methylation

The RRBS results denoted a large amount of "epivariability" present in the sperm methylome, which masked possible signals from dietary manipulation. Although background epivariability is high, we wanted to further characterize the sperm methylome with the intention of discovering diet-induced differences. For this purpose, we performed Whole Genome Bisulfite Sequencing (WGBS) to characterize the entire methylome at base pair resolution. To guard against the inherent variability of sperm methylomes, we utilized a matched pooling strategy that should be robust against the individual variation uncovered by our RRBS data. We generated pooled WGBS libraries from equal amounts of sperm DNA from seven animals maintained on the same diet, and the matched pool from siblings maintained on another diet. These libraries were sequenced to an average depth of 1.3 billion reads, producing 47-fold mean coverage.
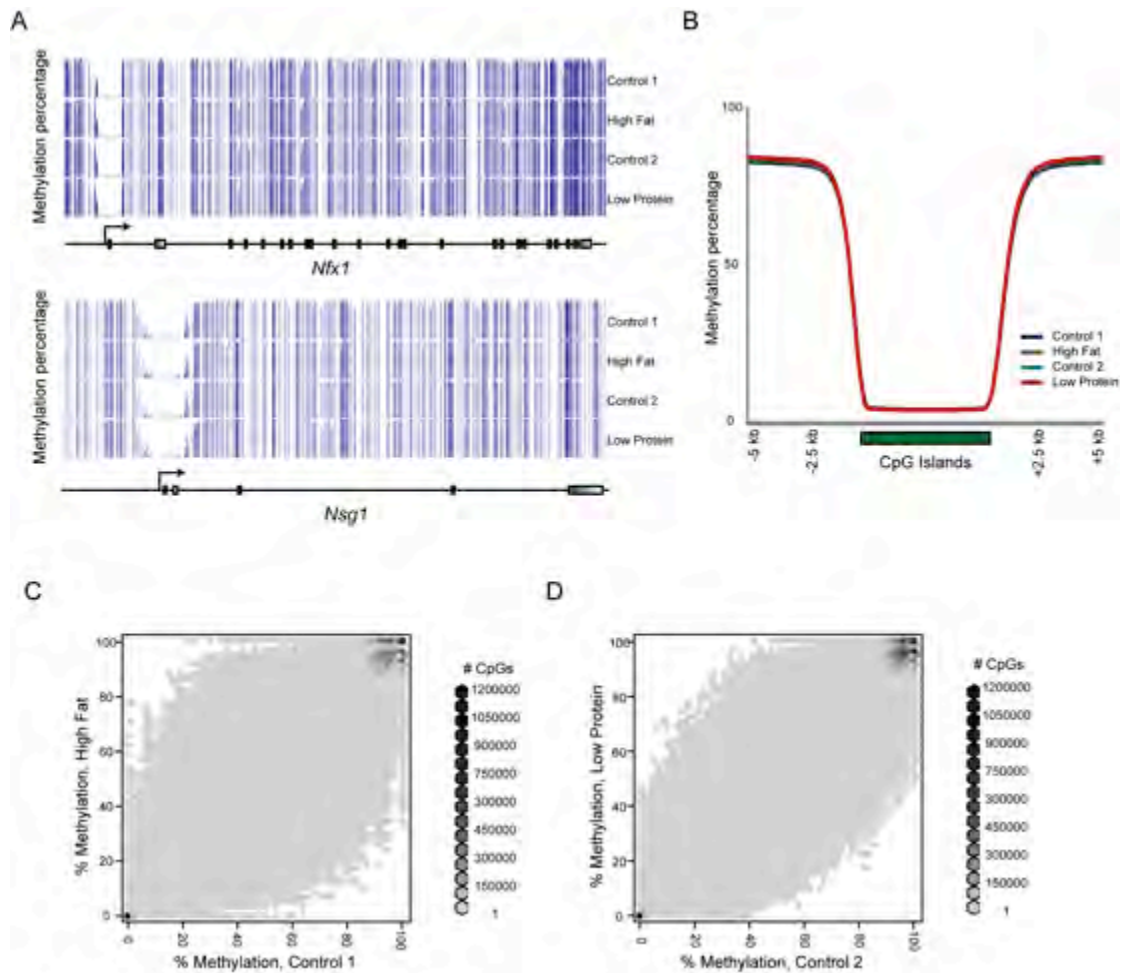
Figure 2.3

Whole genome cytosine methylation in murine sperm

A. Examples of typical methylation profiles. For the two genes shown, cytosine methylation data for each of the four libraries – Control 1, High Fat, Control 2, and Low Protein – are shown. Each vertical bar represents the methylation percentage for a single CpG. There is a general background of complete methylation, with hypomethylation occurring at CpG islands such as promoters. B. Average cytosine methylation for each of the four libraries plotted over CpG islands and surrounding DNA. CpG islands were length-normalized for this visualization.

C-D. Scatterplots for individual CpG methylation levels between the matched Control and High Fat pools (C) or between matched Control and Low Protein pools (D). Data are shown only for the 80% CpGs with the greatest read depth (n=16.1 and 16.6 million CpGs for C and D).

107

These high-resolution sperm methylome datasets confirm that sperm are highly methylated with regions of hypomethylation occurring over regulatory elements (Figure 2.3A). Sperm have an average methylation (average of all CpGs) of 68%, which is similar to hypermethylated somatic lineages (Ziller et al., 2013). Additionally, many regulatory elements contain CpG islands, and these islands maintain identical hypomethylated architectures in all four pools (Figure 2.3B). These WGBS pools demonstrate that sperm are hypermethylated with focal regions of hypomethylation over regulatory elements, similar to somatic cells.

As suggested by the DNA methylation landscape, most of the ~20 million individual CpGs were either fully methylated or fully unmethylated (Figure 2.3C and S2.5A-C). The CpGs that were fully methylated or unmethylated were unaffected by paternal diet (Figure 2.3C and S2.5A-C). However, some CpGs with intermediate methylation levels (20-80%) were enriched for differences between the dietary interventions. However, methylation differences at these individual CpGs rarely exceeded 20% absolute difference. Theoretically, these modest differences at individual CpGs cannot account for penetrant offspring phenotypes because of the "digital" nature of sperm. A single sperm fertilizes a single oocyte, so a CpG with 20% methylation merely means that 1 in 5 sperm are methylated at that CpG. A change from 20% to 40% methylation at a single CpG alters the frequency of sperm with the methylated CpG from 1 in 5 sperm to

Figure S2.5
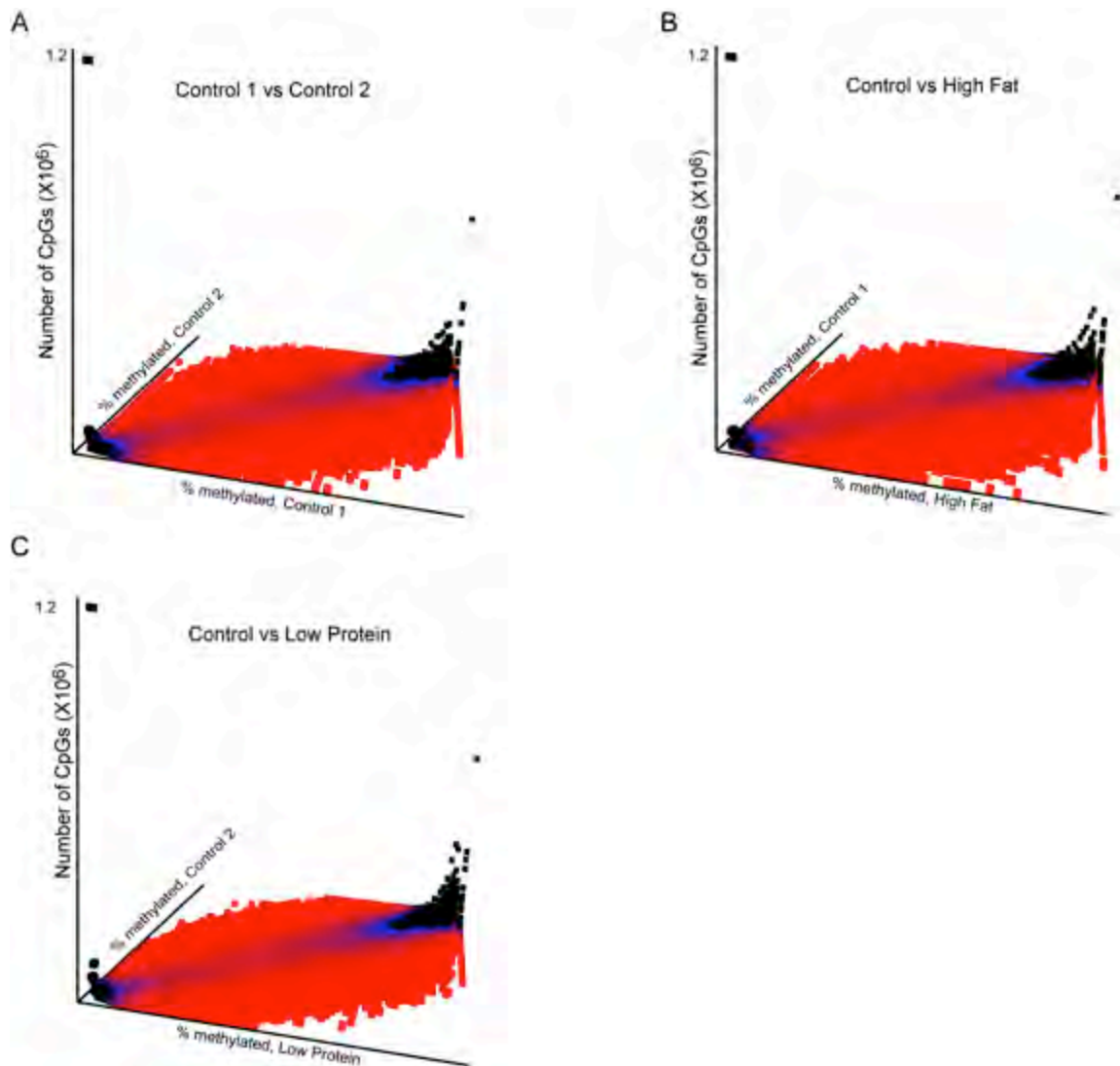
Scatterplots of methylation levels between WGBS pools

A-C. 3D plots showing cytosine methylation percentage for two WGBS pools (x and y axes, scale is from 0 to 100%), with z axis showing the number of individual CpGs. For each scatterplot only the 80% of genomic CpGs with the highest sequencing depth were used. The vast majority of individual CpGs are clustered around (0,0) and (100,100).

2 in 5 sperm. Therefore, these modest changes at individual CpGs cannot account for the penetrant nature of diet-induced phenotypes.

However, integration of larger regions of modest, yet consistent changes in methylation can theoretically result in penetrant phenotypes. The digital nature of cytosine methylation would then be able to exert continuous control of a quantitative phenotype. Therefore, we searched for diet-induced differences over 300 base pair windows that incorporated several adjacent CpGs, using read depth to calculate significance. We discovered hundreds of differentially methylated regions (DMRs) between our matched sperm methylomes (Table S2.2). Most of these DMRs were located in tandemly repeated regions, including both the 5S and 45S rDNA loci (Figure 2.4A-E). Other significantly altered tandem repeat DMRs included the interferon zeta (*Ifnz*) cluster (Figure 2.4F), *defensins*, *Skint* genes, and *Mrgpra/b* genes (Figure S2.6A-C). Many significant regions were shared between Low Protein and High Fat pools, and these regions were generally hypermethylated in Low Protein and hypomethylated in High Fat. Very few large DMRs were found outside of these tandem repeat families, with the few remaining DMRs located in CpG island shores (Figure S2.6D-E). Although repetitive elements have been implicated in transgenerational phenomena, we did not observe significant dietary effects on cytosine methylation over these regions (Table S2.3), including IAP elements that are responsible for epiallelic regulation (Lane et al., 2003). Interestingly, at most of
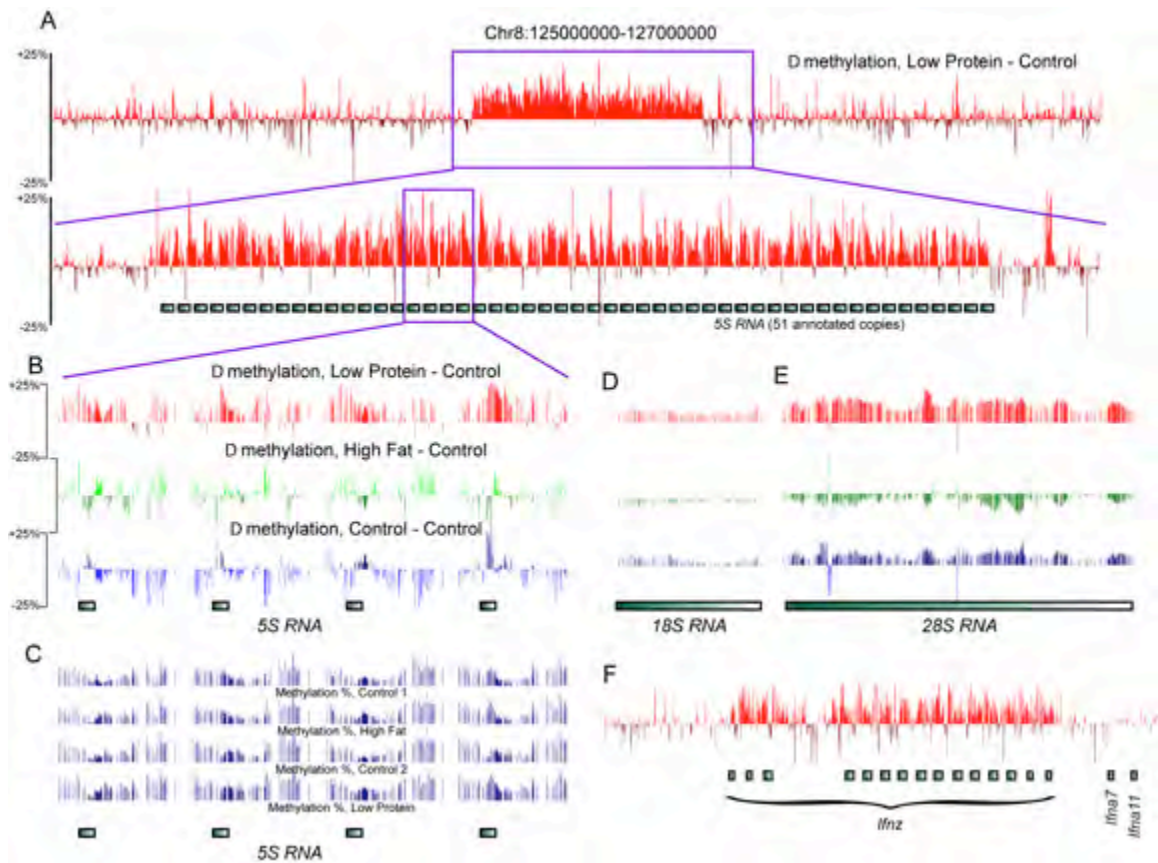
Figure 2.4

Methylation differences primarily occur over repeated gene families

A. Methylation changes over the 5S rDNA locus on chromosome 8, shown as Low Protein minus its matched Control pool – positive values indicate hypermethylated loci in Low Protein diet. Top panel shows 2 MB of chromosome 8 surrounding the 5S rDNA repeats, bottom panel is a zoom-in as indicated.

B. Zoom-in on 4 repeats of the 5S rDNA locus, showing Low Protein minus Control, High Fat minus Control, and Control 1 minus Control 2, as indicated. For all three images, the scale runs from -25% to +25%.

C. Absolute methylation levels (from 0 to 100%) for the loci shown in (B).

D-F. Both the 45S (18S and 28S) rDNA and *Ifnz* cluster are hypermethylated in the Low Protein pool.

111
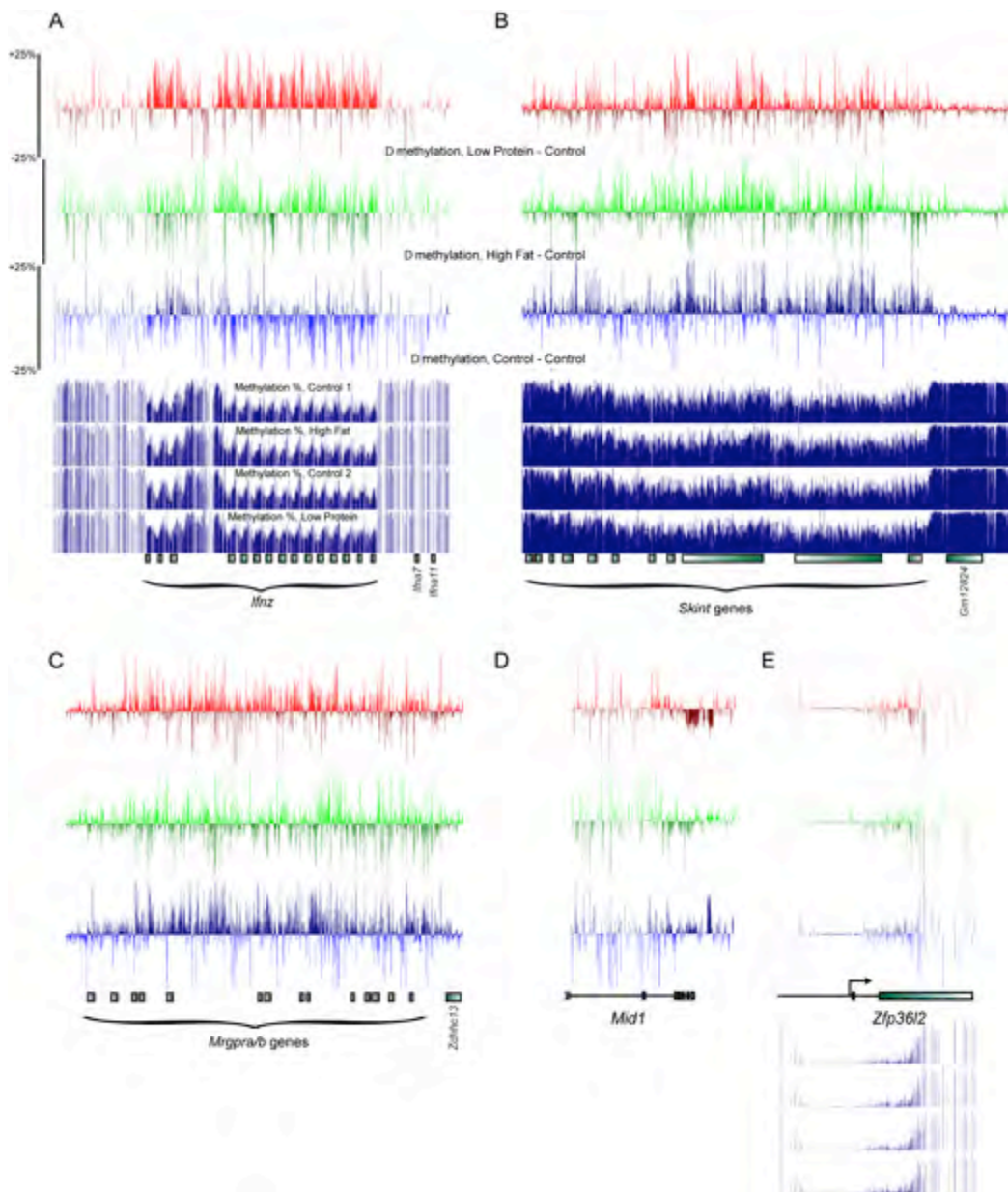
Figure S2.6

WGBS examples of diet-induced differential methylation

A-E. For each locus, data are shown for Low Protein minus Control 2 (top panel, red), High Fat minus Control 1 (middle panel, green), and Control 1 minus Control 2 (bottom panel, blue). Y axis is consistently -25% to +25%. For some loci, absolute methylation levels are also shown. Loci include multigene families *Ifnz* (A), *Skint* (B), and *Mgrpra/b* (C), the beginning of the X chromosome pseudoautosomal region (D), and an example of an epivariable CpG island shore (E).

the tandem repeat clusters reported as DMRs between diets, we also saw differences between the two control pools. So, these regions might be highly variable in general, which is a good indication of environmental lability.

Our discovery that a number of our WGBS DMRs differed between control pools (Figure S2.7) argues that these methylation changes might be driven by the epivariability discovered in our RRBS dataset. To address this issue, we returned to our RRBS datasets to decipher whether the WGBS pools reported *bona fide* dietary-induced changes or penetrant epivariability. Although RRBS covers only a subset of the genome, we found several CpGs located in the rDNA and *Ifnz* clusters represented in the RRBS dataset that were some of the most variable CpGs in the RRBS dataset (Figure 2.2C and S2.8). This indicates that our WGBS pools may have captured high epivariability as diet-induced changes at these loci.

Epivariation at Ribosomal DNA

The combined results from our RRBS and WGBS datasets suggest that epivariation dominates any changes elicited by paternal diet. The relative contributions of diet and epivariability to sperm methylation were subsequently determined at the 45S rDNA locus, one of the most significantly altered tandem repeat regions in both the RRBS and WGBS datasets. We investigated the methylation level at the 45S rDNA with bisulfite pyrosequencing to gain insight
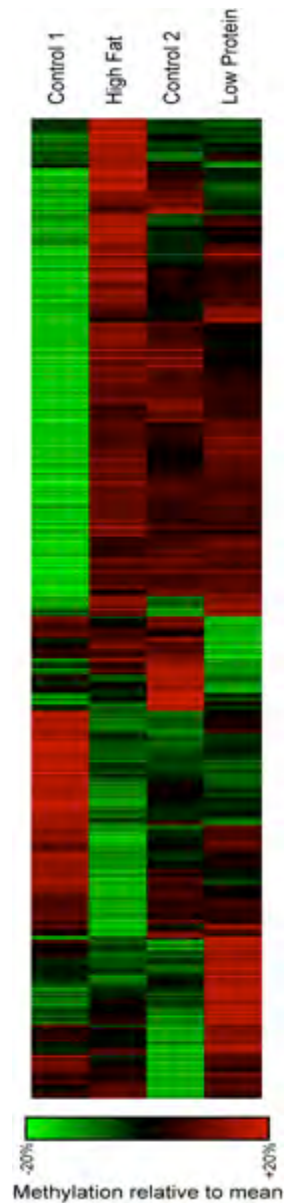
Figure S2.7

Differentially-methylated regions are generally epivariable

All 300 bp windows with significant (q value < 0.05) differences between Control 1 and High Fat, or Control 2 and Low Protein WGBS datasets, were selected. Data were zero centered, and the 2427 windows with at least one dataset exhibiting a 10% methylation difference were selected and clustered. Importantly, for the majority of windows exhibiting a difference between High Fat and Control 1, or between Low Protein and Control 2, Control 1 and Control 2 datasets also differed.
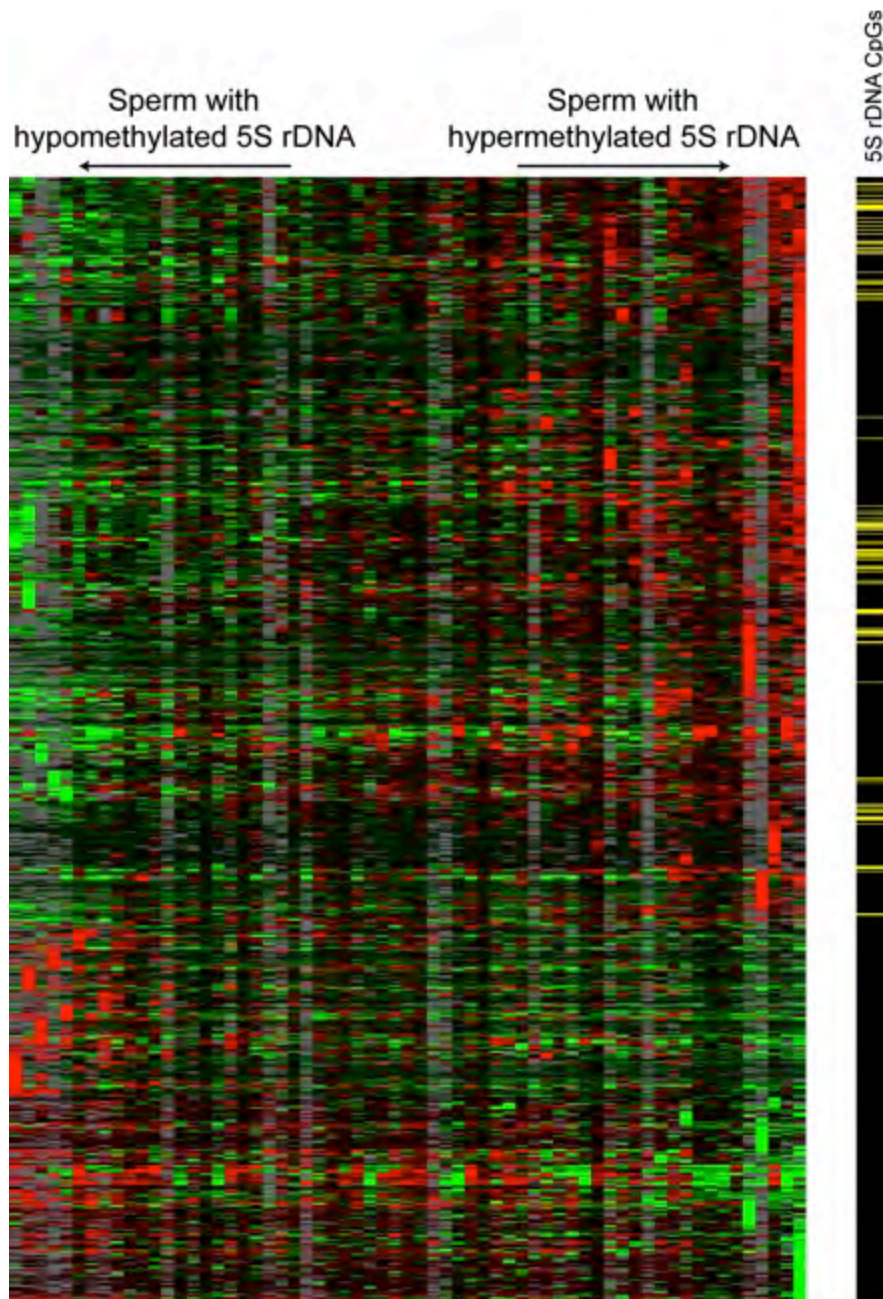
Figure S2.8

Epivariation at 5S rDNA locus

RRBS data are shown for all CpGs exhibiting a correlation/anticorrelation of over 0.3/-0.3 to the averaged RRBS data for the 5S rDNA cluster. Multiple CpGs that are unlinked to the rDNA cluster exhibit correlated methylation patterns, consistent with two loci responding either to the same genetic (rDNA copy number) or environmental (number of littermates) cue.

into the function of cytosine methylation over tandemly repeated regions. Extensive epivariability at the 45S rDNA locus was confirmed at two locations in 219 animals (Figure S2.9A-C). Technical replicates from different bisulfite conversions were highly correlated (Figure S2.9D) and the methylation status of whole testis and sperm from the same animals were very similar (Fig2.9E). Extensive cytosine methylation variation at the 45S rDNA is observed, with methylation levels at individual cytosines ranging from ~5% to ~50% between Control sperm samples.

Pyrosequencing of the 45S rDNA locus from matched littermates maintained on different diets revealed no consistent dietary effect on cytosine methylation at this locus (Figure 2.5A). While some discordance exists between siblings, none of the dietary interventions consistently altered methylation in a uniform direction. Again, the overall methylation levels vary greatly between animals, ranging from 5% to 50% average methylation, even among control animals (Figure 2.5A). However, 45S rDNA methylation was most similar among siblings (Figure 2.5B) - like many other components of the sperm methylome, indicating that 45S rDNA methylation at this locus is either inherited or established early in development.

Inheritance of Epivariation

Our IVF paradigm allows us to directly test if the epigenetic differences in sperm influence offspring phenotypes. First, we made use of our IVF paradigm to
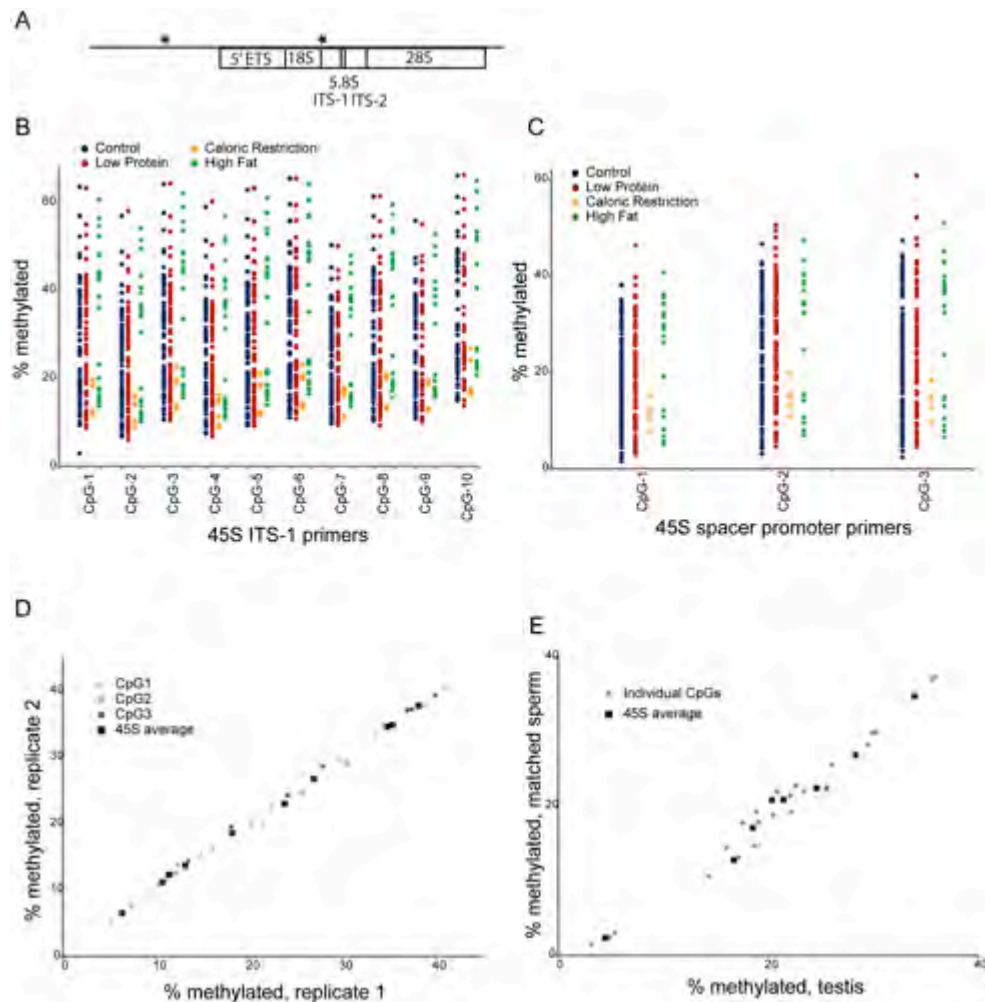
Figure S2.9

Epivariation at the 45S rDNA locus

A. Schematic of murine 45S repeat, with primer pairs used for pyrosequencing indicated as stars.

B-C. Methylation levels over 10 CpGs at the 45S ITS-1 sequence (B), and 3 CpGs at the over the 45S spacer promoter (C). For each sperm sample, genomic DNA was bisulfite converted and amplified with bisulfite-appropriate primers. Methylation levels were quantitated via pyrosequencing. At each CpG, samples are separated according to the four dietary regimes.

D. Replicate bisulfite conversions were pyrosequenced, and methylation levels for replicates are scatterplotted for individual CpGs and for the 45S promoter average.

E. 45S promoter methylation is highly correlated between sperm and matched testis samples.

A

B

C

D
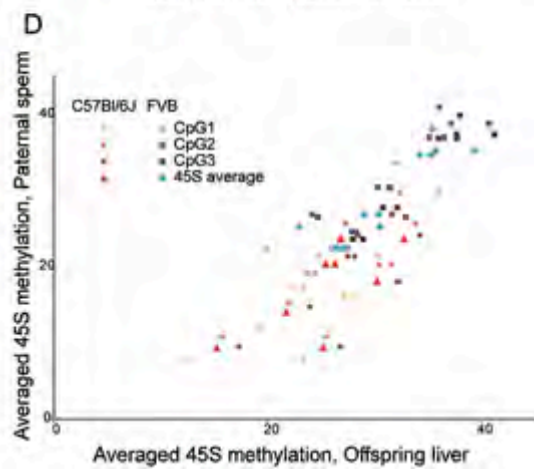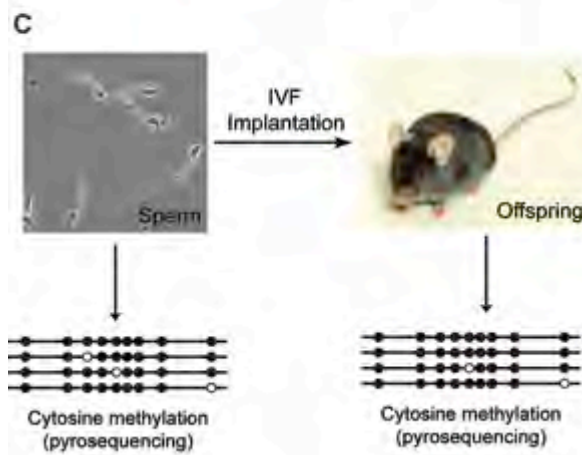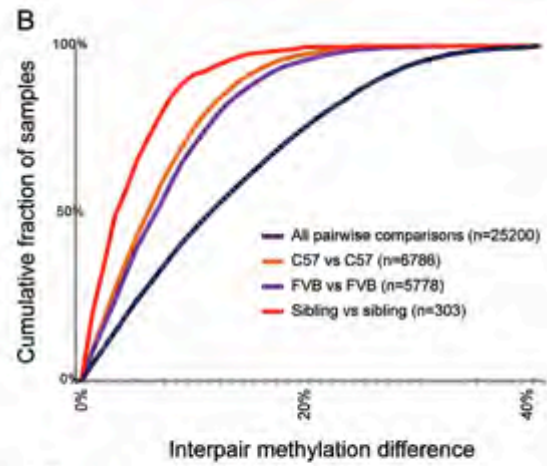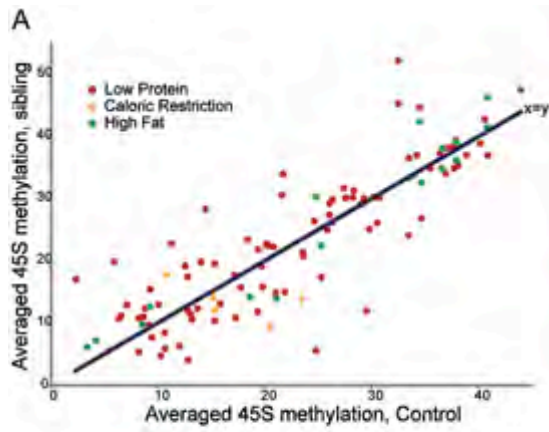
Figure 2.5

Heritability of 45S rDNA Epivariation

A. Average methylation levels over the 45S rDNA spacer promoter are scatterplotted for siblings split to Control diet (X axis) and either Low Protein, Caloric Restriction, or High Fat diet (Y axis). Each point shows the average methylation level for the 3 CpGs interrogated by pyrosequencing. While many siblings exhibit discordant methylation levels, none of the dietary interventions consistently altered methylation in a uniform direction.

B. Differences in 45S rDNA methylation between all pairs of animals, between pairs of the same strain background, or only between siblings. Shown as cumulative distribution plots.

C. Schematic of system used to link the paternal sperm epigenome to offspring phenotype. For a given sperm sample, 5% was used to generate offspring via IVF and surgical implantation. Ninety-five percent of the sperm sample was used for analysis of cytosine methylation, and methylation and mRNA abundance data from matched offspring were obtained. Importantly, the very same sample used to generate offspring was also used for molecular analysis.

D. Heritability of rDNA methylation patterns. 45S promoter methylation was analyzed by pyrosequencing for 11 sperm samples and 20 matched offspring livers. Data are shown for individual CpGs (circles) as well as averaged for the 3 CpGs.

determine if rDNA methylation is inherited from the father (Figure 2.5C). We found that cytosine methylation at the 45S rDNA locus was correlated between paternal sperm and matched offspring livers, indicating that methylation of 45S rDNA is inherited (Figure 2.5D). Since sperm and liver are separated by numerous cell fate transitions, this reveals stability of inherited rDNA methylation throughout development. Analysis of early embryonic methylome profiling by others reveal that rDNA maintains cytosine methylation at these stages (Radford et al., 2014; Smith et al., 2012). On the other hand, tandem repeats are subjected to relatively rapid copy number changes (She et al., 2008), which could signal that our epigenetic measurements are capturing genetic changes.

Inheritance of Copy Number Variation

Considering the relatively rapid copy variation of tandem repeat genes, we determined if the heritability of epivariable rDNA methylation was related to rDNA copy number in the sperm of mice. We utilized qPCR (and digital droplet PCR (see (Hindson et al., 2011)) to determine the copy number of the 45S rDNA locus, finding that the locus was hypervariable, with calculated copy number ranging from ~100 copies to ~300 copies per sperm; consistent with previous reports for murine rDNA content (Veiko et al., 2007). Diet did not cause these changes in rDNA copy number; but like rDNA methylation, copy number was similar among siblings (Figure 2.6A). Interestingly, the genetic background of the mice influenced the methylation status of the 45S rDNA (Figure 2.6B) – another
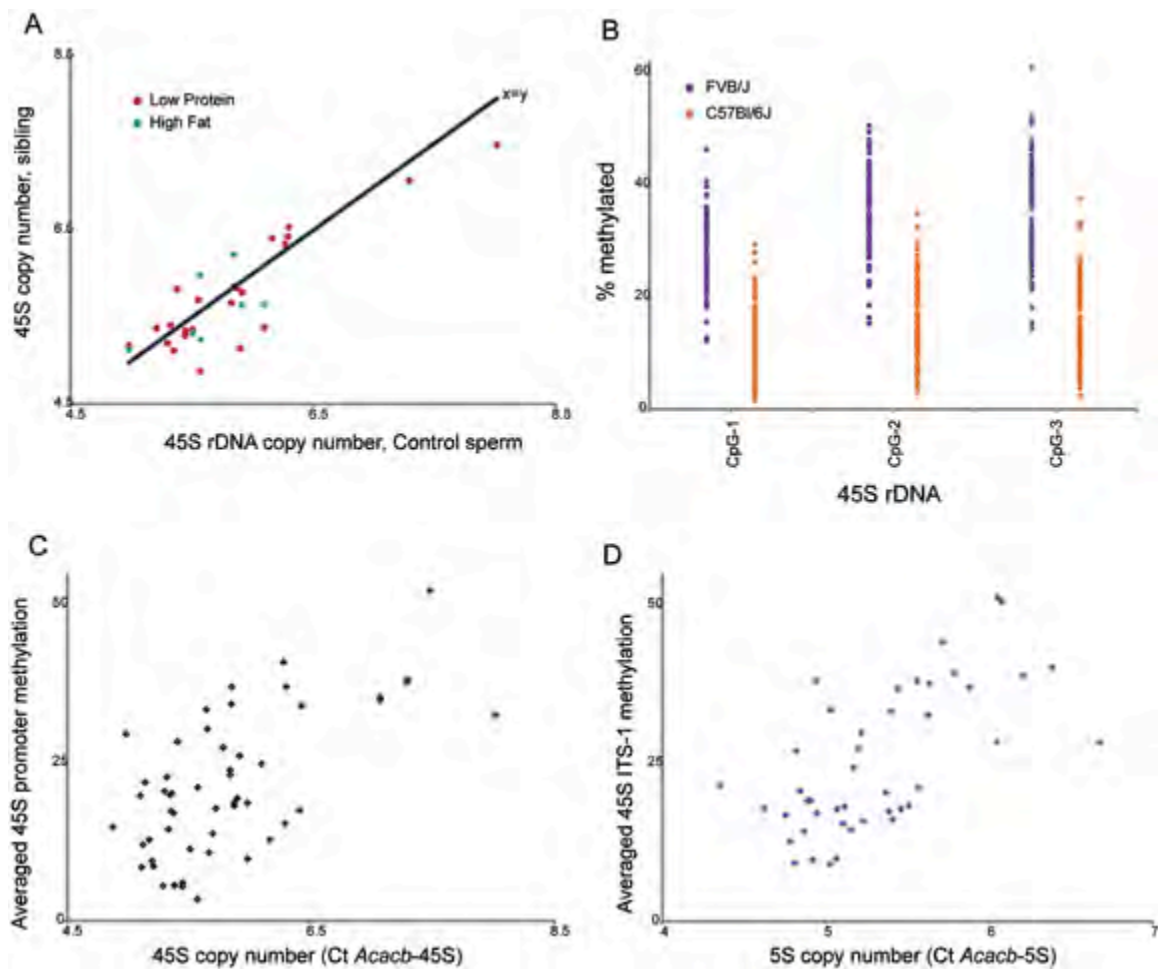
Figure 2.6

Genetic Modifiers of rDNA Epivariation

A. Comparison of rDNA copy number in siblings split to different diets. rDNA copy number was assayed by q-PCR, and relative copy number is shown – copy number here is given as 45S q-PCR Ct cycle subtracted from Ct cycle for q-PCR against the single copy *Acacb* gene. Similar results were obtained using a multi-copy tRNA locus for normalization (not shown).

B. Inbred mouse strains exhibit differential rDNA methylation. Animals separated by strain background as indicated.

C-D. rDNA copy number is correlated with rDNA methylation. 45S and 5S rDNA copy number were quantitated for sperm samples as in (A), and 45S or 5S copy number were scatterplotted against averaged 45S promoter (C) or ITS-1 (D) methylation levels (Y axis), respectively.

indication that a genetic component is controlling methylation at this locus. Comparing rDNA methylation with copy number, we found that these two measures were correlated (Figure 2.6C). These results suggest that our rDNA methylation differences are caused by differences in rDNA copy number; a function that might be conserved among tandem repeats (Brahmachary et al., 2014). Interestingly, the unlinked 5S and 45S rDNA loci are coregulated, as 5S copy number is correlated with 45S methylation (Figure 2.6D) (Gibbons et al., 2015). Although we don't know the mechanism that DNA methylation uses to count copies, rDNA dosage is homeostatically regulated by cytosine methylation (Grummt and Pikaard, 2003). When rDNA copy number increases, cytosine methylation increases to maintain the same number of active copies.

Link between Metabolism and rDNA Variation

Since rDNA plays a central role in cellular metabolism, modulation of its activity influences a number of metabolic processes (Oie et al., 2014; Paredes et al., 2011). We wanted to know the effects of inherited changes in rDNA architecture on offspring metabolism; specifically, the cholesterol biosynthetic pathway. Astonishingly, both rDNA methylation and rDNA copy number in sperm were anti-correlated with Sqle expression in IVF offspring livers (Figure 2.7A,B). While extemporaneous pre-rRNA expression was not significantly correlated with *Sqle* expression (R = 0.13), these relationships could point to differential regulation of rRNA at an earlier time point in development or altered nuclear architecture

A

Sqle expression level, IVF offspring

R = -0.72

Average 45S rDNA methylation in dad's sperm

B

Sqle epression level, IVF offpring

R = -0.57

45S rDNA copy number in dad's sperm

C

Averaged 45S promoter methylation, sibling

Low Protein
Caloric Restriction
High Fat

Averaged 45S promoter methylation, Control

D

Control   Low Protein   High Fat

Methylation level, individual CpGs

Hepatic *Sqle* mRNA abundance

Offspring mRNA abundance

Promoter   *ITS1*

Paternal 45S rDNA methylation

123

Figure 2.7

rDNA Variation Influences Offspring Metabolism Independent of Paternal Diet

A. Paternal rDNA methylation was anticorrelated with offspring cholesterol gene expression. Here, the y axis shows expression level of *Sqle* in offspring liver (shown as $10\text{-}(Ct^{Actb}\text{-}Ct^{Sqle})$).

B. Paternal rDNA copy number and offspring cholesterol genes expression. Same as in (A), except x axis represents 45S rDNA copy number determined by ddPCR (using β-*Act* for normalization).

C. 45S methylation levels for Control animals (x axis) along with matched siblings raised on various diets (y axis). Circles show pairs (or trios) of sperm samples used for IVF.

D. rDNA methylation data (left panels) and *Sqle* expression levels (right panels). Individual bars in the left panel show individual CpGs (differing numbers of ITS-1 CpGs result from CpGs failing pyrosequencing quality control). Individual bars in the right panel represent individual offspring of a given IVF experiment.

caused by this change in rDNA methylation. Both rDNA methylation and copy number in sperm affects expression of metabolically relevant genes in offspring.

Although unlikely, the relationship between rDNA regulation and offspring metabolism made us ask if rDNA was carrying paternal dietary information. We didn't see any directed change in rDNA methylation with mice subjected to different diets (Figure 2.7C). However, since our paternal diet-induced phenotype is not completely penetrant (Carone et al., 2010), we tested the possibility that the sperm samples that showed decreases in rDNA methylation were the ones that produced offspring with elevated cholesterol metabolism. We utilized our IVF system to determine if changes in paternal rDNA methylation cause dietary effects on offspring metabolism. As seen in Figure 2.7C-D, the changes in rDNA methylation are inconsistent with changes in Sqle expression. While some matched litters have increased rDNA methylation and increased Sqle expression, others show the opposite effect (decreased rDNA methylation and increased Sqle). Therefore, it is highly unlikely that rDNA mediates the paternal dietary effects on offspring metabolism, while still influencing metabolism through stochastic variation.

Discussion

Taken together, these results prove that sperm carries the diet-induced epigenetic information that alters offspring metabolism. Furthermore, our extensive methylome studies reveal prevalent epivariability over copy number variable regions. Although we failed to find consistent dietary effects on the methylome that explain offspring phenotypes, sperm epivariability is a major contributor to offspring metabolism on its own. The epivariable rDNA locus has coordinate methylation and copy number changes that influence offspring cholesterol metabolism. Therefore, our results reveal an unappreciated role for copy number variation-coupled epivariability in regulating inherited phenotypes; which acts independently of diet-induced reprogramming of sperm.

Although cytosine methylation has been implicated in regulating transgenerational environmental effects, no conclusive evidence has been obtained. Methylation changes in many of these studies were modest over several CpGs (Dias and Ressler, 2014; Ng et al., 2010), which should preclude these changes from being the causative agent for penetrant phenotypes. We find large amounts of epivariation in our sperm methylomes with little contribution from diet-induced differences. This indicates that post-weaning diet has little effect on the methylome, which is largely established prior to this period. Interestingly, the Ferguson-Smith group found more consistent dietary-induced methylation differences in their model, which probably derives from their

manipulation of *in utero* development, when gametic methylomes are established (Radford et al., 2014). Therefore, the sperm methylome is largely refractory to reprogramming by chronic dietary changes after being established *in utero*.

Despite the lack of diet-induced changes, our broad analysis of sperm methylomes enabled the discovery of drivers of epivariation between animals. CpG island shores and tandem repeat regions display the greatest amounts of epivariation in inbred strains of mice. The variation at CpG island shores is likely attributed to differential usage of these elements during spermatogenesis, as the width of the shore that remains unmethylated is linked to regulatory activity in cell type specific lineages (Irizarry et al., 2009; Ji et al., 2010). Variation at CpG island shores is probably lost after fertilization, since most methylation surrounding CpG islands is erased during pre-implantation development (Smith et al., 2012). On the other hand, tandem repeat genes represent a unique class of genes that have extensive copy number variation between animals, indicating dynamic instability at these loci. These characteristics of tandem repeats lead to epivariation at these loci, since we show that changes in copy number are correlated with methylation changes at rDNA loci. We show that this form of epivariation is paternally inherited, so cytosine methylation over tandem repeats represents heritable regulatory variation.

We find that combined genetic and epigenetic variation at the rDNA loci in fathers is linked to offspring metabolism. Since rDNA represents a central node in metabolic control (Grummt, 2013), altering the regulation of rDNA should have a dominant effect on cholesterol metabolism. Interestingly, the copy number of rDNA is variable within inbred strains of mice, showing that these mice are not genetically identical, as is assumed in numerous studies. This stochastic copy number variation of rDNA also leads to heritable genetic variation that strongly influences offspring metabolism. Since rDNA shares many characteristics with other tandem repeats, heritable copy number variations of other tandem repeats might have large effects on quantitative traits. Copy number variation at other tandem repeats have been shown to effect diverse phenotypes, such as immune function and starch digestion (Hollox et al., 2008; Perry et al., 2007). Therefore, previously unaccounted for phenotypic differences may be attributable to the combined effects of heritable genetic and epigenetic variation at tandem repeat regions (Manolio et al., 2009). Tandem repeats function in numerous pathways, from totipotency to olfaction to metabolism, so their variation likely has wide ranging effects.

Conclusion

We show that sperm transmit paternal dietary effects, a major step in finding the elusive mechanism of paternal effects. Our focus on the sperm methylome uncovered genetic and epigenetic variation in sperm that leads to heritable

changes in offspring metabolism. Our inability to find considerable diet-induced changes in sperm methylation strongly suggests that cytosine methylation does not carry paternal dietary information. Future studies searching for the mechanism behind environmentally-induced intergenerational phenomena should focus on other potential epigenetic carriers of information, including chromatin and small RNA. The combined effects of sperm epivariation and diet-induced changes in sperm contributes to a large range of metabolic outcomes in offspring.

Methods

Animal husbandry

Mice used in this study included C57Bl6/J and FVB/NJ strains from Jackson Laboratories. Animals were maintained on-site in accordance with approved IACUC protocols.

Dietary regimens

The 63 animals in the epivariation dataset included animals consuming standard laboratory chow, a defined Control diet (Bioserv AIN-93g), a Low Protein diet based on AIN-93g (10% of protein rather than 19%, remaining mass made up with sucrose), a High Fat diet 60% Fat based on Ain-93g (Bioserv S3282), as well as animals provided with nicotine hydrogen tartrate (200 mg/mL nicotine,

reported as free base) in drinking water sweetened with 2% saccharine to increase palatability, or animals provided with tartaric acid and saccharine water alone. Animals were placed on diet at weaning (21 days) until mating or sacrifice (10-12 weeks).

Isolation of epididymal sperm DNA

Animals were sacrificed by isoflurane administration followed by cervical dislocation. For sperm isolation, cauda epididymis and epididymis were rapidly dissected, and punctured in 500 mL of human tubal fluid (HTF – Millipore MR-070-D) and incubated at 37 C for 30 minutes to capacitate. Supernatant was removed, and 500 mL of PBS was added to bring total volume to 1 mL. Somatic cell fragments were removed and 0.1% Triton-X 100 (Sigma) was added with vortexing. Sperm were pelleted at 2000 g for 5 minutes, washed once with 1 mL water and then pelleted. They were washed again with 1 mL PBS, and pelleted again. Sperm was resuspended in 400 mL DNA Lysis Buffer (10mM Tris, 5mM EDTA, 0.5% SDS, and 200mM NaCl) with 10mM DTT, and incubated at 37°C for 30 minutes. Sperm were subjected to needle homogenization. 20 mg/mL Proteinase K was added to the homogenate and incubated at 55°C for 16 hours. DNA was extracted with Phenol:Chloroform:Isoamyl Alcohol and precipitated with isopropanol.

In vitro fertilization

In vitro fertilization was performed according to "Manipulating the Mouse Embryo" Second Edition (Hogan, 1994). Superovulated FVB/NJ mice were used as egg donors and sperm was isolated from males fed dietary regimes as above. Isolated sperm were capacitated in HTF for 30 minutes at 37°C. Fertilization took place in 250 mL HTF media covered in mineral oil, pre-gassed in 5% CO_2 at 37°C. Fertilized embryos were developed in KSOM (Millipore - MR-020-P) until the 2-cell stage. Swiss Webster Females between 25 and 35 grams were used as 2-call stage embryo recipients via unilateral oviduct transfer.

Microarray Experiments.

Affymetrix mouse GeneChip ST arrays were used for whole gene expression analysis. Specifically, RNA was prepared by Trizol extraction, prepared with Ambion WT expression kit, and hybridized to Mouse 1.0 ST and Mouse 2.0 ST arrays according to manufacturer's instructions.

Reduced Representation Bisulfite Sequencing (RRBS)

RRBS was carried out as previously described (Meissner et al., 2008). Briefly, genomic DNA was digested with MspI, ends were filled-in, and fragments were A-tailed. DNA fragments were ligated to methylated barcoded adaptors. DNA was subjected to bisulfite conversion and PCR amplified. Cleanup and size selection were performed with SPRI AMPure XP beads.

Whole Genome Bisulfite Sequencing (WGBS)

Control and Low Protein WGBS pools were generated from 8 paired animal samples in which one sibling was weaned to Control diet and the other sibling was weaned to Low Protein diet. Control/High Fat pools were generated using a similar approach using 7 animal pairs. For each pool, each animal contributed 1 ug of genomic DNA. DNA was sheared to an average length of ~100-500 bp with a Covaris sonicator, fragment ends were cleaned up and A-tailed. Fragments were ligated to pre-methylated Illumina paired-end adaptors, bisulfite converted, and PCR-amplified. Libraries were subjected to paired-end 50 bp sequencing on Illumina HiSeq sequencers, yielding an average of 1.4 billion reads and 140 billion base pairs of sequence.

Data Processing and Analysis

Technical replicates were merged together and only the first member of a read pair was used. Data were mapped against mm9 mouse genome with bsmap software v2.73 (Xi and Li, 2009). Default parameters were used for error rate and maximum number of equal best hits was selected as default as well. Two versions of mapping were performed: all reads; non-identical reads. "Non-identical" means that if more than one read had the same initial sequence, only one was used for mapping and further analysis. The same software was used to perform methylation calls. To get a methylation level for a given CpG, information from C from both strands was combined together.

Differentially-methylated CpGs were discovered with the methylKit R package (Akalin et al., 2012). Discovered CpGs were merged using tiling arrays with 300nt windows to calculate p-values, q-values, and fold enrichment factors for High Fat and its Control, Low Protein and its Control. The data can be found in Table S2.2.

Pyrosequencing

Cytosine methylation data for individual loci were generated using a Qiagen Pyromark Q24 pyrosequencer. Genomic DNA was bisulfite converted, and loci to be analyzed were amplified by PCR – primers are listed in Table S2.4. Amplified DNA was cleaned up and analyzed using the manufacturer's protocol for the Pyromark Q24.

rDNA copy number analysis

rDNA copy number was measured using q-PCR and digital droplet PCR (ddPCR). Briefly, genomic DNA was subject to q-PCR using primers located within the rDNA locus, at the *Acacb* single copy locus, and at the multi-copy tRNA-Lys genes. While both of these normalization controls were highly-correlated with rDNA data, *Acacb* primers were in the linear range over a greater fraction of the dataset and so were chosen.

ddPCR was performed according to manufacturer's protocol (Bio-Rad). Briefly, DNA was digested with DpnI. Samples were mixed with primers for 45S rDNA

and B-Actin followed by droplet generation and PCR. Quantitation of products was performed on ddPCR machine.

Table S2.4

Primer List

Primer Name	Target	Assay	Sequence	Reference
45S_rDNA_Bio-BF2	45S rDNA	Pyro	Bio-GAATTTGATTATTTAGAGGAAGTAAAAGT	
45S_rDNA_BR2	45S rDNA	Pyro	CACCCACCCCTTCTCT	
45S-rDNA_Seq2	45S rDNA	Pyro	CTCACTCCAAACACC	
45s_Spacer_BF2	45S rDNA	Pyro	GGAAGTGTTTGTGGTGAGG	Shiao *et al.*
45s_Spacer_Bio_BR2	45S rDNA	Pyro	BIO-CACCAACCCTAACATTTTTCC	Shiao *et al.*
45s_Spacer_Seq 3	45s rDNA	Pyro	GTTTTGGAGATGGTGT	Shiao *et al.*
rDNA_Term_F1	45S rDNA	qPCR	GAACCTTTAGGTCGACCAGTTG	
rDNA_Term_R1	45S rDNA	qPCR	ACAAAGTACCACCCGGAGTA	
Lys(CTT)_F1	Lys tDNA	qPCR	CGGCTAGCTCAGTCGGTAGA	
Lys(CTT)_R1	Lys tDNA	qPCR	AACCCACGACCCTGAGATTA	
Acacb_F1	Acacb	qPCR	TGCTCATAGGCCAAGAGAAAGGCT	
Acacb_R1	Acacb	qPCR	AGTGCTGGGATTACAGGCATGAGT	
5S rDNA F2	5S rDNA	qPCR	CCCGATCTCGTCTGATCTC	
5S rDNA R2	5S rDNA	qPCR	CCTACAGCACCCGGTATT	
Sqle_F1	Sqle	RT-qPCR	TCCTTGCATCAGCTCCGAAA	Carone *et al.*
Sqle_R1	Sqle	RT-qPCR	CGGTCAAAGCAACCCAACAGGACCG	Carone *et al.*
B-Actin_F1	B-Actin	RT-qPCR	GGCTGTATTCCCCTCCATCG	
B-Actin_R1	B-Actin	RT-qPCR	CCAGTTGGTAACAATGCCATGT	

CONCLUSION

Summary

The preceding work demonstrates that paternal diet controls medically important metabolic phenotypes in offspring. We observe transmission of dietary information to the zygote via sperm, and this information evades reprogramming that typically occurs after fertilization. Since cytosine methylation is implicated as a major contributor to meiotic epigenetic inheritance in several transgenerational phenomena, we profiled cytosine methylation of sperm treated with different dietary regimens. Our extensive characterization of the sperm methylome reveals that diet does not significantly affect methylation patterns. However, we find that extensive epivariability in the sperm epigenome makes important contributions to offspring variation. Importantly, coordinate cytosine methylation and copy number changes over the ribosomal DNA (rDNA) locus contributes to changes in offspring metabolism. Thus, rDNA variability acts independently of post-adolescent paternal diet to influence offspring metabolism. Therefore, at least two mechanisms exist for epigenetically controlling offspring metabolism: stochastic epivariation and diet-directed changes. This work argues that an offspring's phenotype can no longer be viewed solely as the result of genetic interactions with the developmental environment - the additional influences of paternal environment and inherited epigenetic variability must also be

considered. These findings reveal novel contributions to metabolism that could revolutionize how we think about the risk factors for human health and disease.

Intergenerational Epigenetic Inheritance

Paternal dietary effects challenge the notion that environmentally-induced epigenetic information is reset between generations. This intergenerational memory demonstrates that epigenetic plasticity, which is a hallmark of mitotic epigenetic mechanisms, increases phenotypic variation through the process of reproduction as well. Concurrent with our finding that paternal low protein diet alters offspring metabolism, other groups realized similar phenotypes in rodents utilizing *in utero* undernutrition, intermittent fasting, and high fat diet (Anderson et al., 2006; Jimenez-Chillaron et al., 2009; Ng et al., 2010). In addition, paternal effects from liver damage and endocrine disrupters imply that environmental factors other than nutritional information are paternally transmitted (Anway et al., 2005; Zeybel et al., 2012). These findings inform us that we need to update inheritance paradigms to include epigenetic contributions from ancestral environments in addition to the genetic and developmental influences on phenotype.

Sperm mediated transmission of paternal environmental response

While a number of non-gametic carriers, such as seminal fluid contents or cryptic maternal responses to mate quality (Bromfield et al., 2014; Velando et al., 2006), are known to affect offspring phenotypes - we conclusively find that sperm carries paternal dietary information. Our finding that sperm transmits this information reveals the incongruity of the soma-germline barrier that was thought to protect the germ line from environmental insults (Weismann et al., 1891). Although we are the first to show that sperm carry nutritional information to offspring, relevant questions persist as to the carrier of other environmental information that is inherited, such as behavioral or chemical stresses. Previous studies attempting to locate the carrier of environmental stress have come to conflicting results (Dias and Ressler, 2014; Dietz et al., 2011). These other studies focused on inheritance of induced neurological phenotypes, which might be transferred by another carrier, such as the aforementioned seminal fluid. In spite of these conflicting reports on the carrier of inherited behavioral responses, we find that sperm conveys heritable nutritional information to subsequent generations. Future studies determining the carrier of other environmental information will give the field a better idea if there is a generalized mechanism for informing offspring of ancestral environmental stress.

We speculate that dietary information reaches sperm through nutrient sensing pathways, and is converted to epigenetic information that is subsequently carried

to the mature oocyte. There are several barriers to sperm passing on epigenetic information intrinsic to the process of fertilization, which make our findings even more interesting. Sperm lack the epigenetic richness of somatic cells, containing minimal amounts of histones (~5% in sperm) and cytoplasmic information carriers (Hammoud et al., 2009). Additionally, the sperm DNA methylome is actively erased at fertilization to more closely resemble the oocyte methylome. However, dietary information survives despite processes that reprogram the remaining sperm epigenome to totipotency after fertilization (Hackett et al., 2013). In spite of these epigenetic barriers, sperm transmit dietary information to offspring, leading to the exigency to characterize the sperm epigenome.

Sperm Methylome

Carriers of epigenetic information in sperm include cytosine methylation, histones, and small RNA species (McLay and Clarke, 2003). Cytosine methylation endures as the best-studied intergenerational information carrier in mammals; essential for the variability of epialleles and the programming of imprints (Goll and Bestor, 2005). We characterized the sperm methylome utilizing several high resolution, quantitative techniques. In depth profiling revealed extensive epivariation between inbred animals, but minimal effects of diet on this epigenome. Interestingly, epivariation was apparent in every technique used to measure cytosine methylation - even when attempting to guard against it by

using pooling strategies that should minimize variation. Therefore, epivariability is the dominant feature controlling differences in sperm methylomes.

The lack of consistent dietary effects in this sea of variability convinces us that chronic dietary change after adolescence has minimal effect on cytosine methylation. Since we see consistent paternal dietary effects on offspring metabolism, this likely informs us that cytosine methylation is not the relevant carrier of hereditary dietary information. This epivariability stands in stark contrast to the consistent sperm methylation changes prompted by manipulating early *in utero* environments - when germ cells are reprogrammed (Radford et al., 2014). These combined results argue that early development establishes sperm methylomes that resist subsequent modest environmental influences. Future studies should be directed at determining if more stringent environmental conditions reprogram the mature sperm methylome, answering whether more severe insults begin to impinge on the established sperm methylome.

Although diet has muted effects on the sperm methylome, the extensive epivariability of sperm indicates unexpected factors shape this epigenome over time. We searched in vain for confounding factors that could explain this epivariation, including seasonal effects, litter size effects, etc. Instead, the sperm methylome stochastically changes over time, with epivariability enriched at tandem repeats and other copy number variable elements, such as the rDNA

locus and *Ifnz* clusters. CpG island shores represent the other minor contributor to epivariability, similar to their roles in methylation variability in somatic lineages (Irizarry et al., 2009). The epivariation in CpG island shores probably results from differential usage during spermatogenesis, but is reprogrammed during early development (Smith et al., 2012), which minimizes any contribution that CpG island shore variability make to offspring phenotypic differences. On the other hand, we have evidence that epivariation at tandem repeats accumulates through multiple generations. The heritability of epivariation at tandem repeats stems from the observations that siblings have more similar methylomes than unrelated animals, and that the methylation patterns of the epivariable rDNA are inherited.

Methylation at Tandem Repeats

The epivariability at tandem repeats presents several interesting possibilities about the function of cytosine methylation at these elements. Threshold models for the targeting of cytosine methylation to repeated elements have been proposed (Goll and Bestor, 2005) that state that repeats need to reach a certain copy number before being recognized as selfish elements and becoming methylated. On the other hand, cytosine methylation represses meiotic recombination between unpaired regions (Colot et al., 1996), so it might be mediating stability of these repetitive regions when they are unsynapsed during meiosis. These models reveal intimate links for cytosine methylation and copy

number variation (CNV), but we propose that cytosine methylation at tandem repeats buffers the number of active copies. Since tandem repeats are subjected to relatively rapid copy number changes, we investigated whether the rDNA loci underwent copy number changes in sperm. We found CNV in the rDNA locus that, like cytosine methylation, was greater between unrelated animals than between siblings. Most interestingly, the copy number changes correlated with cytosine methylation changes, which presented us with this interesting concept. The changes in cytosine methylation might occur in response to copy number changes to homeostatically regulate the activity of rDNA. In other words, an increase in the number of copies of rDNA would only increase the number of silenced copies - thus maintaining relatively constant dosage. We find minimal difference in levels of rRNA between mice with different rDNA copy numbers in qPCR results, arguing for this homeostatic control. Recently, another group found correlations between CNVs at other tandem repeats and cytosine methylation in human populations, arguing that this regulation tandem repeats is likely conserved (Brahmachary et al., 2014).

The lability of the rDNA locus demonstrates how combined genetic and epigenetic variation can be inherited. Others have shown that rDNA copy number is different between inbred mice, but the extent of this variation was unknown (Shiao et al., 2012; Veiko et al., 2007). We find that rDNA copy number is highly variable between mice - rDNA copy number can double within a strain of inbred

mice by our estimates. There were even differences of up to ~50 copies (~25% difference) between some siblings. These large differences in copy number could be detrimental if they were not homeostatically regulated, thus the concordant increase in cytosine methylation.

Effects of Tandem Repeat Variation

Do these variations manifest as phenotypic effects? Does rDNA escape homeostatic regulation to alter phenotypes? rDNA resides at the nexus of metabolic regulation, so changes in rDNA dosage will influence metabolic phenotypes (Grummt, 2013). The observed variations in rDNA appear to be intimately linked to heritable control of metabolism, as paternal methylation and copy number of rDNA are anti-correlated with cholesterol gene expression in offspring. Interestingly, another group recently showed that artificially increasing levels of rRNA in livers by perturbing the function of a negative regulator of rRNA synthesis led to multiple metabolic phenotypes, including a decrease in the level of cholesterol (Oie et al., 2014). Also, rDNA control over metabolism appears to be conserved between species. Artificially reducing rDNA copy number in flies largely affects metabolic genes, arguing the metabolic effects of rDNA CNVs are maintained throughout metazoan (Paredes et al., 2011). This metabolic association with rDNA variability is probably just the tip of the iceberg for the ability of tandem repeats to affect phenotypes, as CNV in other tandem repeat

genes have pleiotropic effects, such as the association of β-defensins with psoriasis and the expansion of amylase genes to extract nutrition from starch rich diets (Hollox et al., 2008; Perry et al., 2007).

Therefore, tandem repeat associated genetic and epigenetic variation contribute to diverse phenotypes. Previous difficulties in ascertaining CNV had prevented in depth studies of subtle copy variations with phenotypes. However, newer technologies enable relatively accurate quantitation (Brahmachary et al., 2014; Hindson et al., 2011), making it possible to investigate the relationship between relatively small copy number changes and phenotypic variation. Research into this area may provide fruitful insights into complex diseases, such as the causes of autism spectrum disorders and metabolic disease (Tang and Amon, 2013). Recently, the use of new technologies has enabled the discovery of a link between rDNA copy number and mitochondrial density, another major contributor to energy homeostasis. Mitochondrial abundance is influenced by several factors - including rDNA copy number (Gibbons et al., 2014), which further implicates copy number variation as a modulator of metabolism. On a related note, Bradford *et al* found consistent changes in rDNA methylation induced by *in utero* undernutrition (Radford et al., 2014), and integrating our findings with theirs indicates that CNV may be directed by environmental stress at more sensitive times during development. In the future, if it is found that environment can direct changes in copy number, then we might be able to alleviate the burden of

143

diseases caused by CNVs, such as Huntington's disease and Fragile X Syndrome (McMurray, 2010), through interventions that control copy number.

Paternal Dietary Effects

On top of the combined effects of epivariability and copy number variation, paternal diet directs cholesterol and lipid biosynthetic gene expression in the livers of adolescent offspring mice. This layered inheritance of metabolic control - one layer stochastic, the other environmentally directed - buffers offspring phenotypes against multiple environments by restraining offspring phenotypic variation. This buffering could ensure offspring survival by hedging against the possibilities of encountering any number of environments (Feinberg and Irizarry, 2010). The directed response to diet informs offspring of prevailing environments, while stochastic variation tempers this phenotypic response.

Offspring Metabolism

These strategies of epigenetic inheritance result in metabolic offspring phenotypes. While our study focused on post-weaning hepatic phenotypes, another study confirmed our results that low protein diet causes metabolic phenotypes in offspring, but instead focused on adult offspring metabolic responses (Watkins and Sinclair, 2014). In the other study, offspring from low protein mice developed diabetic phenotypes (glucose intolerance and insulin

resistance) and cardiovascular deficiencies during adulthood. Although our studies didn't directly analyze diabetes-related phenotypes, liver health and diabetes are intimately linked (Perry et al., 2014). In addition to this link, cholesterol represents a metabolite that is strongly implicated in cardiovascular disease in humans (O'Donnell and Elosua, 2008). Paternal low protein diet studies suggest that epigenetically inherited metabolic dysfunction early in life evolves into metabolic disease later in life. These findings could be essential for combating metabolic diseases - with interventions designed for ancestral generations.

Functionally, we discovered seemingly paradoxical regulatory modules in our offspring hepatic studies. Whereas we detected increased expression of cholesterol biosynthetic genes in livers from low protein fathers, those same livers had decreased levels of cholesterol metabolites. Therefore, increased levels of cholesterol biosynthesis genes did not lead to an increase in the concentration of their products for mice at weaning age. One rationale for this "mismatch" could be that low levels of cholesterol in livers from low protein result from increased efflux or reduced absorption of cholesterol, and cholesterol gene expression is responding to these low levels of cholesterol through SREBP function (Brown and Goldstein, 1997). If cholesterol gene expression becomes locked in at these levels, then cholesterol metabolite levels would be increased later in life when either efflux is reduced or absorption increased (Holzer et al.,

2013; Sene et al., 2013). The observation that low protein offspring suffer from diabetes later in life argues that there is a mismatch between molecular expectations and realities.

This brings us to another interesting point about the outcomes of environmentally induced transgenerational inheritance. A hypothesis exists for the cause of metabolic disease deriving from a mismatch between the expectations and realities of future environments. This has been called the "Thrifty" hypothesis, but is also known as "the developmental origins of health and disease" (Hales and Barker, 2001). In this hypothesis, early environments set later phenotypes to match those early environments - in other words, greater plasticity exists early in development, and established epigenomes are insensitive to later environments. One of the major barriers to reprogramming differentiated cells to pluripotent cells is the epigenetic memory that needs to be erased (De Carvalho et al., 2010). As a case in point, early undernutrition increases energy storage at later stages of life (Meaney et al., 2007), and if abundant nutrition is encountered later on, a "mismatch" will occur that causes metabolic disease through excessive energy storage. Now, recent transgenerational epigenetics studies imply that the signals for this mismatch arise in the parental generation, and that periods of environmental sensitivity extend to the parental generation. A test to determine the relevance of this "thrifty" hypothesis to our work would be to change the diets of offspring to match the parental diet. If the "thrifty" hypothesis pertains to cases

of environmentally induced transgenerational phenomena, then matching the parental and offspring environments would alleviate disease relevant phenotypes.

Our results implicate numerous environmentally sensitive molecular pathways in the regulation of offspring metabolic phenotypes. The liver methylome and transcriptome of our offspring mice indicate that several metabolic pathways are affected by paternal diet, which are controlled by master regulators, such as PPARA and SREBP. For instance, PPARA could be altering up to 15% of the gene expression changes observed in the livers of our low protein offspring, while our strong cholesterol phenotypes could be mediated by changes in SREBP. Although *Ppara* expression is changed in our offspring, and this correlates with changes in enhancer methylation, no changes were observed at the *Ppara* locus in sperm. This implies that the changes to *Ppara* arose at some developmental stage, and that other upstream regulators signal the change in *Ppara*. This could be a common theme for other regulators of the liver phenotype (such as *Srebp*), as we did not observe cytosine methylation changes in sperm for any of the master regulators that control large portions of our liver phenotype.

Offspring Epigenetic Differences

While hundreds of genes were altered in our low protein offspring, many with greater than 2-fold changes in abundance, our epigenetic analysis of livers found

147

modest changes in the methylome. The methylation differences observed at promoters did not correlate with these gene expression differences, implying that cytosine methylation does not fine tune expression responses at every individual gene. This finding is in line with other studies finding relatively modest correlations between cytosine methylation and global expression differences (Bock et al., 2012). On the other hand, differences in cytosine methylation at important regulators, such as the previously discussed *Ppara*, could affect downstream effectors. The differences in cytosine methylation arise at some point during development, since sperm do not exhibit dietary differences in cytosine methylation. The difficulty with trying to discover the origin of these differences is that there are numerous times in development that could elicit this difference: from early cell fate specification to later hepatogenesis. Several transitions occur in the DNA methylome during development, and whether these modest differences result from changes in establishment or erasure would be a difficult to ascertain. Liver methylation differences may have arisen from intercellular interactions, further complicating extraction of the epigenetic pathways. The complexity of cellular and molecular interaction precludes us from formalizing a conclusion with regards to the synthesis of our liver phenotype at this point.

Paternal Effects on Early Embryogenesis

We are actively investigating the ontogeny of our offspring phenotypes with the knowledge that sperm carry ancestral dietary information. This enables us to focus our attentions on differences at fertilization and the early embryonic development bottleneck. Our efforts comprise characterization of the transcriptional changes elicited by paternal diet at these stages of development as a means to discover the affected pathways. We are focusing on early embryonic gene expression differences because alterations in transcriptional programs during pre-implantation development are known to lead to downstream metabolic phenotypes (Lane et al., 2014). For instance, culturing mouse embryos *in vitro* for longer periods leads to metabolic phenotypes (Feuer et al., 2014), indicating that early environmental stresses have long term consequences - by setting the epigenetic state at these points according to the "Thrifty" hypothesis.

Our preliminary results (included in Appendix I) on the pre-implantation embryo transcriptome show that paternal diet alters a number of genes important for early cell fate specification programs - possibly changing the relative contribution to extraembryonic tissues (Macfarlan et al., 2012). A number of these cell fate specification genes are controlled by MERVL elements in an interesting example of co-option of retroelements to regulate cellular processes. Chimeric transcripts of MERVL with endogenous genes comprise a coordinately expressed network, and expression of these genes correlates with increased totipotency. Therefore,

our later metabolic phenotypes might arise from altering early cell fate decisions to affect the relative size of the placenta versus the embryo proper. Since this finding is preliminary, work needs to done to determine the functions of early gene expression changes.

Future Directions

Continuing research needs to be undertaken to find the mechanisms by which paternal environmental information is transmitted to offspring. Since we discovered that sperm transmits this information, focus should be paid to additional carriers of epigenetic information in sperm - most prominently histone modifications and small RNAs (Rando, 2012). While diet minimally influences the sperm methylome, cytosine methylation might still be essential to the process of informing offspring of paternal environments. For example, H3K9me2 protects cytosine methylation in the early embryo (Nakamura et al., 2012), and diet might cause differential retention of H3K9me2, which still needs the underlying cytosine methylation to provide robust epigenetic memory (Hathaway et al., 2012). For all potential epigenetic carriers, a signature of the difference should be able to be captured by transcriptional differences in early development. After paternal diet-influenced transcriptional programs are found during pre-implantation development, these programs should be manipulated to determine their downstream effects.

Conclusion

In total, the research has made a significant contribution to the understanding of the factors influencing offspring variability. Changes in paternal diet alter medically important metabolic phenotypes in offspring. Sperm carries this environmental information. While dietary epigenetic information is not contained in cytosine methylation, epivariation in the sperm methylome contributes to offspring variation. Most interestingly, this epivariation is intimately linked to copy number variation, regulating the dosage of associated genes - such as ribosomal DNA. Intergenerational regulation of ribosomal DNA influences offspring metabolic variation, on top of which, diet acts through an independent mechanism. All of these findings add to the determinants of phenotype. Previously, it was thought that genetic interactions with the environment in one's lifetime determined their phenotype. Now, a parent's environment, along with rapid genetic and epigenetic stochastic variation, increases phenotypic variation in offspring.

Early embryonic effects of paternal

dietary manipulation

Preface

Most of this work is preliminary, and has not coalesced into a coherent story. However, an early result from this work is included in a manuscript under review. Here is the current title and author list on the submitted manuscript:

Paternal diet alters tRNA fragment levels throughout the male reproductive tract in mammals

Upasna Sharma, Colin C. Conine, Jeremy M. Shea, Benjamin R. Carone, Clemence Belleannee, Xin Li, Xin Y. Bing, Lucas Fauquier, Weifeng Gu, Philip D. Zamore, Robert Sullivan, Craig C. Mello, Manuel Garber, and Oliver J. Rando

The result included in both this chapter and the manuscript is the analysis on MERVL elements represented in Fig.A1.3. This analysis derives from data generated by Jeremy Shea.

Abstract

Paternal dietary interventions influence offspring metabolic phenotypes, altering biomarkers involved in diabetes and metabolic syndrome. Using IVF, we demonstrated that sperm carry the effects of altered paternal diets to the oocyte. Since the information needed for paternal diet-induced offspring phenotypes is contained in sperm, we hypothesized that this information may cause transcriptional changes in the pre-implantation embryo. Therefore, we have undertaken an exhaustive transcriptome profiling screen on hundreds of IVF-derived embryos to search for dietary-induced expression changes in pre-implantation development. We profiled single embryos at several stages of pre-implantation development (2-cell, 4-cell, morulae, and blastocyst) to determine stage-specific changes caused by paternal diet. So far, this screen has revealed that paternal diet contributes to early embryonic regulation of totipotency factors driven by MERVL retroelements at the 2-cell stage.

Introduction

Reprogramming gametes after fertilization restores totipotency to the developing organism, enabling it to follow proper developmental trajectories. This reprogramming ensures that most somatic insults acquired by a parent do not afflict offspring phenotypes. However, accumulating examples of phenomena

153

that escape reprogramming reveal that not all transgenerational information is lost, and that numerous environments encountered by parents influence offspring phenotypes (Heard and Martienssen, 2014; Rando, 2012). Several studies have shown that paternal diet is a major modifier of offspring metabolism (Carone et al., 2010; Jimenez-Chillaron et al., 2009; Ng et al., 2010). The highly specialized sperm, which undergoes extensive reprogramming after fertilization, carries this hereditary information. Sperm commits little more than its nuclear content to the zygote, and this nuclear material undergoes extensive DNA demethylation and replacement of the ~95% of the genome packaged in protamines with new maternally contributed histones (McLay and Clarke, 2003). Although a majority of information is lost, dietary information contained in sperm affects offspring metabolic phenotypes. A fingerprint of this information should be apparent in early embryogenesis, possibly as diet-altered chromatin or changes in gene expression.

Changes in the early embryo have substantial effects on later health, and these observations have been integrated into the Developmental Origins of Health and Disease (DOHaD) or "Thrifty" hypothesis (Hales and Barker, 2001). Early embryogenesis represents an environmentally sensitive stage of development where the epigenome is established, which becomes progressively less plastic at later stages of development. This means that the environment encountered by the early embryo determines the equilibrium of the later epigenome. So if the

environment drastically changes after early development, the epigenome may be mismatched to the new environment.

One of the most labile periods for setting later homeostasis appears to be the pre-implantation embryo. The pre-implantation stage is extremely sensitive to *in utero* conditions, as solely restricting a mother's diet during this period has lasting consequences on offspring health (Lane et al., 2014), with those offspring from restricted mothers showing symptoms of metabolic disease. Manipulating a mother's diet at this point alters *in utero* conditions for longer periods, so IVF experiments that change the conditions of the early embryo are useful in determining exactly how much this period affects later health. Humans and mice derived from IVF have altered metabolism later in life, thus this pre-implantation stage is a critical window for setting later metabolism (Chen et al., 2014). Additionally, others have shown that altering the media conditions used for IVF influence offspring phenotypes (Feuer et al., 2014), which points to the ability of the pre-implantation environment to determine later phenotypes. Thus, characterizing environmental effects on pre-implantation embryos will help in determining the changes that cause later phenotypes.

Alterations to the embryonic environment during pre-implantation development may change the kinetics of genome activation and epigenome reprogramming. Since paternal diet influences offspring metabolic phenotypes similarly to pre-

implantation manipulation, both of these perturbations might elicit similar changes in pre-implantation development. Early embryonic gene expression transitions from being dominated by oocyte-specific factors to those needed for pluripotency (Deng et al., 2014). Some of the earliest factors activated from the paternal genome include Pol I directed transcription of rRNA needed for the ensuing rapid development (Lin et al., 2014). Shortly afterwards, a burst of totipotency-associated expression ensues, including release of several retrotransposon families and chimaeric transcripts driven by retrotransposons (such as MERVL elements) (Macfarlan et al., 2012). The first differentiation decisions follow during the morulae stage, with outer cells preferentially activating trophectoderm/extra-embryonic programs, and inner cell expressing pluripotency markers, such as *Oct4* (Guo et al., 2010). Paternal dietary manipulations influence the relative contributions of the trophectoderm versus the inner cell mass (ICM), possibly by changing the dynamics of transcriptional transitions (McPherson et al., 2013). Thus, paternal dietary effects should be present in the pre-implantation embryo transcriptome.

In the study presented here, we have undertaken extensive single embryo transcriptome profiling to determine the effects of paternal diet on early-embryo gene expression. We profiled hundreds of IVF-derived embryos at various pre-implantation stages generated from sperm isolated from fathers on various diets. These experiments reveal a complexity derived from the transition from oocyte-

dominated expression to embryonic activation, with a number of expression changes driven by paternal diet.

Results

Recent technological advances have enabled capturing the transcriptome of single-cells. Previously, only collections of cells could be profiled, preventing analysis of individual variability. Because of these recent advances, fleeting amounts of input material can be analyzed, and individual variation can be measured. We utilized the Smart-Seq2 protocol to profile the transcriptomes of individual embryos derived by IVF (Figure SA1.1) (Picelli et al., 2014). Smart-Seq2 relies on the template switching capabilities of some MMLV RTs to insert PCR primer binding sites at both the 5' and 3' end of cDNA for subsequent PCR pre-amplification of material. Tagmentation by Tn5 transposase (Nextera, Illumina) subsequently adds deep sequencing adapters to the amplified cDNA, so that libraries can be constructed.

We performed pilot experiments to determine the feasibility of capturing pre-implantation embryo transcriptomes. Pilot experiments revealed that the zona pellucida doesn't need to be removed to capture the pre-implantation embryonic transcriptome (not shown). Sequencing at very low depth (~200,000 mapped reads) recapitulated the findings of previous studies on the early embryonic
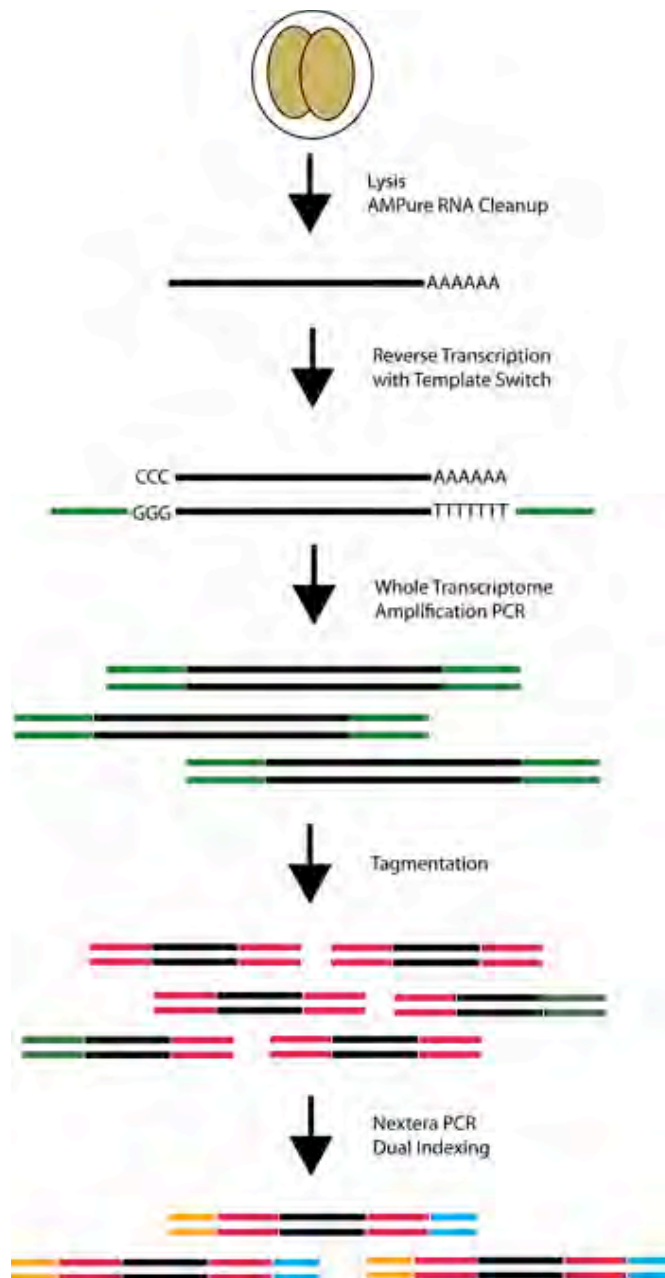
Figure SA1.1

Smart-Seq2 Overview

Embryos are lysed, and RNA is captured using RNA AMPure beads. After AMPure cleanup, reverse transcription is performed with primer that contains a PCR primer binding site and an oligo-switching oligonucleotide with a binding site for the same PCR primer. Single-primer PCR is carried out to amplify material. Tagmentation with Tn5 transposase inserts adapters for library amplification. Final library amplification uses dual indexes.

transcriptome (Deng et al., 2014). These pilot studies affirmed that numerous oocyte-contributed transcripts are present in the 2C embryo (*Spin1*, *Omt2b*, and *Obox5*), while the 4C embryo begins expressing pluripotency markers, such as *Klf2* and *Myc*. Even at very low read depth, 2C embryos are clearly distinct from 4C embryos and morulae (Figure A1.1A). At these depths, diets did not influence the clustering of samples. Subsequently, we tested the ability of greater read depth to give us additional information about the early embryos. Correlations increased between read depths as gene expression increased (Figure A1.1B). Many more genes were captured at high read depths for lowly expressed genes, convincing us that saturation was not achieved at these low read depths. Although relevant information about cell-state is determined at low read depths, we proceeded to achieve higher read depths for each library by running sets of libraries (96-samples per pool) on a HiSeq instead of a MiSeq.

Over 800 libraries were constructed and deep sequenced (TableSA1.1). Numerous combinations of embryos generated from sperm from males on different diets that were collected at different embryonic stages were included in the experiment, including embryos from sperm that fertilized oocytes from different sets of female donors to parse out oocyte contribution to variability. Previously, we had seen that oocytes from females of older mothers affected expression in offspring livers (data not shown), so we attempted to restrict ourselves to using oocytes from females of young mothers (first breeding). As the
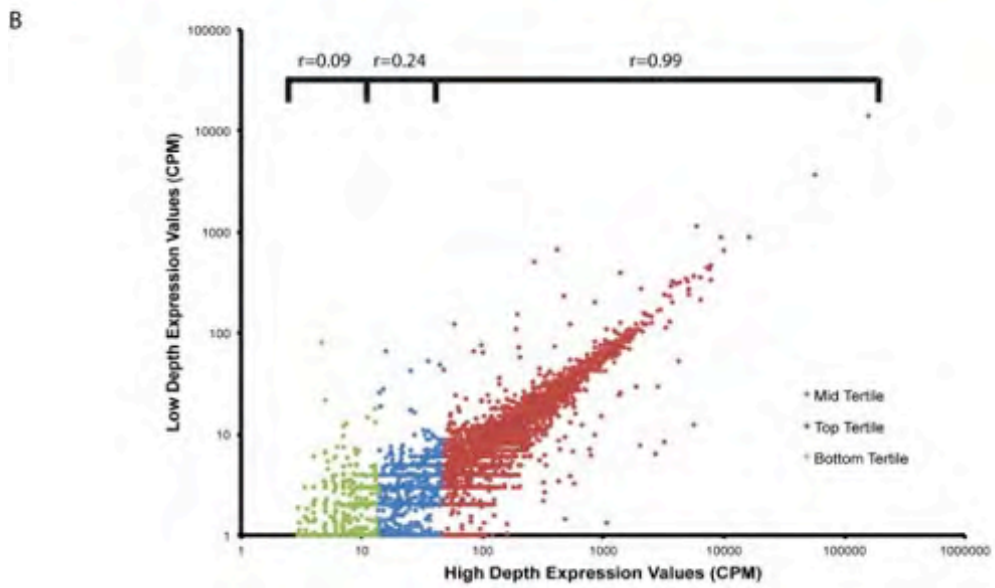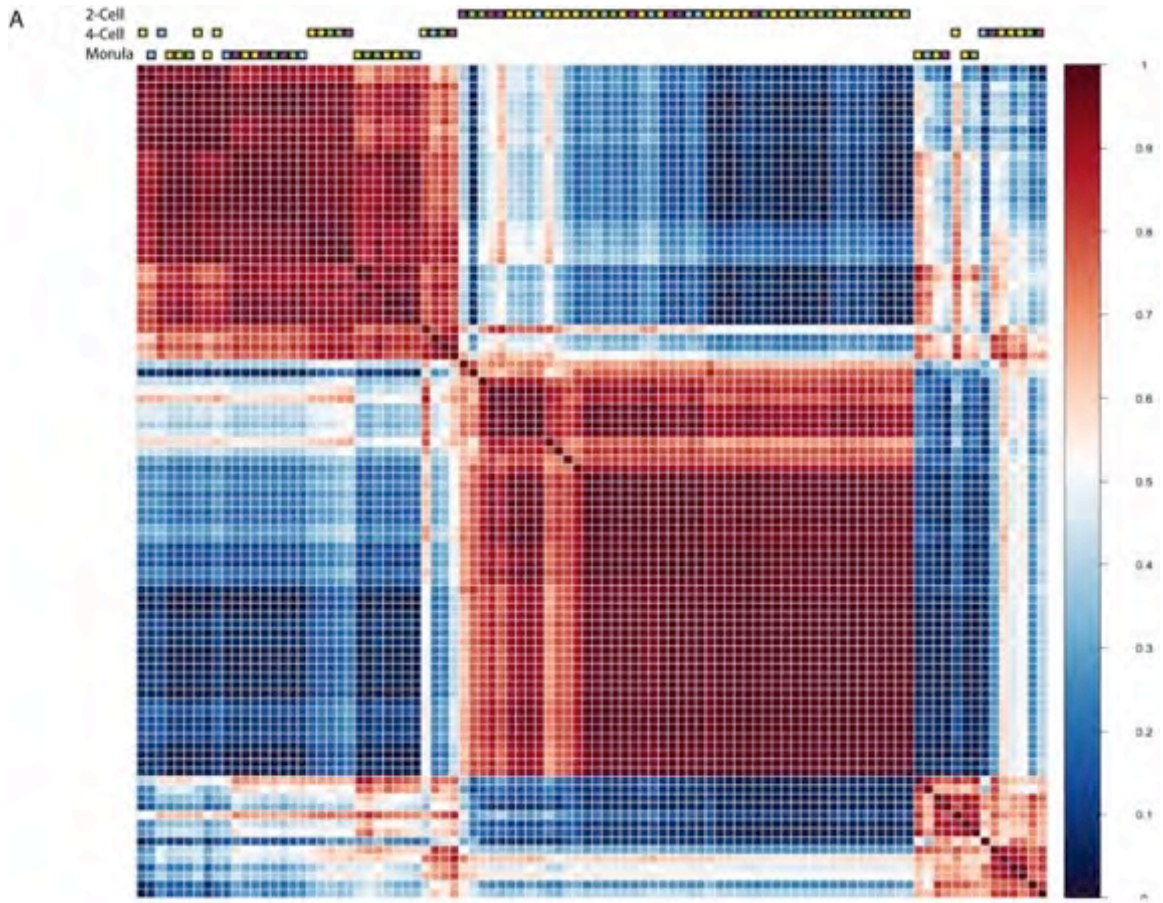
Figure A1.1

Low Depth Sequencing Captures Embryonic Stages

A. Correlation Matrix of pilot experiment. Boxes at the top represent paternal diet (yellow for Control, green for Low Protein, purple for High Fat, and blue for Caloric Restriction) with the row being the embryonic stage.

B. Correlation between a library sequenced at low depth (~200,000 reads) on the y axis and high depth (>1,000,000 reads) on the x axis. Correlations between high depth tertiles and low depth of those same genes shown above the scatter plot.

table depicts, several sample sets had numerous embryos from each stage (2C, 4C, M, and B) and from fathers on different diets, which makes these sets extremely useful for determining kinetics of development for embryos with shared backgrounds.

Principal component analysis (PCA) was performed on the first 241 samples sequenced at high depth. It reveals a trajectory of gene expression changes from the 2-cell stage to the morulae stage based off the genes determining the most variance (PC1), while the third principal component (PC3) segregated these early stage embryos further from blastocysts (Figure A1.2A). Gene ontology analysis finds an enrichment of reproductive annotations for two cell embryos, while 4-cell and morula samples become enriched for metabolic processes and cell cycle annotations. In the gene ontology analysis, blastocysts are enriched for developmental categories. When separating samples out by stage, trends of separation can be seen between different diets, but a clear distinction is not prevalent (Figure A1.2B for morulae stage embryos). This clustering analysis reveals that pre-implantation stages are transcriptionally distinct, while diet provides minimal means to separate staged embryos.

The clustering analysis revealed that paternal diet did not induce substantial perturbations to development. Rather, the effects of diet were subtler, potentially
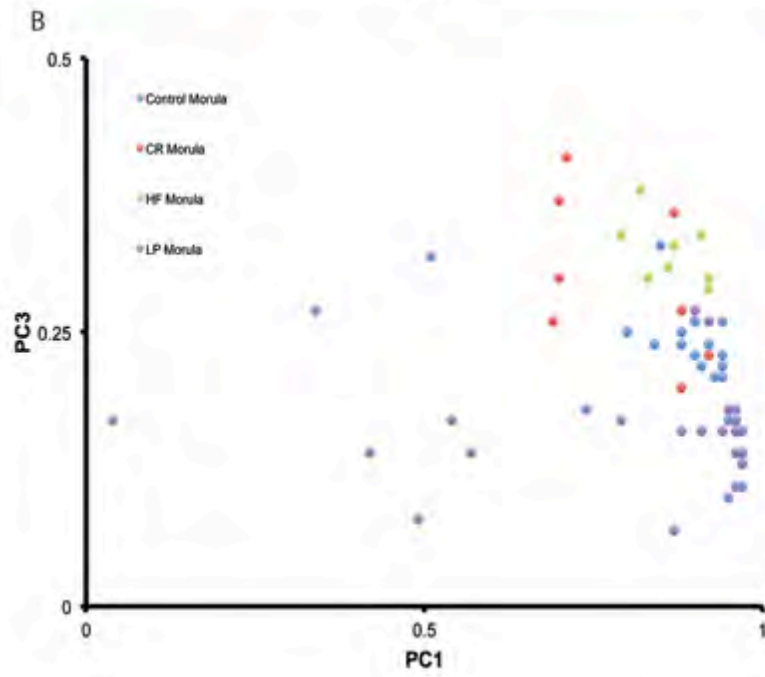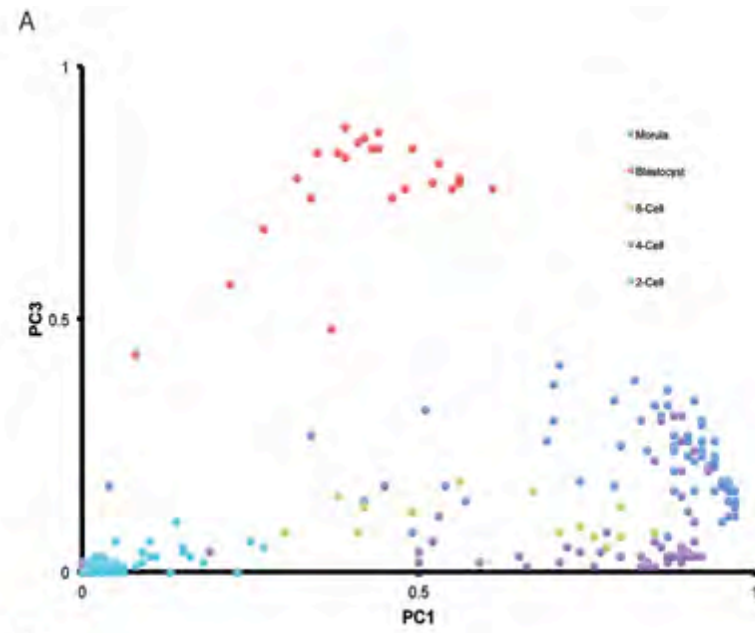
Figure A1.2

Principal Component Analysis of 241 Single Embryo RNA-Seq Samples

A. Scatterplot of PC1 (x axis) and PC3 (y axis) from PCA (using 10 principal components). Embryonic stages are denoted by different color dots. PC2 was not used to plot results against PC1 because it also largely separated 2 cell embryos from every other stage.

B. Scatterplot (as in (A)) of all Morula stage embryos separated by diet. Notice overlap between different diets.

propagated through development to result in offspring metabolic phenotypes (Lane et al., 2014). To find early embryonic transcriptional changes that might result in later metabolic phenotypes, we searched for genes that were altered at the 2-cell stage of development (Figure A1.3A). Interestingly, a number of genes altered by Low Protein diet were located adjacent to MERVL elements, and are controlled by the MERVL LTR (Figure A1.3B-D). MERVL forms chimaeric transcripts with its associated genes. These genes are specifically expressed at the 2C stage and considered to be demonstrative of the totipotent state (Macfarlan et al., 2012). Included in these genes is *Trim43c* (Figure A1.3C), which is only expressed in pre-implantation embryos (Stanghellini et al., 2009). Additionally, some MERVL controlled genes are essential for embryogenesis, such as *Zscan4b*, which is responsible for telomere maintenance in early embryos (Dan et al., 2014). This discovery phase is ongoing, and will be important for determining paternally effected transcripts in the early embryo.
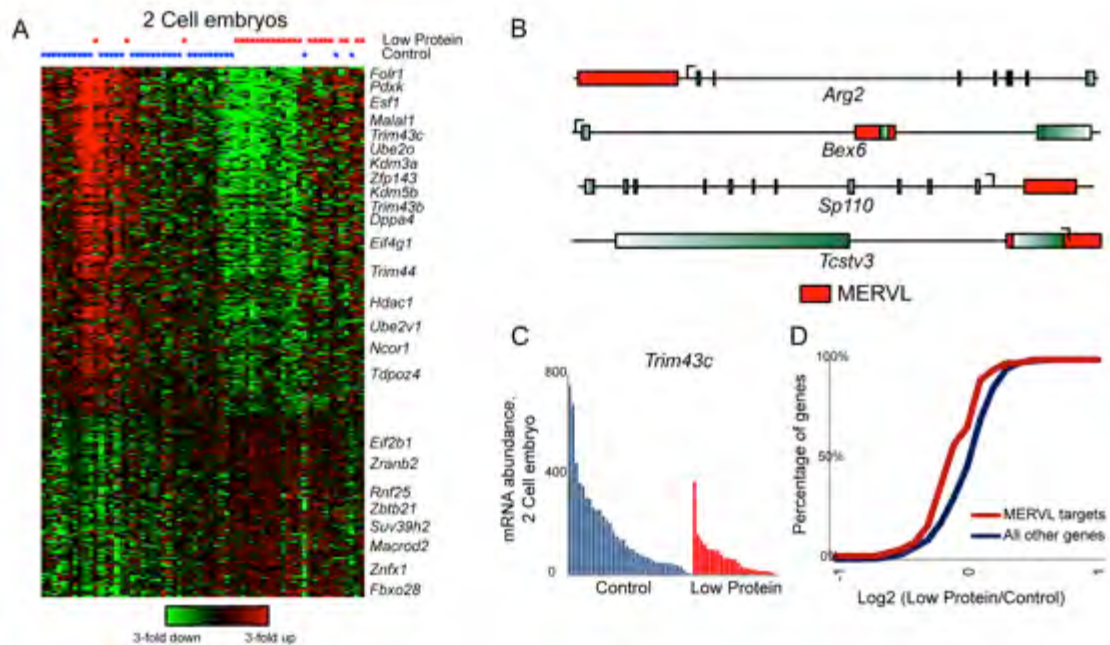
Figure A1.3

Paternal Diet-Induced Expression Changes in 2-cell Embryos

A. Single-embryo RNA-Seq data for 62 2 Cell stage embryos – 37 Control and 25 Low Protein embryos – generated via in vitro fertilization. Data here were zero centered according to embryo batches, and transcripts with >25 ppm average abundance were assessed for significant effects of diet on mRNA abundance.

B. Schematic showing genomic context for four MERVL-associated genes. Other MERVL-associated genes, such as the Tdpoz cluster, are not as close to MERVL LTRs, but instead are located in large MERVL-rich genomic clusters, and have also been shown to be part of the MERVL-regulated gene expression program (Macfarlan et al., 2011).

C. Low Protein embryos exhibit significant repression of Trim43c, a known MERVL target. Bars show Trim43c expression (cpm), for 2C embryos derived from Control or Low Protein sperm, as indicated.

D. Cumulative distribution plots for MERVL targets (defined as those upregulated >2- fold in MERVL-positive ES cells (Macfarlan et al., 2012)), and all remaining genes, showing the percentage of all genes with the average Log2 (Low Protein/Control) as indicated on the x axis. The separation between these curves shows a significant overall shift to lower expression of MERVL targets in 2C embryos derived from Low Protein sperm.

Conclusion

We have shown that the transcriptomes of each stage of pre-implantation development are distinct. These transcriptomes transition from oocyte dominated expression patterns through totipotency to the expression of pluripotent and extra-embryonic markers. Others have shown that individual cells from morulae express either pluripotent or extra-embryonic markers, revealing that these cells have initiated the first differentiation steps (Guo et al., 2010). The cells that contribute to the trophectoderm are located on the exterior of the morula, while those destined to become the inner cell mass are internal.

Our analysis has revealed that diet alters totipotency genes in the early embryo, with many of these being controlled by MERVL elements (Macfarlan et al., 2012). The fact that we see differences in expression of these genes at the 2-cell stage suggests that the kinetics of early embryonic development are altered by paternal diet. Since MERVL elements are silenced by cytosine methylation independent mechanisms in ESC, including Kap1 and G9a, our results intimate that diet alters repression by these factors during early embryogenesis (Maksakova et al., 2013). Diet could possibly alter the transition between cytosine methylation dependent silencing of these elements in sperm to the H3K9me2 mediated early embryonic repression. Interestingly, numerous 2C-specific genes increase expression in TET TKO cells, and hydroxymethylation has been mapped to these

regions, again leading to the inference that the kinetics of reprogramming are altered by paternal diet (Lu et al., 2014). Future studies will be needed to determine if demethylation over MERVL elements following fertilization is changed by paternal diet.

Subsequently, altered developmental kinetics during early embryogenesis can change the ratio of trophectoderm to ICM (Binder et al., 2012; Lee et al., 2015). Reduction of trophectoderm leads to less placentation during uterine development, and has been associated with lower embryonic and birth weights. Therefore, a number of responses to manipulating pre-implantation kinetics may lead to disrupted development, and ultimately metabolic phenotypes in adults. The link between changes in pre-implantation development and later alterations in placentation needs to be investigated in the future.

Methods

In vitro fertilization

In vitro fertilization was performed according to "Manipulating the Mouse Embryo" Second Edition (Hogan, 1994). Superovulated FVB/NJ mice were used as egg donors and sperm was isolated from males fed dietary regimes as above. Isolated sperm were capacitated in HTF for 30 minutes at 37°C. Fertilization took place in 250 mL HTF media covered in mineral oil, pre-gassed in 5% CO_2 at

37°C. Fertilized embryos were developed in KSOM (Millipore - MR-020-P) until the 2-cell stage. Swiss Webster Females between 25 and 35 grams were used as 2-call stage embryo recipients via unilateral oviduct transfer.

Single-Embryo RNA-Seq

Single-Embryo RNA-Seq was performed according to the homemade Smart-Seq2 protocol (Trombetta et al., 2014). Briefly, embryos were lysed in buffer containing TCL (Qiagen #1031576) containing 1% BME for 10 minutes at room temperature. RNA was isolated with 2.2X RNA-SPRI beads (Beckman Coulter #A63987). Reverse transcription was performed on isolated RNA using Superscript II (Life Technologies #18064-014) with additional betaine and $MgCl_2$, using a 3' poly-dT primer that attaches a PCR adapter and a 5' template switching oligonucleotide that attaches a homotypic PCR adapter. PCR amplification used a single PCR primer and KAPA HiFi Hotstart (KAPA #KK2601). PCR cleanup was performed with XP-SPRI (Beckman Coulter #A63881). Product input for Nextera library construction was normalized with measurements from qPCR for *ActB* and DNA concentration from Nanodrop. Library construction was carried out with Nextera XT DNA Sample Prep kits (Illumina #FC-131-1096), except that quarter reactions were used. Final libraries were sequenced on MiSeq and HiSeq machines.

Data Analysis

Adapter sequences were trimmed. Bowtie was used to map paired end reads. RSEM was used to calculate expression levels. R packages were used to perform correlation analysis and PCA analysis. Gene Ontology results from the server at geneontology.org. TTests were used to look for differential gene expression.

BIBLIOGRAPHY

(2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57-74.

Akalin, A., Kormaksson, M., Li, S., Garrett-Bakelman, F.E., Figueroa, M.E., Melnick, A., and Mason, C.E. (2012). methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. Genome biology *13*, R87.

Anderson, L.M., Riffle, L., Wilson, R., Travlos, G.S., Lubomirski, M.S., and Alvord, W.G. (2006). Preconceptional fasting of fathers alters serum glucose in offspring of mice. Nutrition (Burbank, Los Angeles County, Calif) *22*, 327-331.

Anway, M.D., Cupp, A.S., Uzumcu, M., and Skinner, M.K. (2005). Epigenetic transgenerational actions of endocrine disruptors and male fertility. Science *308*, 1466-1469.

Arpanahi, A., Brinkworth, M., Iles, D., Krawetz, S.A., Paradowska, A., Platts, A.E., Saida, M., Steger, K., Tedder, P., and Miller, D. (2009). Endonuclease-sensitive regions of human spermatozoal chromatin are highly enriched in promoter and CTCF binding sequences. Genome research *19*, 1338-1349.

Avital, E., and Jablonka, E. (2000). Animal traditions : behavioural inheritance in evolution (Cambridge, UK ; New York: Cambridge University Press).

Barber, R., Plumb, M.A., Boulton, E., Roux, I., and Dubrova, Y.E. (2002). Elevated mutation rates in the germ line of first- and second-generation offspring of irradiated male mice. Proceedings of the National Academy of Sciences of the United States of America *99*, 6877-6882.

Barlow, D.P., Stoger, R., Herrmann, B.G., Saito, K., and Schweifer, N. (1991). The mouse insulin-like growth factor type-2 receptor is imprinted and closely linked to the Tme locus. Nature *349*, 84-87.

Bartolomei, M.S., Webber, A.L., Brunkow, M.E., and Tilghman, S.M. (1993). Epigenetic mechanisms underlying the imprinting of the mouse H19 gene. Genes Dev *7*, 1663-1673.

Barton, S.C., Surani, M.A., and Norris, M.L. (1984). Role of paternal and maternal genomes in mouse development. Nature *311*, 374-376.

Baubec, T., Ivanek, R., Lienert, F., and Schubeler, D. (2013). Methylation-dependent and -independent genomic targeting principles of the MBD protein family. Cell *153*, 480-492.

Bhutani, N., Brady, J.J., Damian, M., Sacco, A., Corbel, S.Y., and Blau, H.M. (2010). Reprogramming towards pluripotency requires AID-dependent DNA demethylation. Nature *463*, 1042-1047.

Bian, C., and Yu, X. (2014). PGC7 suppresses TET3 for protecting DNA methylation. Nucleic acids research *42*, 2893-2905.

Binder, N.K., Hannan, N.J., and Gardner, D.K. (2012). Paternal diet-induced obesity retards early mouse embryo development, mitochondrial activity and pregnancy health. PloS one *7*, e52304.

Blaschke, K., Ebata, K.T., Karimi, M.M., Zepeda-Martinez, J.A., Goyal, P., Mahapatra, S., Tam, A., Laird, D.J., Hirst, M., Rao, A*., et al.* (2013). Vitamin C induces Tet-dependent DNA demethylation and a blastocyst-like state in ES cells. Nature *500*, 222-226.

Blewitt, M.E., Vickaryous, N.K., Paldi, A., Koseki, H., and Whitelaw, E. (2006). Dynamic reprogramming of DNA methylation at an epigenetically sensitive allele in mice. PLoS genetics *2*, e49.

Bock, C., Beerman, I., Lien, W.H., Smith, Z.D., Gu, H., Boyle, P., Gnirke, A., Fuchs, E., Rossi, D.J., and Meissner, A. (2012). DNA methylation dynamics during in vivo differentiation of blood and skin stem cells. Molecular cell *47*, 633-647.

Borgel, J., Guibert, S., Li, Y., Chiba, H., Schubeler, D., Sasaki, H., Forne, T., and Weber, M. (2010). Targets and dynamics of promoter DNA methylation during early mouse development. Nature genetics *42*, 1093-1100.

Bourc'his, D., Xu, G.L., Lin, C.S., Bollman, B., and Bestor, T.H. (2001). Dnmt3L and the establishment of maternal genomic imprints. Science *294*, 2536-2539.

Boylston, W.H., Gerstner, A., DeFord, J.H., Madsen, M., Flurkey, K., Harrison, D.E., and Papaconstantinou, J. (2004). Altered cholesterologenic and lipogenic transcriptional profile in livers of aging Snell dwarf (Pit1dw/dwJ) mice. Aging cell *3*, 283-296.

Brahmachary, M., Guilmatre, A., Quilez, J., Hasson, D., Borel, C., Warburton, P., and Sharp, A.J. (2014). Digital genotyping of macrosatellites and multicopy genes reveals novel biological functions associated with copy number variation of large tandem repeats. PLoS genetics *10*, e1004418.

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

Brinkman, A.B., Gu, H., Bartels, S.J., Zhang, Y., Matarese, F., Simmer, F., Marks, H., Bock, C., Gnirke, A., Meissner, A*., et al.* (2012). Sequential ChIP-bisulfite sequencing enables direct genome-scale investigation of chromatin and DNA methylation cross-talk. Genome research *22*, 1128-1138.

Bromfield, J.J., Schjenken, J.E., Chin, P.Y., Care, A.S., Jasper, M.J., and Robertson, S.A. (2014). Maternal tract factors contribute to paternal seminal fluid impact on metabolic phenotype in offspring. Proceedings of the National Academy of Sciences of the United States of America *111*, 2200-2205.

Brown, M.S., and Goldstein, J.L. (1997). The SREBP pathway: regulation of cholesterol metabolism by proteolysis of a membrane-bound transcription factor. Cell *89*, 331-340.

Brykczynska, U., Hisano, M., Erkek, S., Ramos, L., Oakeley, E.J., Roloff, T.C., Beisel, C., Schubeler, D., Stadler, M.B., and Peters, A.H. (2010). Repressive and active histone methylation mark distinct promoters in human and mouse spermatozoa. Nature structural & molecular biology *17*, 679-687.

Campanero, M.R., Armstrong, M.I., and Flemington, E.K. (2000). CpG methylation as a mechanism for the regulation of E2F activity. Proceedings of the National Academy of Sciences of the United States of America *97*, 6481-6486.

Carone, B.R., Fauquier, L., Habib, N., Shea, J.M., Hart, C.E., Li, R., Bock, C., Li, C., Gu, H., Zamore, P.D*., et al.* (2010). Paternally induced transgenerational environmental reprogramming of metabolic gene expression in mammals. Cell *143*, 1084-1096.

Cavalli, G., and Paro, R. (1998). The Drosophila Fab-7 chromosomal element conveys epigenetic inheritance during mitosis and meiosis. Cell *93*, 505-518.

Champagne, F., and Meaney, M.J. (2001). Like mother, like daughter: evidence for non-genomic transmission of parental behavior and stress responsivity. Progress in brain research *133*, 287-302.

Chandler, V.L., and Stam, M. (2004). Chromatin conversations: mechanisms and implications of paramutation. Nature reviews Genetics *5*, 532-544.

Chen, M., Wu, L., Zhao, J., Wu, F., Davies, M.J., Wittert, G.A., Norman, R.J., Robker, R.L., and Heilbronn, L.K. (2014). Altered glucose metabolism in mouse and humans conceived by IVF. Diabetes *63*, 3189-3198.

Chodavarapu, R.K., Feng, S., Bernatavichute, Y.V., Chen, P.Y., Stroud, H., Yu, Y., Hetzel, J.A., Kuo, F., Kim, J., Cokus, S.J.*, et al.* (2010). Relationship between nucleosome positioning and DNA methylation. Nature *466*, 388-392.

Chong, S., Vickaryous, N., Ashe, A., Zamudio, N., Youngson, N., Hemley, S., Stopka, T., Skoultchi, A., Matthews, J., Scott, H.S.*, et al.* (2007). Modifiers of epigenetic reprogramming show paternal effects in the mouse. Nature genetics *39*, 614-622.

Chuang, L.S., Ian, H.I., Koh, T.W., Ng, H.H., Xu, G., and Li, B.F. (1997). Human DNA-(cytosine-5) methyltransferase-PCNA complex as a target for p21WAF1. Science *277*, 1996-2000.

Clouaire, T., Webb, S., Skene, P., Illingworth, R., Kerr, A., Andrews, R., Lee, J.H., Skalnik, D., and Bird, A. (2012). Cfp1 integrates both CpG content and gene activity for accurate H3K4me3 deposition in embryonic stem cells. Genes Dev *26*, 1714-1728.

Colot, V., Maloisel, L., and Rossignol, J.L. (1996). Interchromosomal transfer of epigenetic states in Ascobolus: transfer of DNA methylation is mechanistically related to homologous recombination. Cell *86*, 855-864.

Conine, C.C., Moresco, J.J., Gu, W., Shirayama, M., Conte, D., Jr., Yates, J.R., 3rd, and Mello, C.C. (2013). Argonautes promote male fertility and provide a paternal memory of germline gene expression in C. elegans. Cell *155*, 1532-1544.

Cortellino, S., Xu, J., Sannai, M., Moore, R., Caretti, E., Cigliano, A., Le Coz, M., Devarajan, K., Wessels, A., Soprano, D.*, et al.* (2011). Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. Cell *146*, 67-79.

Crews, D., Gore, A.C., Hsu, T.S., Dangleben, N.L., Spinetta, M., Schallert, T., Anway, M.D., and Skinner, M.K. (2007). Transgenerational epigenetic imprints on mate preference. Proceedings of the National Academy of Sciences of the United States of America *104*, 5942-5946.

Cropley, J.E., Suter, C.M., Beckman, K.B., and Martin, D.I. (2006). Germ-line epigenetic modification of the murine A vy allele by nutritional supplementation. Proceedings of the National Academy of Sciences of the United States of America *103*, 17308-17312.

Dan, J., Liu, Y., Liu, N., Chiourea, M., Okuka, M., Wu, T., Ye, X., Mou, C., Wang, L., Wang, L.*, et al.* (2014). Rif1 maintains telomere length homeostasis of ESCs by mediating heterochromatin silencing. Developmental cell *29*, 7-19.

Danchin, E., Charmantier, A., Champagne, F.A., Mesoudi, A., Pujol, B., and Blanchet, S. (2011). Beyond DNA: integrating inclusive inheritance into an extended theory of evolution. Nature reviews Genetics *12*, 475-486.

Daxinger, L., Harten, S.K., Oey, H., Epp, T., Isbel, L., Huang, E., Whitelaw, N., Apedaile, A., Sorolla, A., Yong, J*., et al.* (2013). An ENU mutagenesis screen identifies novel and known genes involved in epigenetic processes in the mouse. Genome biology *14*, R96.

De Carvalho, D.D., You, J.S., and Jones, P.A. (2010). DNA methylation and cellular reprogramming. Trends in cell biology *20*, 609-617.

Deng, Q., Ramskold, D., Reinius, B., and Sandberg, R. (2014). Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. Science *343*, 193-196.

Dias, B.G., and Ressler, K.J. (2014). Parental olfactory experience influences behavior and neural structure in subsequent generations. Nature neuroscience *17*, 89-96.

Diehn, M., Alizadeh, A.A., Rando, O.J., Liu, C.L., Stankunas, K., Botstein, D., Crabtree, G.R., and Brown, P.O. (2002). Genomic expression programs and the integration of the CD28 costimulatory signal in T cell activation. Proceedings of the National Academy of Sciences of the United States of America *99*, 11796-11801.

Dietz, D.M., Laplant, Q., Watts, E.L., Hodes, G.E., Russo, S.J., Feng, J., Oosting, R.S., Vialou, V., and Nestler, E.J. (2011). Paternal transmission of stress-induced pathologies. Biological psychiatry *70*, 408-414.

Epsztejn-Litman, S., Feldman, N., Abu-Remaileh, M., Shufaro, Y., Gerson, A., Ueda, J., Deplus, R., Fuks, F., Shinkai, Y., Cedar, H*., et al.* (2008). De novo DNA methylation promoted by G9a prevents reprogramming of embryonically silenced genes. Nature structural & molecular biology *15*, 1176-1183.

Feinberg, A.P., and Irizarry, R.A. (2010). Evolution in health and medicine Sackler colloquium: Stochastic epigenetic variation as a driving force of development, evolutionary adaptation, and disease. Proceedings of the National Academy of Sciences of the United States of America *107 Suppl 1*, 1757-1764.

Feldman, N., Gerson, A., Fang, J., Li, E., Zhang, Y., Shinkai, Y., Cedar, H., and Bergman, Y. (2006). G9a-mediated irreversible epigenetic inactivation of Oct-3/4 during early embryogenesis. Nature cell biology *8*, 188-194.

Feldmann, A., Ivanek, R., Murr, R., Gaidatzis, D., Burger, L., and Schubeler, D. (2013). Transcription factor occupancy can mediate active turnover of DNA methylation at regulatory regions. PLoS genetics *9*, e1003994.

Feuer, S.K., Liu, X., Donjacour, A., Lin, W., Simbulan, R.K., Giritharan, G., Piane, L.D., Kolahi, K., Ameri, K., Maltepe, E.*, et al.* (2014). Use of a mouse in vitro fertilization model to understand the developmental origins of health and disease hypothesis. Endocrinology *155*, 1956-1969.

Finnegan, E.J., Genger, R.K., Kovac, K., Peacock, W.J., and Dennis, E.S. (1998). DNA methylation and the promotion of flowering by vernalization. Proceedings of the National Academy of Sciences of the United States of America *95*, 5824-5829.

Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., and Mello, C.C. (1998). Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. Nature *391*, 806-811.

Fricke, C., Bretman, A., and Chapman, T. (2008). Adult male nutrition and reproductive success in Drosophila melanogaster. Evolution; international journal of organic evolution *62*, 3170-3177.

Gal-Yam, E.N., Egger, G., Iniguez, L., Holster, H., Einarsson, S., Zhang, X., Lin, J.C., Liang, G., Jones, P.A., and Tanay, A. (2008). Frequent switching of Polycomb repressive marks and DNA hypermethylation in the PC3 prostate cancer cell line. Proceedings of the National Academy of Sciences of the United States of America *105*, 12979-12984.

Ghildiyal, M., Seitz, H., Horwich, M.D., Li, C., Du, T., Lee, S., Xu, J., Kittler, E.L., Zapp, M.L., Weng, Z.*, et al.* (2008). Endogenous siRNAs derived from transposons and mRNAs in Drosophila somatic cells. Science *320*, 1077-1081.

Gibbons, J.G., Branco, A.T., Godinho, S.A., Yu, S., and Lemos, B. (2015). Concerted copy number variation balances ribosomal DNA dosage in human and mouse genomes. Proceedings of the National Academy of Sciences of the United States of America *112*, 2485-2490.

Gibbons, J.G., Branco, A.T., Yu, S., and Lemos, B. (2014). Ribosomal DNA copy number is coupled with gene expression variation and mitochondrial abundance in humans. Nature communications *5*, 4850.

Gifford, C.A., Ziller, M.J., Gu, H., Trapnell, C., Donaghey, J., Tsankov, A., Shalek, A.K., Kelley, D.R., Shishkin, A.A., Issner, R.*, et al.* (2013). Transcriptional and epigenetic dynamics during specification of human embryonic stem cells. Cell *153*, 1149-1163.

Gill, S., Stevenson, J., Kristiana, I., and Brown, A.J. (2011). Cholesterol-dependent degradation of squalene monooxygenase, a control point in cholesterol synthesis beyond HMG-CoA reductase. Cell metabolism *13*, 260-273.

Goll, M.G., and Bestor, T.H. (2005). Eukaryotic cytosine methyltransferases. Annual review of biochemistry *74*, 481-514.

Gray, S., Wang, B., Orihuela, Y., Hong, E.G., Fisch, S., Haldar, S., Cline, G.W., Kim, J.K., Peroni, O.D., Kahn, B.B.*, et al.* (2007). Regulation of gluconeogenesis by Kruppel-like factor 15. Cell metabolism *5*, 305-312.

Greer, E.L., Maures, T.J., Ucar, D., Hauswirth, A.G., Mancini, E., Lim, J.P., Benayoun, B.A., Shi, Y., and Brunet, A. (2011). Transgenerational epigenetic inheritance of longevity in Caenorhabditis elegans. Nature *479*, 365-371.

Grummt, I. (2013). The nucleolus-guardian of cellular homeostasis and genome integrity. Chromosoma *122*, 487-497.

Grummt, I., and Pikaard, C.S. (2003). Epigenetic silencing of RNA polymerase I transcription. Nature reviews Molecular cell biology *4*, 641-649.

Guo, G., Huss, M., Tong, G.Q., Wang, C., Li Sun, L., Clarke, N.D., and Robson, P. (2010). Resolution of cell fate decisions revealed by single-cell gene expression analysis from zygote to blastocyst. Developmental cell *18*, 675-685.

Hackett, J.A., Sengupta, R., Zylicz, J.J., Murakami, K., Lee, C., Down, T.A., and Surani, M.A. (2013). Germline DNA demethylation dynamics and imprint erasure through 5-hydroxymethylcytosine. Science *339*, 448-452.

Hales, C.N., and Barker, D.J. (2001). The thrifty phenotype hypothesis. British medical bulletin *60*, 5-20.

Hammoud, S.S., Nix, D.A., Zhang, H., Purwar, J., Carrell, D.T., and Cairns, B.R. (2009). Distinctive chromatin in human sperm packages genes for embryo development. Nature *460*, 473-478.

Hansen, K.D., Timp, W., Bravo, H.C., Sabunciyan, S., Langmead, B., McDonald, O.G., Wen, B., Wu, H., Liu, Y., Diep, D.*, et al.* (2011). Increased methylation variation in epigenetic domains across cancer types. Nature genetics *43*, 768-775.

Hark, A.T., Schoenherr, C.J., Katz, D.J., Ingram, R.S., Levorse, J.M., and Tilghman, S.M. (2000). CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. Nature *405*, 486-489.

Harris, A., and Seckl, J. (2011). Glucocorticoids, prenatal stress and the programming of disease. Hormones and behavior *59*, 279-289.

Hashimoto, H., Liu, Y., Upadhyay, A.K., Chang, Y., Howerton, S.B., Vertino, P.M., Zhang, X., and Cheng, X. (2012). Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. Nucleic acids research *40*, 4841-4849.

Hathaway, N.A., Bell, O., Hodges, C., Miller, E.L., Neel, D.S., and Crabtree, G.R. (2012). Dynamics and memory of heterochromatin in living cells. Cell *149*, 1447-1460.

Heard, E., and Martienssen, R.A. (2014). Transgenerational epigenetic inheritance: myths and mechanisms. Cell *157*, 95-109.

Heijmans, B.T., Tobi, E.W., Stein, A.D., Putter, H., Blauw, G.J., Susser, E.S., Slagboom, P.E., and Lumey, L.H. (2008). Persistent epigenetic differences associated with prenatal exposure to famine in humans. Proceedings of the National Academy of Sciences of the United States of America *105*, 17046-17049.

Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A.*, et al.* (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nature genetics *39*, 311-318.

Hendrickson, D.G., Hogan, D.J., Herschlag, D., Ferrell, J.E., and Brown, P.O. (2008). Systematic identification of mRNAs recruited to argonaute 2 by specific microRNAs and corresponding changes in transcript abundance. PloS one *3*, e2126.

Hindson, B.J., Ness, K.D., Masquelier, D.A., Belgrader, P., Heredia, N.J., Makarewicz, A.J., Bright, I.J., Lucero, M.Y., Hiddessen, A.L., Legler, T.C.*, et al.* (2011). High-throughput droplet digital PCR system for absolute quantitation of DNA copy number. Analytical chemistry *83*, 8604-8610.

Hirasawa, R., Chiba, H., Kaneda, M., Tajima, S., Li, E., Jaenisch, R., and Sasaki, H. (2008). Maternal and zygotic Dnmt1 are necessary and sufficient for the maintenance of DNA methylation imprints during preimplantation development. Genes Dev *22*, 1607-1616.

Hogan, B. (1994). Manipulating the mouse embryo : a laboratory manual, 2nd edn (Plainview, N.Y.: Cold Spring Harbor Laboratory Press).

Holliday, R. (1987). The inheritance of epigenetic defects. Science *238*, 163-170.

Hollox, E.J., Huffmeier, U., Zeeuwen, P.L., Palla, R., Lascorz, J., Rodijk-Olthuis, D., van de Kerkhof, P.C., Traupe, H., de Jongh, G., den Heijer, M., *et al.* (2008). Psoriasis is associated with increased beta-defensin genomic copy number. Nature genetics *40*, 23-25.

Holzer, M., Trieb, M., Konya, V., Wadsack, C., Heinemann, A., and Marsche, G. (2013). Aging affects high-density lipoprotein composition and function. Biochimica et biophysica acta *1831*, 1442-1448.

Horsthemke, B., and Wagstaff, J. (2008). Mechanisms of imprinting of the Prader-Willi/Angelman region. American journal of medical genetics Part A *146a*, 2041-2052.

Horton, J.D., Shah, N.A., Warrington, J.A., Anderson, N.N., Park, S.W., Brown, M.S., and Goldstein, J.L. (2003). Combined analysis of oligonucleotide microarray data from transgenic and knockout mice identifies direct SREBP target genes. Proceedings of the National Academy of Sciences of the United States of America *100*, 12027-12032.

Horton, T.H. (2005). Fetal origins of developmental plasticity: animal models of induced life history variation. American journal of human biology : the official journal of the Human Biology Council *17*, 34-43.

Hu, J.L., Zhou, B.O., Zhang, R.R., Zhang, K.L., Zhou, J.Q., and Xu, G.L. (2009). The N-terminus of histone H3 is required for de novo DNA methylation in chromatin. Proceedings of the National Academy of Sciences of the United States of America *106*, 22187-22192.

Hu, S., Wan, J., Su, Y., Song, Q., Zeng, Y., Nguyen, H.N., Shin, J., Cox, E., Rho, H.S., Woodard, C., *et al.* (2013). DNA methylation presents distinct binding sites for human transcription factors. eLife *2*, e00726.

Ieda, M., Fu, J.D., Delgado-Olguin, P., Vedantham, V., Hayashi, Y., Bruneau, B.G., and Srivastava, D. (2010). Direct reprogramming of fibroblasts into functional cardiomyocytes by defined factors. Cell *142*, 375-386.

Irizarry, R.A., Ladd-Acosta, C., Wen, B., Wu, Z., Montano, C., Onyango, P., Cui, H., Gabo, K., Rongione, M., Webster, M., *et al.* (2009). The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. Nature genetics *41*, 178-186.

Jablonka, E., and Lamb, M.J. (1995). Epigenetic inheritance and evolution : the Lamarckian dimension (Oxford ; New York: Oxford University Press).

Jacinto, F.V., Ballestar, E., and Esteller, M. (2008). Methyl-DNA immunoprecipitation (MeDIP): hunting down the DNA methylome. BioTechniques *44*, 35, 37, 39 passim.

Jackson-Grusby, L., Beard, C., Possemato, R., Tudor, M., Fambrough, D., Csankovszki, G., Dausman, J., Lee, P., Wilson, C., Lander, E.*, et al.* (2001). Loss of genomic methylation causes p53-dependent apoptosis and epigenetic deregulation. Nature genetics *27*, 31-39.

Jasielec, J., Saloura, V., and Godley, L.A. (2014). The mechanistic role of DNA methylation in myeloid leukemogenesis. Leukemia *28*, 1765-1773.

Jerome, T., Laurie, P., Louis, B., and Pierre, C. (2007). Enjoy the Silence: The Story of let-7 MicroRNA and Cancer. Current genomics *8*, 229-233.

Ji, H., Ehrlich, L.I., Seita, J., Murakami, P., Doi, A., Lindau, P., Lee, H., Aryee, M.J., Irizarry, R.A., Kim, K.*, et al.* (2010). Comprehensive methylome map of lineage commitment from haematopoietic progenitors. Nature *467*, 338-342.

Jiang, J., Gusev, Y., Aderca, I., Mettler, T.A., Nagorney, D.M., Brackett, D.J., Roberts, L.R., and Schmittgen, T.D. (2008). Association of MicroRNA expression in hepatocellular carcinomas with hepatitis infection, cirrhosis, and patient survival. Clinical cancer research : an official journal of the American Association for Cancer Research *14*, 419-427.

Jimenez-Chillaron, J.C., Isganaitis, E., Charalambous, M., Gesta, S., Pentinat-Pelegrin, T., Faucette, R.R., Otis, J.P., Chow, A., Diaz, R., Ferguson-Smith, A.*, et al.* (2009). Intergenerational transmission of glucose intolerance and obesity by in utero undernutrition in mice. Diabetes *58*, 460-468.

Kaati, G., Bygren, L.O., and Edvinsson, S. (2002). Cardiovascular and diabetes mortality determined by nutrition during parents' and grandparents' slow growth period. European journal of human genetics : EJHG *10*, 682-688.

Kaelin, W.G., Jr., and McKnight, S.L. (2013). Influence of metabolism on epigenetics and disease. Cell *153*, 56-69.

Kaplan, N., Moore, I.K., Fondufe-Mittendorf, Y., Gossett, A.J., Tillo, D., Field, Y., LeProust, E.M., Hughes, T.R., Lieb, J.D., Widom, J.*, et al.* (2009). The DNA-encoded nucleosome organization of a eukaryotic genome. Nature *458*, 362-366.

Kato, Y., Kaneda, M., Hata, K., Kumaki, K., Hisano, M., Kohara, Y., Okano, M., Li, E., Nozaki, M., and Sasaki, H. (2007). Role of the Dnmt3 family in de novo methylation of imprinted and repetitive sequences during male germ cell development in the mouse. Human molecular genetics *16*, 2272-2280.

Klose, R.J., and Bird, A.P. (2006). Genomic DNA methylation: the mark and its mediators. Trends in biochemical sciences *31*, 89-97.

Lane, M., Robker, R.L., and Robertson, S.A. (2014). Parenting from before conception. Science *345*, 756-760.

Lane, N., Dean, W., Erhardt, S., Hajkova, P., Surani, A., Walter, J., and Reik, W. (2003). Resistance of IAPs to methylation reprogramming may provide a mechanism for epigenetic inheritance in the mouse. Genesis (New York, NY : 2000) *35*, 88-93.

Le Guezennec, X., Vermeulen, M., Brinkman, A.B., Hoeijmakers, W.A., Cohen, A., Lasonder, E., and Stunnenberg, H.G. (2006). MBD2/NuRD and MBD3/NuRD, two distinct complexes with different biochemical and functional properties. Mol Cell Biol *26*, 843-851.

Lee, J.T., and Bartolomei, M.S. (2013). X-inactivation, imprinting, and long noncoding RNAs in health and disease. Cell *152*, 1308-1323.

Lee, Y.S., Thouas, G.A., and Gardner, D.K. (2015). Developmental kinetics of cleavage stage mouse embryos are related to their subsequent carbohydrate and amino acid utilization at the blastocyst stage. Human reproduction (Oxford, England) *30*, 543-552.

Lehnertz, B., Ueda, Y., Derijck, A.A.H.A., Braunschweig, U., Perez-Burgos, L., Kubicek, S., Chen, T., Li, E., Jenuwein, T., and Peters, A.H.F.M. (2003). Suv39h-Mediated Histone H3 Lysine 9 Methylation Directs DNA Methylation to Major Satellite Repeats at Pericentric Heterochromatin. Current Biology *13*, 1192-1200.

Li, E., Beard, C., and Jaenisch, R. (1993). Role for DNA methylation in genomic imprinting. Nature *366*, 362-365.

Li, H., Rauch, T., Chen, Z.X., Szabo, P.E., Riggs, A.D., and Pfeifer, G.P. (2006). The histone methyltransferase SETDB1 and the DNA methyltransferase DNMT3A interact directly and localize to promoters silenced in cancer cells. The Journal of biological chemistry *281*, 19489-19500.

Libby, R.T., Hagerman, K.A., Pineda, V.V., Lau, R., Cho, D.H., Baccam, S.L., Axford, M.M., Cleary, J.D., Moore, J.M., Sopher, B.L.*, et al.* (2008). CTCF cis-regulates trinucleotide repeat instability in an epigenetic manner: a novel basis for mutational hot spot determination. PLoS genetics *4*, e1000257.

Lim, J.P., and Brunet, A. (2013). Bridging the transgenerational gap with epigenetic memory. Trends in genetics : TIG *29*, 176-186.

Lin, C.J., Koh, F.M., Wong, P., Conti, M., and Ramalho-Santos, M. (2014). Hira-mediated H3.3 incorporation is required for DNA replication and ribosomal RNA transcription in the mouse zygote. Developmental cell *30*, 268-279.

Liu, X., Gao, Q., Li, P., Zhao, Q., Zhang, J., Li, J., Koseki, H., and Wong, J. (2013). UHRF1 targets DNMT1 for DNA methylation through cooperative binding of hemi-methylated DNA and methylated H3K9. Nature communications *4*, 1563.

Lu, F., Liu, Y., Jiang, L., Yamaguchi, S., and Zhang, Y. (2014). Role of Tet proteins in enhancer activity and telomere elongation. Genes Dev *28*, 2103-2119.

Lumey, L.H., Stein, A.D., Kahn, H.S., van der Pal-de Bruin, K.M., Blauw, G.J., Zybert, P.A., and Susser, E.S. (2007). Cohort profile: the Dutch Hunger Winter families study. International journal of epidemiology *36*, 1196-1204.

Macfarlan, T.S., Gifford, W.D., Driscoll, S., Lettieri, K., Rowe, H.M., Bonanomi, D., Firth, A., Singer, O., Trono, D., and Pfaff, S.L. (2012). Embryonic stem cell potency fluctuates with endogenous retrovirus activity. Nature *487*, 57-63.

Madsen, M.A., Hsieh, C.C., Boylston, W.H., Flurkey, K., Harrison, D., and Papaconstantinou, J. (2004). Altered oxidative stress response of the long-lived Snell dwarf mouse. Biochemical and biophysical research communications *318*, 998-1005.

Maksakova, I.A., Thompson, P.J., Goyal, P., Jones, S.J., Singh, P.B., Karimi, M.M., and Lorincz, M.C. (2013). Distinct roles of KAP1, HP1 and G9a/GLP in silencing of the two-cell-specific retrotransposon MERVL in mouse ES cells. Epigenetics & chromatin *6*, 15.

Manolio, T.A., Collins, F.S., Cox, N.J., Goldstein, D.B., Hindorff, L.A., Hunter, D.J., McCarthy, M.I., Ramos, E.M., Cardon, L.R., Chakravarti, A.*, et al.* (2009). Finding the missing heritability of complex diseases. Nature *461*, 747-753.

Mathieu, O., Reinders, J., Caikovski, M., Smathajitt, C., and Paszkowski, J. (2007). Transgenerational stability of the Arabidopsis epigenome is coordinated by CG methylation. Cell *130*, 851-862.

McLay, D.W., and Clarke, H.J. (2003). Remodelling the paternal chromatin at fertilization in mammals. Reproduction *125*, 625-633.

McMurray, C.T. (2010). Mechanisms of trinucleotide repeat instability during human development. Nature reviews Genetics *11*, 786-799.

McPherson, N.O., Bakos, H.W., Owens, J.A., Setchell, B.P., and Lane, M. (2013). Improving metabolic health in obese male mice via diet and exercise restores embryo development and fetal growth. PloS one *8*, e71459.

McRae, A.F., Powell, J.E., Henders, A.K., Bowdler, L., Hemani, G., Shah, S., Painter, J.N., Martin, N.G., Visscher, P.M., and Montgomery, G.W. (2014). Contribution of genetic variation to transgenerational inheritance of DNA methylation. Genome biology *15*, R73.

Meaney, M.J., Szyf, M., and Seckl, J.R. (2007). Epigenetic mechanisms of perinatal programming of hypothalamic-pituitary-adrenal function and health. Trends in molecular medicine *13*, 269-277.

Meissner, A., Mikkelsen, T.S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B.E., Nusbaum, C., Jaffe, D.B.*, et al.* (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature *454*, 766-770.

Messerschmidt, D.M., Knowles, B.B., and Solter, D. (2014). DNA methylation dynamics during epigenetic reprogramming in the germline and preimplantation embryos. Genes Dev *28*, 812-828.

Migeon, B.R., Jan de Beur, S., and Axelman, J. (1989). Frequent derepression of G6PD and HPRT on the marsupial inactive X chromosome associated with cell proliferation in vitro. Experimental cell research *182*, 597-609.

Mikkelsen, T.S., Hanna, J., Zhang, X., Ku, M., Wernig, M., Schorderet, P., Bernstein, B.E., Jaenisch, R., Lander, E.S., and Meissner, A. (2008). Dissecting direct reprogramming through integrative genomic analysis. Nature *454*, 49-55.

Miller, B.H., McDearmon, E.L., Panda, S., Hayes, K.R., Zhang, J., Andrews, J.L., Antoch, M.P., Walker, J.R., Esser, K.A., Hogenesch, J.B.*, et al.* (2007). Circadian and CLOCK-controlled regulation of the mouse transcriptome and cell proliferation. Proceedings of the National Academy of Sciences of the United States of America *104*, 3342-3347.

Mohandas, T., Sparkes, R.S., and Shapiro, L.J. (1981). Reactivation of an inactive human X chromosome: evidence for X inactivation by DNA methylation. Science *211*, 393-396.

Molaro, A., Hodges, E., Fang, F., Song, Q., McCombie, W.R., Hannon, G.J., and Smith, A.D. (2011). Sperm methylation profiles reveal features of epigenetic inheritance and evolution in primates. Cell *146*, 1029-1041.

Molinier, J., Ries, G., Zipfel, C., and Hohn, B. (2006). Transgeneration memory of stress in plants. Nature *442*, 1046-1049.

Moore, T., and Haig, D. (1991). Genomic imprinting in mammalian development: a parental tug-of-war. Trends in genetics : TIG *7*, 45-49.

Morgan, H.D., Sutherland, H.G., Martin, D.I., and Whitelaw, E. (1999). Epigenetic inheritance at the agouti locus in the mouse. Nature genetics *23*, 314-318.

Muotri, A.R., Marchetto, M.C., Coufal, N.G., Oefner, R., Yeo, G., Nakashima, K., and Gage, F.H. (2010). L1 retrotransposition in neurons is modulated by MeCP2. Nature *468*, 443-446.

Nakamura, T., Liu, Y.J., Nakashima, H., Umehara, H., Inoue, K., Matoba, S., Tachibana, M., Ogura, A., Shinkai, Y., and Nakano, T. (2012). PGC7 binds histone H3K9me2 to protect against conversion of 5mC to 5hmC in early embryos. Nature *486*, 415-419.

Nan, X., Ng, H.H., Johnson, C.A., Laherty, C.D., Turner, B.M., Eisenman, R.N., and Bird, A. (1998). Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. Nature *393*, 386-389.

Neri, F., Krepelova, A., Incarnato, D., Maldotti, M., Parlato, C., Galvagni, F., Matarese, F., Stunnenberg, H.G., and Oliviero, S. (2013). Dnmt3L antagonizes DNA methylation at bivalent promoters and favors DNA methylation at gene bodies in ESCs. Cell *155*, 121-134.

Ng, S.F., Lin, R.C., Laybutt, D.R., Barres, R., Owens, J.A., and Morris, M.J. (2010). Chronic high-fat diet in fathers programs beta-cell dysfunction in female rat offspring. Nature *467*, 963-966.

Nishiyama, A., Yamaguchi, L., Sharif, J., Johmura, Y., Kawamura, T., Nakanishi, K., Shimamura, S., Arita, K., Kodama, T., Ishikawa, F.*, et al.* (2013). Uhrf1-dependent H3K23 ubiquitylation couples maintenance DNA methylation and replication. Nature *502*, 249-253.

O'Donnell, C.J., and Elosua, R. (2008). [Cardiovascular risk factors. Insights from Framingham Heart Study]. Revista espanola de cardiologia *61*, 299-310.

Oie, S., Matsuzaki, K., Yokoyama, W., Tokunaga, S., Waku, T., Han, S.I., Iwasaki, N., Mikogai, A., Yasuzawa-Tanaka, K., Kishimoto, H.*, et al.* (2014). Hepatic rRNA transcription regulates high-fat-diet-induced obesity. Cell reports *7*, 807-820.

Ooi, S.K., Qiu, C., Bernstein, E., Li, K., Jia, D., Yang, Z., Erdjument-Bromage, H., Tempst, P., Lin, S.P., Allis, C.D.*, et al.* (2007). DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. Nature *448*, 714-717.

Ooi, S.L., and Henikoff, S. (2007). Germline histone dynamics and epigenetics. Current opinion in cell biology *19*, 257-265.

Paredes, S., Branco, A.T., Hartl, D.L., Maggert, K.A., and Lemos, B. (2011). Ribosomal DNA deletions modulate genome-wide gene expression: "rDNA-sensitive" genes and natural variation. PLoS genetics *7*, e1001376.

Pembrey, M.E., Bygren, L.O., Kaati, G., Edvinsson, S., Northstone, K., Sjostrom, M., Golding, J., and Team, A.S. (2006). Sex-specific, male-line transgenerational responses in humans. European journal of human genetics : EJHG *14*, 159-166.

Perry, G.H., Dominy, N.J., Claw, K.G., Lee, A.S., Fiegler, H., Redon, R., Werner, J., Villanea, F.A., Mountain, J.L., Misra, R.*, et al.* (2007). Diet and the evolution of human amylase gene copy number variation. Nature genetics *39*, 1256-1260.

Perry, R.J., Samuel, V.T., Petersen, K.F., and Shulman, G.I. (2014). The role of hepatic lipids in hepatic insulin resistance and type 2 diabetes. Nature *510*, 84-91.

Picelli, S., Faridani, O.R., Bjorklund, A.K., Winberg, G., Sagasser, S., and Sandberg, R. (2014). Full-length RNA-seq from single cells using Smart-seq2. Nature protocols *9*, 171-181.

Prokhortchouk, A., Hendrich, B., Jorgensen, H., Ruzov, A., Wilm, M., Georgiev, G., Bird, A., and Prokhortchouk, E. (2001). The p120 catenin partner Kaiso is a DNA methylation-dependent transcriptional repressor. Genes Dev *15*, 1613-1618.

Pryke, S.R., and Griffith, S.C. (2009). Genetic incompatibility drives sex allocation and maternal investment in a polymorphic finch. Science *323*, 1605-1607.

Qian, Q., Qian, H., Zhang, X., Zhu, W., Yan, Y., Ye, S., Peng, X., Li, W., Xu, Z., Sun, L.*, et al.* (2012). 5-Azacytidine induces cardiac differentiation of human umbilical cord-derived mesenchymal stem cells by activating extracellular regulated kinase. Stem cells and development *21*, 67-75.

Quenneville, S., Verde, G., Corsinotti, A., Kapopoulou, A., Jakobsson, J., Offner, S., Baglivo, I., Pedone, P.V., Grimaldi, G., Riccio, A.*, et al.* (2011). In embryonic stem cells, ZFP57/KAP1 recognize a methylated hexanucleotide to affect

chromatin and DNA methylation of imprinting control regions. Molecular cell *44*, 361-372.

Radford, E.J., Ito, M., Shi, H., Corish, J.A., Yamazawa, K., Isganaitis, E., Seisenberger, S., Hore, T.A., Reik, W., Erkek, S.*, et al.* (2014). In utero effects. In utero undernourishment perturbs the adult sperm methylome and intergenerational metabolism. Science *345*, 1255903.

Rakhshandehroo, M., Sanderson, L.M., Matilainen, M., Stienstra, R., Carlberg, C., de Groot, P.J., Muller, M., and Kersten, S. (2007). Comprehensive analysis of PPARalpha-dependent regulation of hepatic lipid metabolism by expression profiling. PPAR research *2007*, 26839.

Rakyan, V.K., Chong, S., Champ, M.E., Cuthbert, P.C., Morgan, H.D., Luu, K.V., and Whitelaw, E. (2003). 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.

Rando, O.J. (2012). Daddy issues: paternal effects on phenotype. Cell *151*, 702-708.

Rando, O.J., and Verstrepen, K.J. (2007). Timescales of genetic and epigenetic inheritance. Cell *128*, 655-668.

Rassoulzadegan, M., Grandjean, V., Gounon, P., Vincent, S., Gillot, I., and Cuzin, F. (2006). RNA-mediated non-mendelian inheritance of an epigenetic change in the mouse. Nature *441*, 469-474.

Reik, W., and Walter, J. (2001). Genomic imprinting: parental influence on the genome. Nature reviews Genetics *2*, 21-32.

Ronn, T., Volkov, P., Davegardh, C., Dayeh, T., Hall, E., Olsson, A.H., Nilsson, E., Tornberg, A., Dekker Nitert, M., Eriksson, K.F.*, et al.* (2013). A six months exercise intervention influences the genome-wide DNA methylation pattern in human adipose tissue. PLoS genetics *9*, e1003572.

Rothbart, S.B., Krajewski, K., Nady, N., Tempel, W., Xue, S., Badeaux, A.I., Barsyte-Lovejoy, D., Martinez, J.Y., Bedford, M.T., Fuchs, S.M.*, et al.* (2012). Association of UHRF1 with methylated H3K9 directs the maintenance of DNA methylation. Nature structural & molecular biology *19*, 1155-1160.

Rowe, H.M., and Trono, D. (2011). Dynamic control of endogenous retroviruses during development. Virology *411*, 273-287.

Santos, F., Hendrich, B., Reik, W., and Dean, W. (2002). Dynamic reprogramming of DNA methylation in the early mouse embryo. Developmental biology *241*, 172-182.

Schmitz, R.J., Schultz, M.D., Urich, M.A., Nery, J.R., Pelizzola, M., Libiger, O., Alix, A., McCosh, R.B., Chen, H., Schork, N.J*., et al.* (2013). Patterns of population epigenomic diversity. Nature *495*, 193-198.

Schuettengruber, B., Chourrout, D., Vervoort, M., Leblanc, B., and Cavalli, G. (2007). Genome regulation by polycomb and trithorax proteins. Cell *128*, 735-745.

Seisenberger, S., Andrews, S., Krueger, F., Arand, J., Walter, J., Santos, F., Popp, C., Thienpont, B., Dean, W., and Reik, W. (2012). The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells. Molecular cell *48*, 849-862.

Sene, A., Khan, A.A., Cox, D., Nakamura, R.E., Santeford, A., Kim, B.M., Sidhu, R., Onken, M.D., Harbour, J.W., Hagbi-Levi, S*., et al.* (2013). Impaired cholesterol efflux in senescent macrophages promotes age-related macular degeneration. Cell metabolism *17*, 549-561.

Seong, K.H., Li, D., Shimizu, H., Nakamura, R., and Ishii, S. (2011). Inheritance of stress-induced, ATF-2-dependent epigenetic change. Cell *145*, 1049-1061.

Sharif, J., Muto, M., Takebayashi, S., Suetake, I., Iwamatsu, A., Endo, T.A., Shinga, J., Mizutani-Koseki, Y., Toyoda, T., Okamura, K*., et al.* (2007). The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. Nature *450*, 908-912.

She, X., Cheng, Z., Zollner, S., Church, D.M., and Eichler, E.E. (2008). Mouse segmental duplication and copy number variation. Nature genetics *40*, 909-914.

Shiao, Y.H., Leighty, R.M., Wang, C., Ge, X., Crawford, E.B., Spurrier, J.M., McCann, S.D., Fields, J.R., Fornwald, L., Riffle, L*., et al.* (2012). Molecular and organismal changes in offspring of male mice treated with chemical stressors. Environmental and molecular mutagenesis *53*, 392-407.

Shukla, S., Kavak, E., Gregory, M., Imashimizu, M., Shutinoski, B., Kashlev, M., Oberdoerffer, P., Sandberg, R., and Oberdoerffer, S. (2011). CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. Nature *479*, 74-79.

Smith, Z.D., Chan, M.M., Mikkelsen, T.S., Gu, H., Gnirke, A., Regev, A., and Meissner, A. (2012). A unique regulatory phase of DNA methylation in the early mammalian embryo. Nature *484*, 339-344.

Smith, Z.D., and Meissner, A. (2013). DNA methylation: roles in mammalian development. Nature reviews Genetics *14*, 204-220.

Solon-Biet, S.M., McMahon, A.C., Ballard, J.W., Ruohonen, K., Wu, L.E., Cogger, V.C., Warren, A., Huang, X., Pichaud, N., Melvin, R.G.*, et al.* (2014). The ratio of macronutrients, not caloric intake, dictates cardiometabolic health, aging, and longevity in ad libitum-fed mice. Cell metabolism *19*, 418-430.

Stanghellini, I., Falco, G., Lee, S.L., Monti, M., and Ko, M.S. (2009). Trim43a, Trim43b, and Trim43c: Novel mouse genes expressed specifically in mouse preimplantation embryos. Gene expression patterns : GEP *9*, 595-602.

Surani, M.A., Barton, S.C., and Norris, M.L. (1984). Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis. Nature *308*, 548-550.

Symonds, M.E., Sebert, S.P., Hyatt, M.A., and Budge, H. (2009). Nutritional programming of the metabolic syndrome. Nature reviews Endocrinology *5*, 604-610.

Tang, Y.C., and Amon, A. (2013). Gene copy-number alterations: a cost-benefit analysis. Cell *152*, 394-405.

Trombetta, J.J., Gennert, D., Lu, D., Satija, R., Shalek, A.K., and Regev, A. (2014). Preparation of Single-Cell RNA-Seq Libraries for Next Generation Sequencing. Current protocols in molecular biology / edited by Frederick M Ausubel [et al] *107*, 4.22.21-24.22.17.

Tsuchiya, T., Dhahbi, J.M., Cui, X., Mote, P.L., Bartke, A., and Spindler, S.R. (2004). Additive regulation of hepatic gene expression by dwarfism and caloric restriction. Physiological genomics *17*, 307-315.

Tsumura, A., Hayakawa, T., Kumaki, Y., Takebayashi, S., Sakaue, M., Matsuoka, C., Shimotohno, K., Ishikawa, F., Li, E., Ueda, H.R.*, et al.* (2006). Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. Genes to cells : devoted to molecular & cellular mechanisms *11*, 805-814.

Veiko, N.N., Shubaeva, N.O., Malashenko, A.M., Beskova, T.B., Agapova, R.K., and Liapunova, N.A. (2007). [Ribosomal genes in inbred mouse strains: interstrain and intrastrain variations of copy number and extent of methylation]. Genetika *43*, 1226-1238.

Velando, A., Beamonte-Barrientos, R., and Torres, R. (2006). Pigment-based skin colour in the blue-footed booby: an honest signal of current condition used by females to adjust reproductive investment. Oecologia *149*, 535-542.

Waddington, C.H. (1953). Genetic assimilation of an acquired character. Evolution; international journal of organic evolution, 118-126.

Wagner, K.D., Wagner, N., Ghanbarian, H., Grandjean, V., Gounon, P., Cuzin, F., and Rassoulzadegan, M. (2008). RNA induction and inheritance of epigenetic cardiac hypertrophy in the mouse. Developmental cell *14*, 962-969.

Waterland, R.A., and Jirtle, R.L. (2003a). Transposable elements: targets for early nutritional effects on epigenetic gene regulation. Mol Cell Biol *23*, 5293-5300.

Waterland, R.A., and Jirtle, R.L. (2003b). Transposable Elements: Targets for Early Nutritional Effects on Epigenetic Gene Regulation. Molecular and Cellular Biology *23*, 5293-5300.

Watkins, A.J., and Sinclair, K.D. (2014). Paternal low protein diet affects adult offspring cardiovascular and metabolic function in mice. American journal of physiology Heart and circulatory physiology *306*, H1444-1452.

Weaver, I.C., Cervoni, N., Champagne, F.A., D'Alessio, A.C., Sharma, S., Seckl, J.R., Dymov, S., Szyf, M., and Meaney, M.J. (2004). Epigenetic programming by maternal behavior. Nature neuroscience *7*, 847-854.

Weber, M., Davies, J.J., Wittig, D., Oakeley, E.J., Haase, M., Lam, W.L., and Schubeler, D. (2005). Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. Nature genetics *37*, 853-862.

Weber, M., Hellmann, I., Stadler, M.B., Ramos, L., Paabo, S., Rebhan, M., and Schubeler, D. (2007). Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. Nature genetics *39*, 457-466.

Weiner, A., Chen, H.V., Liu, C.L., Rahat, A., Klien, A., Soares, L., Gudipati, M., Pfeffner, J., Regev, A., Buratowski, S.*, et al.* (2012). Systematic dissection of roles for chromatin regulators in a yeast stress response. PLoS biology *10*, e1001369.

Weismann, A., Poulton, E.B., Schönland, S., and Shipley, A.E. (1891). Essays upon heredity and kindred biological problems, 2d edn (Oxford,: Clarendon press).

Whitelaw, N.C., and Whitelaw, E. (2008). Transgenerational epigenetic inheritance in health and disease. Current opinion in genetics & development *18*, 273-279.

Wolfner, M.F. (2002). The gifts that keep on giving: physiological functions and evolutionary dynamics of male seminal proteins in Drosophila. Heredity *88*, 85-93.

Wu, H., and Zhang, Y. (2014). Reversing DNA methylation: mechanisms, genomics, and biological functions. Cell *156*, 45-68.

Xi, Y., and Li, W. (2009). BSMAP: whole genome bisulfite sequence MAPping program. BMC bioinformatics *10*, 232.

Yamaguchi, R., Nakagawa, Y., Liu, Y.J., Fujisawa, Y., Sai, S., Nagata, E., Sano, S., Satake, E., Matsushita, R., Nakanishi, T.*, et al.* (2010). Effects of maternal high-fat diet on serum lipid concentration and expression of peroxisomal proliferator-activated receptors in the early life of rat offspring. Hormone and metabolic research = Hormon- und Stoffwechselforschung = Hormones et metabolisme *42*, 821-825.

Yang, X., Deignan, J.L., Qi, H., Zhu, J., Qian, S., Zhong, J., Torosyan, G., Majid, S., Falkard, B., Kleinhanz, R.R.*, et al.* (2009). Validation of candidate causal genes for obesity that affect shared metabolic pathways and networks. Nature genetics *41*, 415-423.

Youngson, N.A., and Whitelaw, E. (2008). Transgenerational epigenetic effects. Annual review of genomics and human genetics *9*, 233-257.

Zemach, A., and Zilberman, D. (2010). Evolution of eukaryotic DNA methylation and the pursuit of safer sex. Current biology : CB *20*, R780-785.

Zeybel, M., Hardy, T., Wong, Y.K., Mathers, J.C., Fox, C.R., Gackowska, A., Oakley, F., Burt, A.D., Wilson, C.L., Anstee, Q.M.*, et al.* (2012). Multigenerational epigenetic adaptation of the hepatic wound-healing response. Nature medicine *18*, 1369-1377.

Zhao, Q., Rank, G., Tan, Y.T., Li, H., Moritz, R.L., Simpson, R.J., Cerruti, L., Curtis, D.J., Patel, D.J., Allis, C.D.*, et al.* (2009). PRMT5-mediated methylation of histone H4R3 recruits DNMT3A, coupling histone and DNA methylation in gene silencing. Nature structural & molecular biology *16*, 304-311.

Zhong, S.H., Liu, J.Z., Jin, H., Lin, L., Li, Q., Chen, Y., Yuan, Y.X., Wang, Z.Y., Huang, H., Qi, Y.J.*, et al.* (2013). Warm temperatures induce transgenerational epigenetic release of RNA silencing by inhibiting siRNA biogenesis in

Arabidopsis. Proceedings of the National Academy of Sciences of the United States of America *110*, 9171-9176.

Ziller, M.J., Gu, H., Muller, F., Donaghey, J., Tsai, L.T., Kohlbacher, O., De Jager, P.L., Rosen, E.D., Bennett, D.A., Bernstein, B.E.*, et al.* (2013). Charting a dynamic DNA methylation landscape of the human genome. Nature *500*, 477-481.

Ziller, M.J., Muller, F., Liao, J., Zhang, Y., Gu, H., Bock, C., Boyle, P., Epstein, C.B., Bernstein, B.E., Lengauer, T.*, et al.* (2011). Genomic distribution and inter-sample variation of non-CpG methylation across human cell types. PLoS genetics *7*, e1002389.