The sperm epigenome in mice with the human equivalent of the *MTHFR* 677C>T variant and the effects of folic acid deficiency and supplementation

Edgar Martínez Duncker Rebolledo

Department of Human Genetics, Faculty of Medicine and Health Sciences,

McGill University, Montreal, Quebec, Canada

OCTOBER 2022

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Master of Science

Table of Contents

Abstract

The one-carbon metabolism pathway enzyme, 5,10-methylenetetrahydrofolate reductase (MTHFR), plays a key role in providing methyl groups for DNA methylation, including during spermatogenesis. A common genetic variant in humans, *MTHFR* 677C>T, results in a thermolabile enzyme with ~50% reduced enzymatic activity and has been linked to various disorders including infertility. A new mouse model has been developed by introducing the human equivalent polymorphism using CRISPR/Cas9. Biochemical parameters in the *Mthfr* 677TT mice recapitulate values found in *MTHFR* 677TT men. The aim of this study was to characterize reproductive parameters and the sperm DNA methylome of the newly developed mice exposed to diets with different levels of folic acid. Adult male mice were fed a folic acid control diet (CD; 2mg folic acid/kg diet), a 7-fold folic acid deficient diet (FD; 0.3mg folic acid/kg diet) and a moderate dose folic acid supplemented diet (FASD; 10mg folic acid/kg diet), for four months. Methylation of imprinted genes was examined using bisulfite pyrosequencing. Genome-wide DNA methylation was studied using the Illumina Infinium™ Mouse Methylation BeadChip for CD and FD samples, and whole-genome bisulfite sequencing (WGBS) for CD and FASD samples. Body and reproductive organ weights, as well as testicular sperm counts, and histology were unaffected by either genotype or diet. Imprinted genes showed correct levels of methylation with bisulfite pyrosequencing, Infinium Array and WGBS. The largest effect out of all comparisons was caused by the TT genotype, showing both hyper- and hypomethylation, with a predominance of the latter. FD did not exacerbate the effects found in the TT genotype. FASD was able to partly rescue the effects caused by the TT genotype. Repeat sequences were not largely affected. Gene ontology enrichment

analyses revealed common terms across different groups, such as cell migration processes and cell adhesion. Other enriched pathways included neurological processes. The results presented in this study are the first to characterize the sperm DNA methylome of the *Mthfr* 677C>T mice exposed to diets containing different amounts of folic acid. Furthermore, this study demonstrates that the newly developed mouse constitutes a promising model for the human *MTHFR* 677C>T polymorphism, enabling the opportunity for further research projects.

Résumé

L'enzyme 5,10-méthylènetétrahydrofolate réductase (MTHFR), impliquée dans le métabolisme du carbone, joue un rôle clé en fournissant les groupements méthyles nécessaires à la méthylation de l'ADN, entre autres durant la spermatogenèse. Un variant génétique relativement commun chez l'humain, *MTHFR* 677C>T, produit une enzyme thermolabile exhibant 50% de l'activité normale, a été associé à plusieurs troubles médicaux, incluant l'infertilité. Un nouveau modèle murin a été récemment développé, utilisant la technologie CRISPR/Cas9 pour introduire ce polymorphisme dans le génome de la souris. Les paramètres biochimiques mesurés chez la souris mâle *Mthfr* 677TT sont similaires à ceux qu'on rapporte chez l'homme *MTHFR* 677TT. Le but de cette étude était de caractériser les paramètres reproducteurs et le méthylome de l'ADN spermique chez ce nouveau modèle murin, en l'exposant à des diètes contenant différents niveaux d'acide folique. Des souris mâles adultes ont été soumis à une diète contrôle (CD; 2 mg acide folique/ kg diète), une diète déficiente en acide folique (FD; 0.3 mg acide folique/kg diète) et une diète modérément supplémentée (FASD; 10 mg acide folique/ diète) pendant 4 mois. Les niveaux de méthylation de gènes à empreinte ont été examinés par pyroséquencage. La méthylation de l'ADN spermique à l'échelle du génome entier fut examinée par deux méthodes différentes : la puce à ADN « Infinium™ Mouse Methylation BeadChip » d'Illumina fut utilisée pour les échantillons CD et FD et la technique de séquençage à haute résolution « Whole-Genome Bisulfite Sequencing (WGBS) » fut appliquée sur les échantillons CD et FASD. L'évaluation du poids corporel, des organes reproducteurs, de la numération des spermatozoïdes et de l'histologie testiculaire n'a identifié aucun effet du génotype ou de la diète. Les gènes à empreinte ont exhibé des

patrons adéquats de méthylation de l'ADN, tel qu'évalués par pyroséquencage, puce à ADN Infinium et séquençage à haute résolution WGBS. L'effet le plus marqué au niveau de la méthylation de l'ADN fut observé dans le groupe variant TT, certaines régions démontrant une diminution et la majorité des régions exhibant une augmentation du niveau de méthylation. Une déficience en acide folique (FD) n'a pas exacerbé cet effet et une supplémentation (FASD) a partiellement corrigé les changements liés au génotype. Les éléments répétés du génome n'ont pas été significativement affectés par la diète ou le génotype. L'analyse de l'ontologie des gènes a identifié des fonctions communes parmi différents groupes tels que « processus de migration cellulaire » et « adhésion cellulaire » et plusieurs autres impliqués dans les processus neurologiques. En conclusion, les résultats présentés dans cette étude sont les premiers à caractériser le méthylome de l'ADN spermique du nouveau modèle murin *Mthfr* 677C>T exposé à différents niveaux d'acide folique nutritionnel. Cette étude démontre aussi que cette souris est un modèle prometteur pour de futurs projets de recherche permettant d'approfondir les connaissances sur le variant humain *MTHFR* 677C>T.

Acknowledgements

This project would not have been possible without the support and collaboration of many people. Before anything else, I would like to express my most sincere gratitude to my supervisor Dr. Jacquetta Trasler for giving me the outstanding opportunity to work and learn so much in her lab. Thank you for seeing the potential in me and for reminding me that a chance not taken is an opportunity missed.

To the current and former lab members of the Trasler lab, thank you all for always creating a lighthearted yet productive work environment. I truly enjoyed working together these past years. A special thanks to Donovan Chan for always having a great disposition and the patience to teach me how to analyze data and together learn the best way to do so on Zoom during uncertain times. Josée Martel, thank you as well for teaching me lab techniques and helping me learn how to be efficient and meticulous, as well as for all the valuable “tips and tricks” that certainly helped me along the way. I would also like to thank Sophia Rahimi for sharing her expertise on how to best perform bisulfite pyrosequencing analyses. To my fellow students, Gloria Ihirwe thank you for always being a text away and helping me troubleshoot problems I encountered during our time together in the lab, and Julia Barnwell for being my cheerful partner during poster presentations at conferences, amazing e-race (which you and I know won) and Pedal for Kids fundraiser!

I must also acknowledge the helpful contributions of my supervisory committee members, Dr. Rima Rozen, and Dr. Cristian O’Flaherty, your helpful input and availability is greatly appreciated.

Thank you to the Human Genetics Department for creating a great learning environment and for being so responsive whenever I needed help. I am proud to be a part of this department. As well, I would like to take this opportunity to thank the RI MUHC/Desjardins, the *Réseau Québecois en Reproduction* and McGill University's Faculty of Medicine, all of which provided financial support throughout my time as a master's student.

To all the good friends I made here from the Human Genetics department, other departments, and even other universities, thank you all for the good times, I wish you all the very best. A special warm thanks to Logan and Coco, the ones who saw me work more than anyone else. Both reminded me that it's okay to take a break, one made sure that I stayed well fed and the other one helped me stay active thanks to his unparalleled bursts of energy (you know which one is which).

Last but certainly not least, I wouldn't be here without the unwavering support of my family, thank you all for help and encouragement, now and always.

Contribution of Authors

The research presented in this thesis was performed by the candidate with assistance from the lab members and collaborators. The project was designed and performed under the supervision of Dr. Jacquetta Trasler.

Mice and tissue collections were performed by Dr. Donovan Chan in collaboration with Dr. Rima Rozen's lab. Sperm DNA isolation was performed by the candidate with the assistance of Dr. Donovan Chan

Genotyping of mouse sperm DNA was performed by the candidate as well as bisulfite conversion of mouse sperm DNA samples. Bisulfite pyrosequencing was taught to the candidate by Ms. Sophia Rahimi who offered assistance with the initial samples. Subsequent analyses were performed by the candidate. Ms. Josée Martel and MSc candidate Gloria Ihirwe helped troubleshoot problems encountered.

Testicular histology slides were prepared at the histopathology platform in the RI-MUHC. Illumina Mouse Methylation Array was performed by the Innovation Centre of Genome Québec and WGBS was conducted at the McGill Genome Centre.

Mouse sperm homogenization, testicular sperm counts, testicular histology analysis and all statistical analyses were performed by the candidate.

Bioinformatic data analyses were all performed by the candidate with the assistance of Dr. Donovan Chan who kindly offered help whenever any obstacles were encountered.

The abstract was kindly translated into French by Ms. Josée Martel.

Abbreviations

5mC	5-methylcytosine
5-methylTHF	5-methyltetrahydrofolate
CD	control diet
CH3	methyl group
CpG	cytosine-phosphate-guanine
CRIPSR	clustered regularly interspaced short palindromic repeats
DMC	differentially methylated cytosine
DMR	differentially methylated region
DNA	deoxyribonucleic acid
DNMT	DNA methyltransferase
DTT	dithiothreitol
E	embryonic day
EtOH	ethanol
FAD	flavin adenine dinucleotide
FASD	folic acid supplemented diet
FD	folate deficient diet
GO	gene ontology
H3K4me3	histone H3 lysine 4 trimethylation
ICR	imprinting control region
MTHFR	5,10-methylenetetrahydrofolate reductase
NaCl	sodium chloride
PBS	phosphate-buffered saline
PND	postnatal day
RRBS	reduced representation bisulfite sequencing
RRVGO	r package reduce and visualize gene ontology
SAH	S-adenosylhomocysteine
SAM	S-adenosylmethionine
SEM	standard error of the mean
THF	tetrahydrofolate
WGBS	whole genome bisulfite sequencing

List of figures

List of tables

CHAPTER 1: INTRODUCTION

For many years it has been known that an alteration of 5,10-methylenetetrahydrofolate reductase (MTHFR) increases the risk for various disorders. MTHFR is a key enzyme in the one-carbon metabolism (Fig 1), which is a series of interlinking metabolic pathways that include the methionine and folate cycles which are critical for normal cellular function. MTHFR catalyzes the synthesis of 5-methyltetrahydrofolate (5-methyl-THF), which in turn, acts as a methyl donor for the synthesis of methionine from homocysteine for the subsequent production of S-adenosylmethionine (SAM). DNA methylation, a crucial epigenetic modification of DNA that regulates gene expression and cellular differentiation, is catalyzed by DNA methyltransferase enzymes (DNMTs) which require methyl groups provided by SAM.

In the human population, there is a common mutation in the gene that encodes *MTHFR* identified as a C to T substitution at nucleotide 677 (*MTHFR* 677C>T), which converts an alanine to valine residue (Frosst et al., 1995) and results in an enzyme with ~50% reduced activity. A study in 2002 (Friso et al., 2002) showed that the *MTHFR C677T* polymorphism influences DNA methylation status through an interaction with folate status. Individuals with a TT genotype with low levels of folate had a diminished level of genomic DNA methylation in peripheral blood, compared to those with the CC wild-type genotype.

More recent studies have shown an association of this polymorphism with various diseases such as vascular and neurological disorders, cancers, diabetes, psoriasis, etc. Epidemiological studies show that these associations vary depending on the geography and ethnicity of the affected individuals (Liew & Gupta, 2015). Furthermore, the mutation

has also been correlated with an increased risk of male infertility in certain populations (Gong et al., 2015). Infertility is a recurring health problem that affects approximately 8-12% of couples globally (Agarwal et al., 2021), and almost 50% cases are due to male factors. Among other causes, methylation abnormalities of sperm DNA have been linked to male infertility (Kobayashi et al., 2007) and it has been shown that the one carbon metabolism can be affected by polymorphisms of related genes, such as *MTHFR* (Lee et al., 2006). The variant *MTHFR* 677 C>T can alter methyl donor availability, disrupting DNA methylation processes during in utero development when critical DNA methylation reprogramming events in the germline occur (Fig. 2), or even during the maintenance of DNA methylation throughout the male's lifetime. The persistence of altered DNA methylation patterns in the germline may therefore increase the risk for infertility.

In contrast, other studies have suggested that this polymorphism might also confer certain protection against other conditions. For instance, in Van Guelpen et al., 2006, they reported a possible decreased risk of colorectal cancer in individuals with the *MTHFR* 677TT polymorphism due to their lower folate status. Similarly, in Saffroy et al., 2008, the authors communicated a possible protective role of *MTHFR* 677TT against alcohol dependence. These types of shielding effects of this polymorphism warrant further investigation.

To better understand the underlying mechanisms of MTHFR deficiency and the epigenetic impact of deficient and high folic acid supplementation on sperm and the testis, mouse studies have been performed. In particular, a knockout mouse model with mild MTHFR deficiency was developed as a model for the human *MTHFR* 677TT genotype. These mice present hyperhomocysteinemia and reduced folate status with global DNA

hypomethylation (Chen et al., 2001), however, even though these mice reflect certain biochemical parameters observed in men with the *MTHFR 677C>T* mutation, they do not produce a thermolabile enzyme. Thanks to the advancement of gene editing technologies, a new mouse model has been developed using CRISPR/cas9 at The Jackson Laboratory by introducing the human equivalent single nucleotide polymorphism (SNP) 677 C>T into the mouse genome and then backcrossed to C57Bl/6J mice (Reagan et al., accepted in *Journal of Cerebral Blood Flow & Metabolism*, 2022).

To better understand the bigger picture, the introduction will detail fascinating mechanisms that have been studied thus far and will help understand the vast trajectory that male germ cells need to go through in order to become normal functioning mature spermatozoa. The introduction will begin with the prenatal development of these highly specialized cells and their path to spermatogenesis. Next, it will revise epigenetic modifications focusing mainly on DNA methylation and their critical role in normal male germ cell development and function. Finally, it will explain the consequences of MTHFR deficiency, along with folic acid deficiency and supplementation.

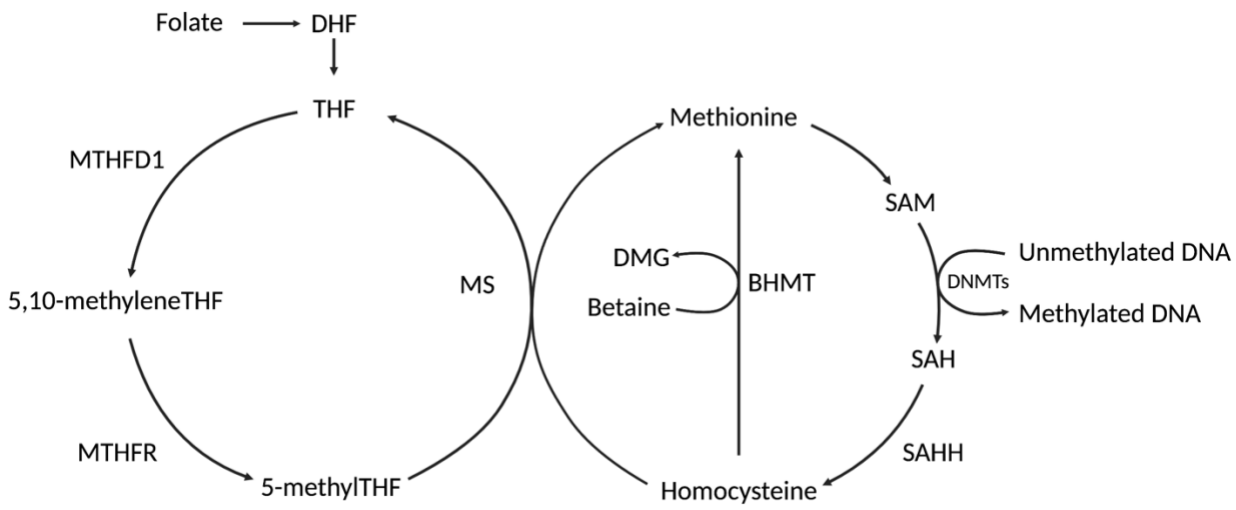


Figure 1. One-carbon metabolism cycle.

The one-carbon metabolism cycle is a series of interlinking pathways that comprise the folate and methionine cycles. This pathway is crucial for cellular function, synthesis of DNA, and providing methyl groups for DNA methylation processes. MTHFR is a key enzyme that reduces 5,10-methyleneTHF to 5-methyTHF, a primary carbon donor for methionine production from homocysteine. Methionine in turn is used to produce SAM, a global methyl donor. DHF, dihydrofolate; THF, tetrahydrofolate; MTHFD1, methylenetetrahydrofolate dehydrogenase, cyclohydrolase and formyltetrahydrofolate synthetase 1; MTHFR, 5,10-methylenetetrahydrofolate reductase; MS, methionine synthase; BHMT, betaine-homocysteine methyltransferase; DMG, dimethylglycine; DNMTs, DNA methyltransferases; SAH, S-adenosylhomocysteine; SAHH, S-adenosylhomocysteine hydrolase; SAM, S-adenosylmethionine (Modified from Miousse et al., 2017)

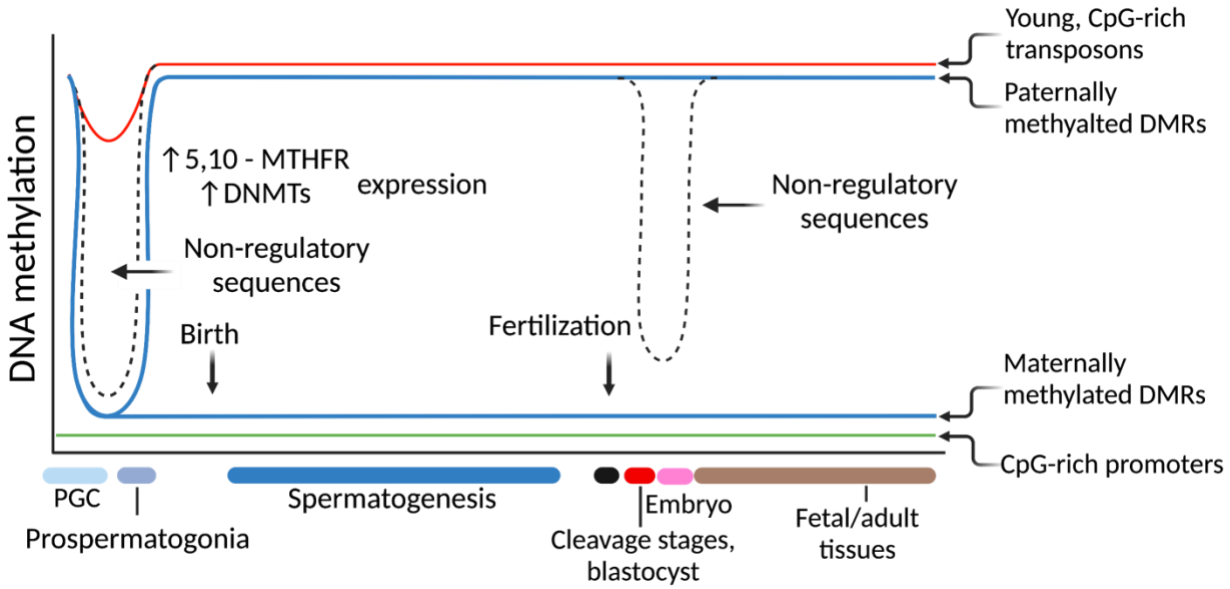


Figure 2. Epigenetic reprogramming in male germ cells.

During development, male germ cells are subjected to global erasure followed by re-establishment of DNA methylation patterns. DNMTs and MTHFR are highly expressed during re-methylation. PGCs, primordial germ cells; MTHFR 5,10-methylenetetrahydrofolate reductase; DNMTs, DNA methyltransferases; CpG, cytosine-phosphate-guanine; m^5C, 5-methylcytosine. (Modified from Edwards et al., 2017)

1.1 Male germ cell development

1.1.1 Prenatal development

The germ cell lineage is imperative for the creation of new individuals, ensuring the preservation of genetic and epigenetic information across generations. In mice, germ cell fate commences in the epiblasts at around embryonic day (E) 6.0 by BMP4 (bone morphogenic signaling protein 4) signaling from the extraembryonic ectoderm (Hayashi et al., 2011). Both male and female gametes originate from a common predecessor, the primordial germ cells (PGCs), these cells can be found at around E7.25 as a cluster of alkaline phosphatase-positive cells in the extraembryonic mesoderm (Saitou et al., 2002). As PGCs migrate through the hindgut, they actively proliferate and colonize the genital ridges around E8-E10.5 (von Meyenn et al., 2016). By E11.5, most of the germ cells can be found in the genital ridges; however, some will remain scattered along the migratory trajectory (Molyneaux et al., 2001). Next, sexual differentiation commences, with seminiferous cords beginning to form around the germ cells in males which are now called gonocytes or prospermatogonia. These cells then enter mitotic arrest and remain inactive until not long after birth.

1.1.2 Postnatal development and spermatogenesis.

The commitment of fetal male germ cells to mitotic arrest is crucial for their subsequent development (Zhao et al., 2021). After birth, gonocytes migrate to the basement membrane of the seminiferous tubules and resume mitosis, giving rise to spermatogonial stem cells (SSCs) (Culty, 2009). These cells have the capacity of self-renewal, allowing them to maintain the population of spermatogonia in the testes. The male germ cells continue to proliferate and mature to then migrate towards the lumen of

the seminiferous tubules. Surrounding the germ cells are the Sertoli cells, these are vital as they tightly control the development of germ cells. These cells also provide them with nutrients and phagocyte apoptotic germ cells (Mruk and Cheng, 2004). Another crucial function of Sertoli cells is the formation of an immune-privileged environment for the germ cells through the formation of tight junctions, between the basal and adluminal compartments of the seminiferous tubule. Located between the seminiferous tubules are the interstitial cells, somatic Leydig cells can be found here and are responsible for producing testosterone and other androgens in males.

Spermatogenesis is a complex process by which mature haploid spermatozoa are produced from diploid spermatogonia and SSCs. In mice this process takes about 35 days, whereas in humans it takes 64 days (Adler, 1996). Notwithstanding, both species share similarities in germ cell development. Spermatogenesis is initiated after birth, where population of undifferentiated, self-renewing prospermatogonia re-enter the cell cycle by resuming mitosis and initiating spermatogenesis. In mice, SSCs initiate meiotic entry at postnatal day 7 (PND7). The complete process of spermatogenesis can be divided into three distinct phases (Bellve et al., 1977) and has been thoroughly studied and outlined (Clermont, Y., 1972). The first one is the proliferative phase, here spermatogonia continuously divide to increase the total population of cells or differentiate into daughter cells that become mature gametes (mitotic differentiation). The second one is the meiotic phase, in which tetraploid (containing four homologous sets of chromosomes) spermatocytes ensure genetic diversity by recombination and production of haploid (having a single set of unpaired chromosomes) germ cells, now termed spermatids. Lastly, spermiogenesis encompasses the third phase, in which substantial morphological

changes and chromatin packing occurs in the round cells to produce spermatozoa. These will eventually migrate to finish maturation and be stored in the epididymides (Hess and Renato de Franca, 2008).

Importantly, during spermiogenesis, round spermatids elongate and shed the majority of their cytoplasm. The nucleus then undergoes remarkable condensation, which is achieved by replacing histones with transition proteins and then protamines. (Marushige and Marushige, 1975; Allen et al., 1993; Haaf and Ward, 1995; Lesch et al., 2014). Interestingly, approximately 1% of sperm histones are retained in mice whereas 15% are retained in men (Hammoud et al., 2009; Erkek et al., 2013). Retained histones are found at regulatory regions of promoters involved in spermatogenesis, sperm function, embryo development, metabolism, and routine cellular processes. (Hammoud et al., 2009; Brykczynska et al., 2010; Lesch et al., 2016). Abnormalities during this process have been known to result in detrimental outcomes. For example, aberrant distribution of genome-wide histone retention was reported in men with infertility (Hammoud et al., 2011). As well, irregular DNA methylation has been found at CpGs located at incorrectly retained histones when compared to fertile men. These abnormalities occurred in areas related to genes responsible for embryonic development and correlated with altered gene expression in the blastocysts. This underscores the importance of normal histone placement and modifications for adequate cellular function (Denomme et al., 2017)

1.2 Epigenetic modifications.

Epigenetics refers to the modifications to DNA and chromatin that lead to changes in gene expression without changing the actual DNA sequence. This is how, despite

having the same DNA content, different cells, and tissues in an individual, express genes differently. Additionally, epigenetics is important for X-chromosome inactivation in female mammals (Lyon, M., 1961; Riggs, A., 1975), which is crucial so that females do not have twice the number of X-chromosome gene products. Within cells, there are three known epigenetic mechanisms that can interact with each other to carefully regulate genes: DNA methylation, histone modifications and non-coding RNAs. Disruptions in normal epigenetic patterns have the potential of causing diseases such as cancer and developmental disorders (Santos-Rebouças and Pimentel, 2007).

As well, these types of modifications play important roles in normal germ cell function and in postfertilization embryonic development. In the context of male germ cells, to form sperm and ensure the totipotency of the zygote, these epigenetic modifications must be carefully regulated. For this reason, alterations in these crucial modifications associated with reproductive processes can also have negative outcomes, such as male infertility or early embryo development failure. Various studies have underscored the importance of DNA methylation in silencing a group of tissue-specific and imprinted genes, as well as repetitive elements during spermatogenesis (Liu et al., 2019). Genomic imprinting is an epigenetic phenomenon in which DNA methylation marks at differentially methylated regions (DMRs) permit the monoallelic expression of genes in a parent-of-origin manner (Reik, Dean at al., 2001). This phenomenon is pivotal for normal embryonic growth and development. Indeed, various disorders are linked to an abnormal expression of imprinted genes, as is the case for syndromes such as Beckwith-Wiedemann, Prader-Willi, Angelman Syndromes and even cancers (Paulsen and Ferguson-Smith, 2001).

1.2.1 Histone modifications

Histone modification is the process of reshaping histone proteins by enzymatic interactions. There are many types of histone modifications, such as acetylation, methylation, ubiquitination, citrullination and phosphorylation of specific amino acids within the histone protein, usually towards the C-terminal ('tail') end of the protein (Kouzarides, 2007; Kebede et al., 2015; Lawrence et al., 2016). These modifications play a critical role in gene expression by regulating the degree of compaction, of chromatin. Methylation, which frequently occurs on histones H3 and H4 on specific lysine (K) and arginine (A) residues, is one important method for the modification of histone proteins (Bannister and Kouzarides, 2011). Histone lysine methylation can lead to activation or inhibition, usually depending on the position in which it is located. For instance, H3K9, H3K27, and H4K20 are well known to be "inactivation" markers because of the interaction between these methylation marks and heterochromatin formation. However, the methylation of H3K4 and H3K36 are considered to be "activation" marks. Acetylation, which often occurs in the N-terminal conserved lysine residues, is also an important way to modify the histone proteins. This modification is often associated with the activation or opening of the chromatin. In contrast, de-acetylation of the lysine residues leads to chromatin compression and inactivation of gene transcription. Different histone modifications can affect each other and can have interactions with DNA methylation (Bannister and Kouzarides, 2011).

1.2.2 Non-coding RNAs

Non-coding RNAs are a cluster of RNAs that do not encode functional proteins and were originally considered to merely regulate gene expression at the post-

transcriptional level. These can be divided into “housekeeping” non-coding RNAs and regulatory non-coding RNAs. RNA that has a regulatory role is mainly divided into two categories based on size: short chain non-coding RNAs (including siRNAs, miRNAs, and piRNAs) and long non-coding RNA (lncRNAs). In recent years, several studies have shown that non-encoding RNAs play a significant role in epigenetic modification and can regulate expression at the level of genes and chromosomes to control cell differentiation (Wei et al., 2017).

1.2.3 DNA methylation

DNA methylation is the best studied epigenetic process, it occurs when a methyl group is added to the fifth carbon of cytosine in DNA to convert cytosine to 5-methylcytosine (5mC), which commonly happens in the context of CpG dinucleotides in the genome (Greenberg and Bourc'his, 2019). The transcription of most protein coding genes is initiated at promoters rich in CG sequences, that is a cytosine that is positioned next to a guanine linked by a phosphate called a CpG site. Individual CpGs are commonly found to be hypermethylated. On the other hand, dense CpG stretches of 500-2000 base-pair-long regions, associated with gene promoters are known as CpG islands and are typically hypomethylated (Meissner et al., 2008; Deaton and Bird, 2011). In the mammalian genome, approximately 20 million and 32.3 million (Gershman et al., 2022) CpG dinucleotides are methylated (Trasler, 2006) in mice and humans, respectively. Methylation of cytosines varies across the genome. Some areas are more methylated than others, and heavily methylated areas tend to be less transcriptionally active. The vast majority of methylated cytosines can be found within repeat sequences and transposable elements such as LINE-1. These elements have the capacity of disrupting

gene expression, but DNA methylation prevents them from doing so. Indeed, demethylation of these sequences can result in transposons reactivation (Walsh and Bestor, 1999).

DNA methylation patterns are known to be established and modified in response to environmental factors by a complex interplay of enzymes called DNA methyltransferases (DNMTs) (Okano et al., 1998). At least three independent DNMTs have been characterized and classified, DNMT1, DNMT3A and DNMT3B (Hermann et al., 2004). These catalyze the reaction in which methyl groups from S-adenosylmethionine (SAM) are transferred to cytosines. DNMT1, the first of these enzymes to be discovered preferentially modifies hemimethylated DNA by transferring patterns of methylation to a newly synthesized strand after DNA replication (maintenance methylation) (Goll and Bestor, 2005). For this reason, DNMT1 is essential for adequate embryonic development, imprinting and X-inactivation. DNMT3 is a family of enzymes that can methylate hemimethylated and unmethylated CpG at the same degree. This family consists of: DNMT3A, DNMT3B, DNMT3C and DNMT3L. DNMT3A and DNMT3B both have DNA methyltransferase activity and can also interact with DNMT1 for methylation events. DNMT3L does not possess any catalytic activity (Aapola, Lyle et al. 2001), nonetheless, it contains DNA methyltransferase motifs. It is expressed during gametogenesis when genomic imprinting takes place and also plays a role in stem cell biology (Liao et al., 2012).

1.3 Epigenetic reprogramming in male germ cells

Male germ cells go through a pattern of global genomic DNA methylation erasure followed by re-establishment in the prenatal period. This process ensures the cellular

totipotency of subsequent generations and prevents the accumulation of epimutations (Ohno et al., 2013). In mice, the erasure phase takes place between E8.0-13.5 in primordial germ cells, when the majority of the ~20 million CpG sites in the genome are demethylated. This process coincides with the time when PGCs migrate and colonize the genital ridge (Sasaki and Matsui, 2008). The global loss of DNA methylation in PGCs entails both the active and passive demethylation pathways (Popp et al., 2010; Seisenberger et al., 2012; Ohno et al., 2013; Dawlaty et al., 2013; Kagiwada et al., 2013; Shirane, 2022).

Passive DNA demethylation is largely responsible for global DNA demethylation, for this, replication-coupled dilution of 5mC must occur (Kagiwada et al., 2013). While DNMT1 shows ubiquitous expression throughout the window of DNA demethylation, the canonical somatic pathway for methylation maintenance in germ cells via UHRF1 (also known as NP95), as well as *de novo Dnmt3a* and *Dnmt3b* are all expressed at a low level in PGCs (Shirane, 2022). The low expression of UHRF1 in PGCs leads to poor recruitment of DNMT1 to the replication loci between E10.5 and E13.5 when these cells divide twice per week (Kagiwada et al., 2013; Shirane, 2022).

Active demethylation requires the modification of 5mC by enzymes in various pathways. For instance, de-amination of 5mC into thymine via certain enzymes like Activation-induced Cytidine Deaminase (AID) and others, or oxidization of 5mC into 5-hydroxymethylcytosine (5hmC) via ten-eleven translocation proteins (TET1, 2, 3). Afterwards, 5hmC is further de-aminated to thymine or oxidized, and thymine DNA glycosylase (TDG) completes the demethylation process via a DNA repair mechanism (Kriaucionis et al., 2009; Ito et al., 2011). *Tet1* and *Tdg* are expressed in PGCs between

E9.5-13.5, during the time of DNA demethylation (Kagiwada et al., 2013). Interestingly, however, *Tet 1* KO PGCs show a slight increase in global DNA methylation levels, suggesting that a fraction of the genome depends on TET1-mediated DNA demethylation in PGCs (Shirane 2022; Yamaguchi et al., 2012).

After global DNA methylation erasure in PGCs, *de* novo DNA methylation takes place for the most part between ~E13.5-E19.0 in mitotically arrested prospermatogonia (Kubo et al. 2015), a time when the DNMTs are all expressed at high levels in male germ cells (La Salle and Trasler, 2006). Mouse studies have also revealed a crucial role for *MTHFR* in male germ cells as it is most highly expressed at the time of DNA methylation acquisition in the embryonic gonad (Garner et al., 2013). The targets of *de novo* DNA methylation in the male germline include retrotransposons, imprinted DMRs and non-promoter regions (Shirane, 2022; Kobayashi et al., 2013; Kubo et al., 2015). DMR methylation of imprinted genes is acquired in an allele- and sex-specific manner. For instance, for the gene *H19,* the paternally inherited allele acquires complete methylation before birth in prenatal germ cells, whereas the maternally inherited allele completes its methylation in postnatal spermatogonia (Davis et al., 2000). The first wave of *de novo* methylation in prenatal male germ cells takes place between E14-E17. During this time, four known DMRs of imprinted genes acquire methylation in male germ cells: *H19, IG, Rasgr1* and *Zdbf2*. These regions are commonly referred to as gametic or germline differentially methylated regions (gDMRs). Next, around E16.5, the majority of retrotransposons and genic regions show intermediate or high average methylation (Kobayashi et al., 2013; Molaro et al, 2014). A portion of retrotransposons, however, avoid this initial remethylation. The second wave of *de novo* methylation is suggested to happen

around E.17 via piRNA pathway and completed by postnatal day 2 (PND 2) (Molaro et al., 2014).

After birth, some remodeling continues, studies have shown that global CpG methylation levels increase from ~55% at E16.5 to ~75% at PND 2.5 (Molaro et al., 2014; Morselli et al., 2015). Another study showed that non-repeat intergenic regions also gain some DNA methylation between E16.5 and PND6 (Niles et al., 2013). Because of their potential to disrupt genetic stability, retrotransposons must continue to remain silenced, this is one of the essential roles of *de novo* DNA methylation in postnatal male germ cell development.

1.4 One-carbon metabolism

The one-carbon metabolism is a series of interlinking metabolic pathways that include the methionine and folate cycles (Clare et al., 2019) which are crucial for the synthesis of nucleotides, amino acids, and SAM. As described before, DNMTs catalyze DNA methylation events, and these enzymes use methyl groups which are provided by S-adenosylmethionine (SAM), a universal methyl donor.

Folate is a B vitamin known to be important for numerous biochemical reactions. It must be acquired through dietary intake as it cannot be internally synthesized by humans. A synthetic form of folate, called folic acid, was synthesized in 1945. When absorbed in the body, folate is converted to tetrahydrofolate (THF) by dihydrofolate reductase (Fig 1). Next, the reduction of 5,10-methylenetetrahydrofolate (5,10-methyleneTHF) to 5-methylTHF irreversibly occurs by 5,10-methyleneTHF reductase (MTHFR) activity, and the following transfer of the methyl group of 5-methylTHF to homocysteine result in the production of methionine. Methionine then provides the methyl

groups required for SAM production which are in turn used for DNA methylation reactions via methyltransferases (Goll et al., 2005; Ducker and Rabinowitz, 2017). Both dietary or genetic disruptions, such as the *MTHFR* 677C>T polymorphism, can negatively impact the folate metabolism and the generation of SAM.

1.5 MTHFR deficiency, folic acid deficiency and supplementation

Mutations to genes of the one carbon metabolism pathway, such as *MTHFR*, have the potential of causing lower enzymatic activity which can lead to a decrease in available SAM pools. A common genetic variant in humans in which cytosine is replaced by thymine at the 677th position (*MTHFR* 677C>T) results in an alanine to valine substitution at amino acid 222 (A222V). This substitution alters the active site of the enzyme rendering it thermolabile and less biologically active (Kang et al., 1991). Individuals with this polymorphism have shown lower folate status, higher concentrations of circulating homocysteine and reduced lymphocyte DNA methylation (Frosst et al., 1995; Friso et al., 2002). As well, individuals who are homozygous to this polymorphism (TT genotype) have specific enzymatic activities at 37º C that range from 40% to 50% of control values; heterozygotes (CT genotype) are in the intermediate range. Homozygous individuals also have residual activities after heating at 46ºC that are approximately 35% of residual activity in controls; heterozygotes fall in an intermediate range (Rozen, 1997).

MTHFR is a flavoprotein, which uses flavin adenine dinucleotide (FAD), a riboflavin derivative, as a cofactor and nicotinamide adenine dinucleotide phosphate (NAD[P]H) as an electron donor. Studies with MTHFR from *E. coli* have shown that a A177V mutation, which is homologous to A222V in humans, yields an enzyme that loses its FAD cofactor on dilution about 11 times faster than the wild-type enzyme. Flavin is necessary for

transferring electrons between NADH and CH2-H4folate, therefore, flavin release is followed by loss of activity (Yamada et al., 2001).

A meta-analysis has found that the global frequency of T allele and TT genotype is around 24% and 7.7%, respectively, with the highest prevalence being found in the Caucasian population (Yadav et al., 2017). The inheritance of this mutation may potentially inhibit methyl donor availability in the conceived offspring. Additionally, this may alter DNA methylation processes during in utero development when critical DNA methylation reprogramming events in the germline occur, or even during the maintenance of DNA methylation throughout the male's lifetime. The persistence of altered DNA methylation patterns in the germline may therefore increase the risk for infertility. Moreover, individuals with the *MTHFR* 677TT genotype are predisposed to various disorders in some populations (Liew & Gupta, 2015; Gong et al., 2015).

To assess the consequence of MTHFR deficiency on fertility in mice, a model with a targeted mutation in the *Mthfr* gene has been used. *Mthfr*$^{-/-}$ mice on a C57BL/6 background were healthy and fertile, but had reproductive abnormalities including lower testis weights, lower sperm counts and increased proportions of abnormal seminiferous tubules (Chan et al., 2010; Lawrance et al., 2011). In contrast, mice on a BALB/c background had more severe reproductive defects and were infertile, revealing an important role of genetic background, as suggested by human studies.

Based on studies that have shown improved sperm counts after 6 months of high dose folic acid supplementation, men with infertility are often prescribed with up to 5mg of daily intake; >10 times the daily recommended dose (Ebisch, et al., 2007). Folic acid supplemented diet (FASD) is known to prevent neural tube defects (NTDs); for this

reason, in many countries grain products and other sources of supplements, such as multivitamins, have been fortified with folic acid. While this has been successful in lowering cases of NTDs, it has also led to higher intake of folic acid than recommended.

Importantly, high dose folic acid supplementation has been shown to induce a "pseudo-MTHFR deficiency" associated with abnormal SAM levels and SAM/SAH ratios in the liver in mice, inducing folate deficiency by over-supplementation (Christensen et al., 2015). Another study using a high dose of folic acid (5mg/day for six months) in men with idiopathic infertility showed an unexpected loss of methylation across the sperm DNA methylome. In addition, individuals with the *MTHFR* 677TT genotype were more susceptible to epigenetic alterations in important disease-related genes after supplementation (Aarabi, et al., 2015). In a different study 5-methylcystine capture sequencing (Chan et al., 2019) was used to evaluate the variation in the human sperm DNA methylome that was representative of altered folate metabolism. Compared to *MTHFR* 677CC men, those with the 677TT genotype had both increases and decreases in sperm DNA methylation, with a predominance of hypermethylation in the TT genotype. Treatment with high-dose folic acid supplementation exacerbated sperm hypomethylation in *MTHFR* 677TT men. In both cases, >80% of altered methylation was found in sites with 20-80% levels of methylation (termed dynamic sperm CpGs), uniquely measured by their assay.

In a follow-up mouse study to investigate the epigenomic impact and underlying effects of folic acid on male germ cells, wild-type and *Mthfr*$^{+/-}$ mice were fed high-dose folic acid or control diets for 6 months (Aarabi et al., 2018). Reduced representation bisulfite sequencing (RRBS) showed sperm DNA hypomethylation in *Mthfr*$^{+/-}$ mice on

10xFASD diets. Wild-type mice showed sperm hypomethylation only with a very high dose (20xFASD) of folic acid administered for 12 months. Testicular MTHFR protein levels decreased significantly in wild-type mice on the 20xFASD diet but not in those on the 10xFASD diet, suggesting a possible role for MTHFR deficiency in sperm DNA hypomethylation. To study the *Mthfr* genotype effects alone, *Mthfr*$^{+/+}$ and *Mthfr*$^{+/-}$ mice on the control diet were compared. No significant difference was detected in global methylation levels, however, sperm DNA of *Mthfr*$^{+/-}$ mice contained 1348 differentially methylated tiles (DMTs), most of which were hypermethylated (Aarabi et al., 2018).

Other adverse effects from high folic acid supplementation and folic acid deficiency (FD) have been described in mouse studies. Paternal lifetime (prenatal and postnatal) FD caused aberrant DNA methylation in sperm and an increase in birth defect rates in the subsequent generation (Lambrot et al., 2013). In a different study, lifetime FASD and FD altered DNA methylation in sperm, decreased sperm counts, had negative reproductive effects. Paternal lifetime FASD and FD resulted in imprinted gene DNA methylation perturbations in offspring (Ly et al.,2017). More recently, postnatal paternal FD was shown to alter histone H3 lysine 4 trimethylation (H3K4me3) in sperm at developmental genes and putative enhancers (Lismer et al.,2021). As well, a study assessed the effect of FD, 10 FASD and 20 FASD on F1 spermatogonia from PND 6 male mice, where the exposure to the different diets would only target the prenatal window of development (Chan et al., 2022 [submitted to Andrology]). The results obtained with RRBS revealed a large number of DMCs, both hyper- and hypomethylated, with the 20FASD group showing vast loss of methylation. In the same study, lifetime exposure to these diets also showed DMCs for all groups with both increases and decreases in mature F1 sperm DNA

methylation, however, the total number of DMCs was fewer than in spermatogonia. These results demonstrated that the folic acid diets impact DNA methylation patterns in germ cells of the F1 male mice, and spermatogonia were more severely affected than sperm.

1.6 Rationale and Hypothesis

Many studies have been interested in assessing the impact caused by MTHFR deficiency on various disorders and the genetic mutations found within this crucial enzyme. The most common mutation of this gene which is involved in significant health risks is the *MTHFR 677 C>T* polymorphism. Thus far, studies have investigated the effect of MTHFR deficiency using mice heterozygous and homozygous for a *Mthfr* null allele and have yielded significant results, however, the mechanisms by which the *MTHFR 677C>T* mutation interacts to increase the risk of certain disorders remains unclear. Therefore, it is an important task to characterize the newly developed mouse model that accurately reflects this polymorphism. This newly developed mouse was generated at The Jackson Laboratory using CRISPR/cas9 and our collaborators have determined that important biochemical parameters accurately mimic what is found in humans, such as decreased MTHFR enzyme activity and increased plasma homocysteine levels (Reagan et al., accepted in *Journal of Cerebral Blood Flow & Metabolism*, 2022).

Proper DNA methylation interactions are crucial for optimal germ cell and embryonic function and development, for this reason, the availability of methyl donors such as dietary folate and the normal functioning of its metabolic pathway are essential. In this study we assessed the impact of this mutation on the sperm epigenome by examining the global DNA methylation levels of *Mthfr 677*CC and TT mice using the Illumina Mouse Methylation BeadChip and WGBS. In addition, we investigated the impact

of a folate deficient diet (7-fold folate deficient, 0.3 mg/kg diet) and a moderate dose of folic acid supplementation (5-fold folic acid supplemented, 10 mg/kg diet) on the sperm epigenome.

The goal of this study was to determine if this newly developed mouse model faithfully reproduces results previously reported in humans and knockout mice. **We hypothesized that the *Mthfr* 677TT mice would show sperm DNA hypermethylation on a control diet, compared to the CC genotype. Moreover, we expected the FD diet to exacerbate the potential detrimental outcomes caused by the TT genotype. Lastly, we hypothesized that the moderate levels of folic acid supplementation would rescue the TT mouse sperm epigenome, causing it to be similar to the wild-type genotype.**

In addition to global sperm DNA methylation, imprinted genes were assessed for any perturbations caused by the different genotypes and diets. As well, body weights and reproductive organ weights were compared among all different groups for any significant differences. Sperm counts and testes histology were also studied in all groups to determine if they were normal.

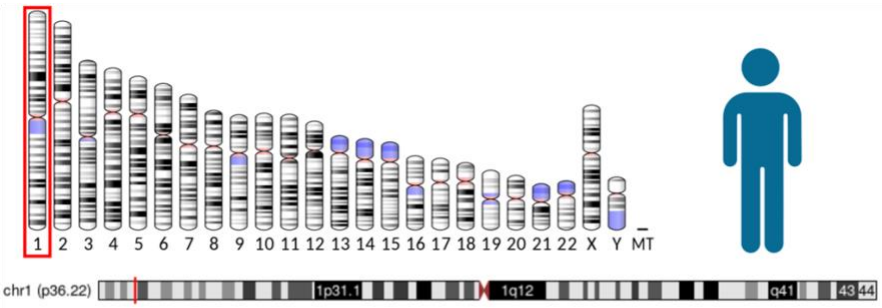
CHAPTER 2: MATERIALS AND METHODS

2.1 Animals and diets

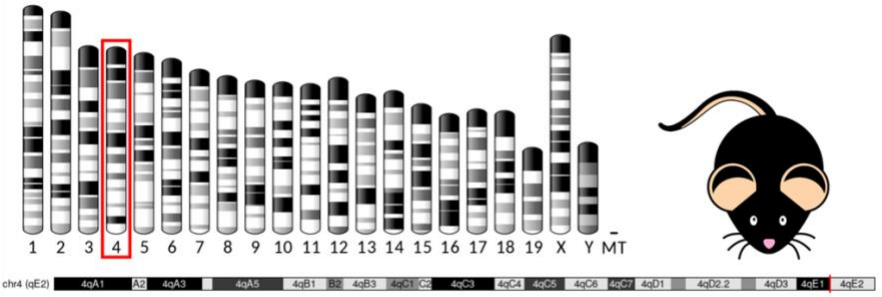
MODEL-AD (model organism development and evaluation for late-onset Alzheimer's disease) created a novel knock-in (KI) strain carrying the *Mthfr* 677C>T allele on the C57BL/6J background. The animals were generated at The Jackson Laboratory (JAX) as part of the IU/JAX/Pitt/Sage MODEL-AD Center. Experiments were conducted in accordance with policies and procedures described in the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health and were approved by the JAX Institutional Animal Care and Use Committee. All mice are congenic on the C57BL/6J (JAX# 000664) (B6) strain and bred and housed in a 12/12-hour light/dark cycle and fed LabDiet 5K52 (LabDiet), a standard natural ingredient diet containing 1.9 mg/kg folic acid, and 9.0 mg/kg riboflavin. In mice, based on numbering from ENSMUST00000069604.15, the 806C>T causes a codon change GCC>GCT, leading to an A262V mutation in the 5,10-methylenetetrahydrofolate reductase gene that is equivalent to the 677C>T polymorphism and corresponding A222V mutation in humans (Fig 3). Initially, the *Mthfr* 677C>T was engineered by CRISPR onto LOAD1 (B6.APOE4.Trem2*R47H) (JAX #30922). To create the *Mthfr* 677C>T strain (JAX #31464), APOE4 and Trem2*R47H were bred out by successive backcrosses to B6. To produce experimental animals, mice heterozygous for *Mthfr* 677C>T were intercrossed to produce litter-matched male and female *Mthfr* CC, *Mthfr* CT and *Mthfr* TT mice (Reagan et al., accepted in *Journal of Cerebral Blood Flow & Metabolism*, 2022).

The experimental design (Fig. 4) consists of six groups of mice that were sorted out according to their genotype and diet (n= 6/group). The groups consist of male

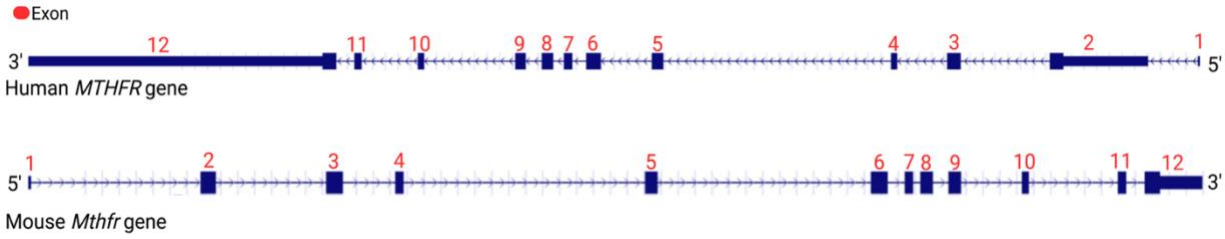
C57Bl/6J *Mthfr* TT (Jackson laboratories) mice and wild-type mice as control (*Mtfhr* CC). These were exposed to control (CD, 2mg folic acid/kg diet), folate-deficient (FD, 0.3mg folic acid/kg diet) and moderate dose (FASD, 5-fold control: 10 mg folic acid/kg diet) folic acid-supplemented diets starting at four weeks of age and maintained on their respective diets for four months; they were then sacrificed, and tissues and organs were collected. The treatment duration was 4 months for liver function studies from our collaborators. For this project, this duration also allowed the exposed mice to complete ~2 rounds of spermatogenesis and sperm maturation. The amino acid defined diets (Envigo) follow AIN93 recommendations. Succinylsulfathiazole was added to inhibit folate production by intestinal bacteria. Diet consumption was measured by weighing food weekly. Experiments and procedures were carried out in conformity with the Canadian Council on Animal Care guidelines and the study was approved by the McGill University Animal Care Committee. Mice were maintained under a 12 Light: 12 Dark cycle in a temperature and humidity-controlled environment at the Research Institute of the MUHC's pathogen-free animal facility with food and water ad libitum.



Position: hg38 chr1:11,785,723-11,805,964
Size: 20,242
Total Exon Count: 12 Strand: -



Position: mm10 chr4:148,041,191-148,059,551
Size: 18,361
Total Exon Count: 12 Strand: +



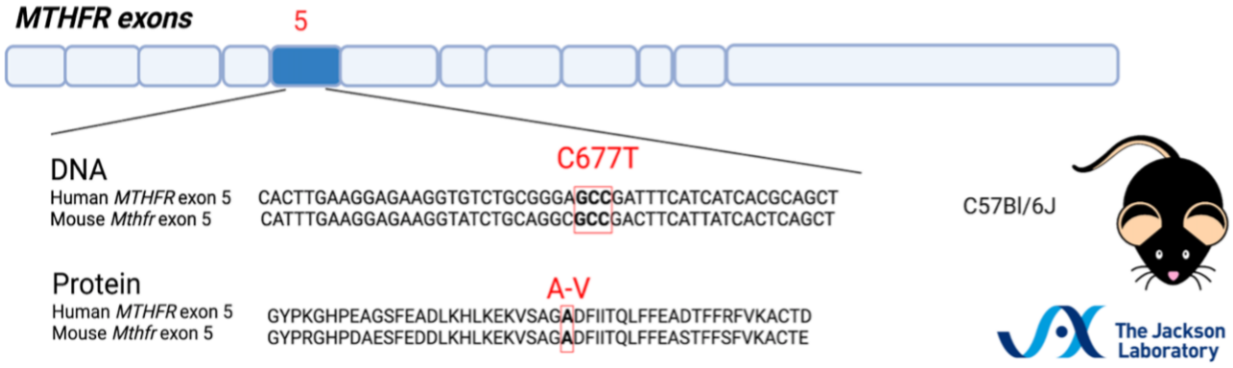
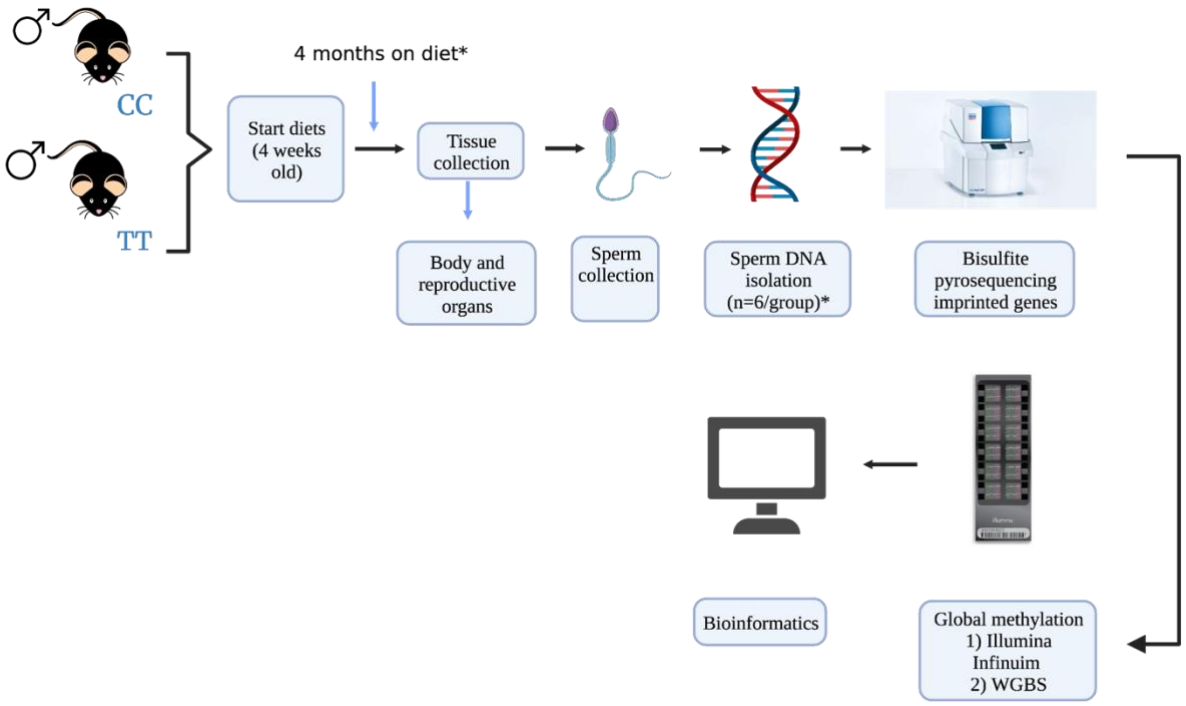


Figure 3. *MTHFR* gene and SNP location.

Mouse and human *MTHFR* DNA and protein sequences are highly conserved around the 677-variant located on exon 5. The *Mthfr* 677C>T variant was introduced into the mouse genome by editing the codon GCC to GCT using CRISPR/cas9 leading to an A262V mutation in the *Mthfr* gene. To create the *Mthfr* 677C>T strain (JAX#31464), APOE4 and Trem2*R47H were bred out by successive backcrosses to C57BL/6 mice. To produce experimental animals, mice heterozygous for *Mthfr* 677C>T (*Mthfr* CT) mice were intercrossed to produce litter-matched male and female *Mthfr* CC, *Mthfr* CT and *Mthfr* TT mice (Reagan et al., accepted in Journal of Cerebral Blood Flow & Metabolism, 2022)



Diet groups	Folic acid content	Human equivalent
Control (CD)	2 mg/kg diet	0.4 mg/day
7x Folate-deficient (FD)	0.3 mg/kg diet	0.06 mg/day
5x Folic acid supplemented (FASD)	10 mg/kg diet	2 mg/day

Figure 4. Experimental design

Male CC and TT mice were weaned onto their respective diets at 4 weeks of age and were kept on that diet for 4 months. They were then sacrificed, and their body and reproductive organs were collected and weighed for comparison. Sperm was collected from all mice and a subset of samples (n=6/group) was selected for sperm DNA isolation. Imprinting control region methylation levels (*H19, Snrpn* and *Peg1)* were analyzed by bisulfite pyrosequencing. Global sperm DNA methylation was analyzed with Illumina Mouse Methylation BeadChip for control (CD) and folate deficient (FD) samples and whole-genome bisulfite sequencing (WGBS) for CD and folic acid supplemented (FASD) samples (n=6/group respectively). Finally, downstream analyses were performed.

2.2 Tissue collection and sperm DNA isolation

Mice were euthanized by being anesthetized with isoflurane followed by CO2 followed by cervical dislocation. Next, emptied seminal vesicles, paired testes and epididymides were removed and weighed. The right testis was frozen in liquid nitrogen, and the left testis was immersed in Bouin's fixative for at least 24h, washed and placed in 70% EtOH. Mature spermatozoa from the paired cauda epididymides were collected as previously described (Chan et al., 2012). Briefly, cauda epididymides were excised, trimmed free of fat, cut in PBS, and agitated on a shaker for 10 mins at room temperature. Supernatant was filtered through a 40µm cell strainer. This step was repeated, and the filtered supernatant was combined with the previous one. Sperm suspension was centrifuged to pellet cells then washed with a hypotonic solution (0.45% NaCl) a minimum of 3 times. Samples were spun down for 5mins in-between washes. After the last spin, supernatant was removed conserving as much as possible and sperm pellets were immediately frozen at -80ºC until use.

2.2.1 Sperm DNA extraction

Sperm were lysed in a buffer containing a final concentration of 150 mM Tris, 10 mM ethylenediaminetetraacetic acid, 40 mM dithiothreitol (DTT), 2 mg/mL proteinase K, and 0.1% sarkosyl detergent and were incubated overnight at 37°C. DNA was then extracted using the QIAamp® DNA Mini kit (Qiagen) according to the manufacturer's protocols.

2.3 Genotyping

Genotype confirmation was done for the mouse sperm DNA samples (n=6/group). Primers (Table 1) were obtained and a PCR with different temperature gradients was

done to find the optimal annealing temperature for the primers. The results showed optimal annealing at 55,57ºC and 60.6ºC. Next, a PCR was run using all 24 samples plus controls, and an agarose gel was run to ensure proper PCR amplification and no contamination. *Sal*1 is a restriction enzyme that recognizes the sequence “GTCGAC”, if a sperm DNA sample had a TT or CT genotype the enzyme would cut the amplified DNA sequence and create two bands of 351bp and 242bp, CC genotype would show bands of 593bp. An enzyme mix was prepared using 1uL of *Sal*1, 2.5uL of buffer, and 16.5 uL water. Next, 20uL were added to 5uL of each individual PCR product. The resulting 25uL were mixed, spun down, and incubated at 37ºC overnight to ensure proper enzyme digestion. PCR amplicons were analyzed by agarose gel electrophoresis and visualized by ethidium bromide. The agarose gel was loaded with 5uL of digested product and run for 30 minutes at 140V to then obtain a gel image.

2.4 Testicular hemocytometric sperm counts and histology analysis

Frozen right testes were weighed and homogenized (Polytron, Brinkmann Instruments) for 2x15s inside a 5ml homogenizing solution on ice (0.9% NaCl, 0.1% thimerosal and 0.5% Tween-20) and elongated spermatozoal heads, which are resistant to homogenization, were counted as described by Kelly et al. (2003). Counts for 4 hemocytometer chambers were averaged. The number of spermatozoal heads per testis and per gram of testis were calculated.

2.4.1 Testicular histology analysis

Fixed left testes were bisected laterally and embedded in paraffin, cut into 5-µm-thick serial transverse cross sections and every sixth section was stained with hematoxylin and eosin (H&E). Four to eight sections (each section with at least 100

tubules) from each animal were examined under an upright fluorescence Leica DM6000 B microscope and imaged using a digital camera with Leica Application Suite with DMControl for Windows software. Normal tubules were identified as having all germ cell types in the majority of the seminiferous epithelium but could contain some mild alterations (e.g., small and few vacuoles). Additionally, complete spermatogenesis was determined by assessing whether all 12 stages of the seminiferous epithelium were present across the sections analyzed, where various cell types created in the differentiation of primordial germ cells to mature sperm should be seen in an organized manner. Abnormal tubules, as described in Karahan et al., 2021, were identified as possessing at least one of the following criteria: (1) asymmetric distribution of germs cells within the tubule, germ cells present in a part of the tubule cross-section but not on the other part; (2) tubule with early germ cells but missing spermatids; (3) tubule with spermatozoa and spermatids but no apparent early germ cells; (4) Sertoli cell only phenotype: tubule has only Sertoli cells, no noticeable germ cells within the tubule. The number of abnormal tubules in 100 seminiferous tubules examined for each mouse was used to calculate the proportion of abnormal tubules.

2.5 Bisulfite pyrosequencing DNA methylation analysis at imprinting control regions (ICRs)

Sperm DNA samples were screened for possible somatic cell contamination through bisulfite pyrosequencing of the imprinting control regions (ICRs) for imprinted paternally methylated (*H19*) and maternally methylated genes (*Snrpn* and *Peg1/Mest).* Assessment of imprinted gene methylation also provided an indication of the impact of

Mthfr genotype and folic acid diets on these genes which are known to be critical for offspring development.

For all pyrosequencing assays, 500 ng of sperm DNA underwent bisulfite conversion with the EpiTect® bisulfite kit (Qiagen) according to the manufacturer's protocol. Bisulfite PCR was conducted using primers and pyrosequencing was performed as described in Dejeux et al. 2009. Briefly, regions of interest were PCR amplified with one of the primers being biotinylated. Capture of the biotinylated strand was performed with streptavidin-coated sepharose beads and washed using the PyroMark® Q24 Vacuum Workstation (Qiagen). A sequencing primer was annealed to the isolated captured template strand and the pyrosequencing reaction was conducted using the PyroMark® Q24 kit (Qiagen) as per the manufacturer's protocol. Primer sequences utilized for PCR amplification were acquired from previously published studies and can be found in Table 1 below. For somatic cell contamination screening, sperm DNA samples were considered to be contaminated if methylation across all analyzed CpGs in imprinted genes deviated from the expected high levels of methylation for *H19* (>90%) and low levels for *Snrpn* and *Peg1*/*Mest* (<10%). All studied samples met the criteria for purity and thus, no sperm DNA samples were excluded.

ICR/Gene	# CpGs analyzed within ICR	Primer sequences	Reference
H19	6	Fwd: 5'-GGGGGGTAGGATATATGTATTTTT Rev: 5'-biot-ACCTCATAAAACCCATAACTATAAAATCAT Seq: 5'-GTGTGTAAAGATTAGGG	Adapted from Susiarjo et al., 2013
Snrpn	5	Fwd: 5'-TTGGTAGTTGTTTTTTGGTAGGAT Rev: 5'-biot-TCCACAAACCCAACTAACCTTC Seq: 5'-GTGTAGTTATTGTTTGGGA	Adapted from Susiarjo et al., 2013
Peg1	5	Fwd: 5'-GGTTGGGTTTGGATATTGTAAAG Rev: 5'-biot-TTCCCTAAAATTCTAACAACCTAAACA Seq: 5'-ATTGTAAAGTTAAAGTTGTAGTAAA	de Waal et al., 2014
Mthfr	-	Fwd: AGCTTGGACACCCACAATTC Rev: AGCTCCAGAGTGCTGGGATA	Dr. Rozen's lab and collaborators

Table 1. Primers for *Mthfr* genotyping and for bisulfite pyrosequencing used to assess DNA methylation at imprinting control regions (ICRs) in mice.

2.6 Illumina Mouse Methylation BeadChip data analysis

2.6.1 Illumina Array data quality control

After confirming that the sperm DNA samples were free of somatic cell contamination, control and folate deficient samples (6/group) were sent to Genome Quebec CES (Montreal, QC, Canada) for DNA methylation analysis using the Illumina Infinium™ Mouse Methylation BeadChip. This recently developed array offers an extensive coverage of genic regions. It interrogates >285K methylation sites per sample at single nucleotide resolution and achieves >98% reproducibility for technical replicates, this array was first validated in Garcia-Prieto et al., 2022. It also provides quantitative methylation measurement at the level of single CpG–sites and offers powerful resolution for understanding epigenetic changes. In addition, it provides balanced coverage of CpG islands, transcription start sites (TSS), enhancers, imprinted loci, gene body regions, repetitive element regions, lamina-attachment domains, CTCF binding sites as described by the manufacturer.

Raw data were normalized, and quality control was performed by using the ENmix package (Bioconductor), this package provides a set of quality control, preprocessing/correction, and data analysis tools for Illumina Methylation Beadchips. It includes functions to read in raw IDAT data, background correction, dye bias correction, probe-type bias adjustment and other tools as described by the manufacturer.

2.6.2 Coverage of Illumina probes within ICRs

Custom tracks were imported into UCSC genome browser for the Genome Reference Consortium Mouse Build 38 (mm10) showing Illumina Mouse Methylation BeadChip probes. Imprinted control regions were searched using the following

coordinates: *H19,* chr7:142578768-142586025; *Snrpn,* chr7: 60002495-60006697; *Peg1/Mest,* chr6: 30734932-30739966. The total number of probes covering single CpGs within each ICR was counted and calculated as a percentage of coverage. Genome browser screenshots were saved from the imprinted gene locations and zoomed out coordinates. These were then edited using PowerPoint software to highlight the corresponding ICR of each imprinted gene.

2.6.3 ICR methylation beta values

ICR locus coordinates were obtained for 24 imprinted genes from a previously published study (Strogantsev, R. et al., 2015). The coordinates were published as NCBI/mm9 genome assembly, so they were converted to mm10 using the Lift Genome Annotations tool from UCSC genome browser. The list included 4 paternally methylated imprinted genes, *Igf2/H19*, *KvDMR1*, *Rasgrf1* and *Zdbf2*, and 20 maternally methylated imprinted genes, *Gpr1*, *Nnat*/*Peg5*, *Mcts2*/*H13*, *Gnas_ProXL*, *Gnas_Ex1A*, *Fkbp6*, *Peg10*, *Mest*, *Nap1l5*, *Snrpn*, *Peg3*, *Inpp5f_v2*, *Cdh15*, *Zac1* / *Plagl1*, *IG-DMR*/*Dlk1-Meg3*, *Grb10*, *Zrsr1*, *Peg13*, *Igf2r*, *Impact*. Illumina Mouse Methylation Array data was then intersected, using Bedtools software (Quinlan and Hall, 2010), with the newly obtained mm10 coordinates to retrieve specific ICR methylation beta values. The resulting list was used to calculate individual and mean beta values for each different group of samples (4 groups with 6 samples each).

2.6.4 Analysis of variance and post hoc analyses.

After normalization and quality control analysis of Illumina Mouse Methylation BeadChip probes, 277,466 of them remained significant. With these probes, an analysis of variance (ANOVA) was performed with RStudio software. Subsequently, multiple

pairwise comparisons were done by performing Tukey's HSD (honestly significant difference). The groups were compared as, CC FD vs CC CD, TT FD vs TT CD, TT CD vs CC CD, TT FD vs CC FD and TT FD vs CC CD, where CC and TT indicate the genotype, CD (control diet) and FD (folate deficient diet). With these analyses we were able to identify CpG sites showing differential methylation (DMCs).

2.6.5 Genome annotation

Significant DMCs were annotated with HOMER software version 4.9.1. Hypermethylated and hypomethylated DMCs were annotated separately. Values were exported into Excel software to calculate genomic annotation values and convert them into percentage values. These were then graphed using GraphPad Prism 9 software.

2.6.6 Repeat sequences and evolutionary young retrotransposable elements

A list of repeat sequences in the mouse genome was obtained from UCSC genome browser using the Table Browser tool which allows one to retrieve and export data from the genome browser annotation track database. Here the genome selected was "Mouse", the selected group was "Variation and Repeats", and the track selected was "RepeatMasker". Next, the "get output" option was selected in order to download the list of repeat sequences. The list was intersected with each different group comparison as well as the background. The resulting lists were opened in Microsoft Excel software to calculate the proportion of each different repeat sequence (LINE, SINE, LTR, simple repeats, other repeats, and non-repeats). A list of young retrotransposable elements subfamilies was obtained from Karahan et al., 2021, specifically L1MdA_I, L1MdTf_I, L1MdTf_II, L1MdA_II, L1MdGf_I, L1MdTf_III, L1MdGf_II and L1MdA_III. These were

filtered and calculated from each different comparison and the results were graphed using GraphPad Prism 9.

2.6.7 Regional analysis of differentially methylated CpG sites (DMCs)

DMCs from all the different Tukey's comparisons were merged into regions to search for larger-scale regional effects. Hypermethylated and hypomethylated DMCs were merged separately. Neighboring DMCs located within 100bp, 250bp, 500bp, 1000bp, 5000bp, and 10,000bp were merged using the "merge" function of Bedtools, a software that allows the combination of overlapping or "book-ended" features in an interval file into a single feature which spans all the combined features, as described by the developer (Quinlan and Hall, 2010). DMCs that lacked any neighboring DMCs remained as single CpGs, and those that merged >1 DMC were identified as differentially methylated regions (DMRs).

2.6.8 Gene ontology enrichment analysis and reduced visualization

Gene ontology (GO) term pathway enrichment was first performed on the list of DMCs using the R package topGO (version 2.42.0) from Bioconductor, then, all significant terms discovered were imported to the R package Reduce + Visualize GO (rrvgo) which helps reduce GO terms by identifying redundant terms based on semantic similarity. The results were graphed using GraphPad Prism 9.

2.7 Whole-genome bisulfite sequencing (WGBS)

For a more comprehensive analysis and to examine the effects of folic acid supplementation on the *Mthfr* 677CC and TT mice, whole-genome bisulfite sequencing (~20 million CpGs covered) was performed at McGill Genome Centre. A representative subset of samples was chosen (in total 24) per group, that is, 6 CC CD, 6 TT CD, 6 CC

folic acid supplemented (FASD), and 6 TT FASD. As well, 6 control diet samples that were analyzed by the Illumina Mouse Methylation BeadChip were included here, 3 CC CD and 3 TT CD (included in the 6 samples per group described above).

Between 2 and 3 µg of sperm DNA per sample were sent to ensure the best quality of library. The WGBS library was created using the KAPA High Throughput Library Preparation kit (Roche/KAPA Biosystems) as previously described (Chan et al., 2019). Briefly, sperm DNA was spiked with 0.1% (w/w) unmethylated λ and pUC19 DNA (Promega). DNA was sonicated and fragment sizes of 300-400bp were controlled on a Bioanalyzer DNA 1000 LabChip (Agilent). Following fragmentation, DNA-end repair of double-stranded DNA breaks, 3'-end adenylation, adaptor ligation, and clean-up steps were done according to KAPA Biosystems's protocol. The resulting bisulfite DNA was quantified with OliGreen (Life Technology) and amplified with 9-12 PCR cycles using the KAPA HiFi HotStart Uracil + DNA Polymerase kit (Roche/KAPA Biosystems) according to suggested protocols. The final WGBS library was purified using Agencout AMPure Beads (Beckman Coulter), validated on Bioanalyzer High Sensitivity DNA LabChip kits (Agilent) and quantified by PicoGreen (Thermofisher). The samples were sequenced over two lanes on the Illumina NovaSeq60000 S4 v1.5, PE150, 12 samples were lane. Sequencing data was processed with the GenPIpes Methyl-Seq pipeline. This includes protocols to align reads to the reference genome, call methylation, remove duplicates, and generate metric files.

2.7.1 Comparing control samples analyzed by both Illumina Array and WGBS

To determine common CpGs overlapped between datasets from control samples analyzed by both technologies (3CC and 3TT samples), they were intersected with the

intersectBed function of Bedtools. The resulting lists were used to perform a Pearson correlation analysis on each sample on RStudio (version 4.2.1) and generate scatterplots.

2.7.2 Differential methylation analysis

DNA methylation analysis was performed using methylKit software (Bioconductor), an R package for DNA methylation analysis and annotation from high-throughput bisulfite sequencing, with the ability to carry out operations such as differential methylation analysis, sample clustering and annotation, and visualization of DNA methylation events (Akalin et a., 2012). Here, methylKit was used for the identification of differentially methylated tiles (DMTs, 100 bp/tile) with ≥10x CpG coverage in all samples and a difference of at least 5% methylation between compared groups. A Benjamini-Hochberg false discovery (FDR)-based method for P-value correction and passing the q-value threshold (q = 0.01) were used.

2.7.3 Downstream analyses of WGBS data

Further analyses with WGBS data were performed as was described previously for the Illumina Mouse Methylation BeadChip data, such as imprinted genes analysis, analysis of DMTs, genomic distribution, proportion of repeat sequences and young retrotransposons, gene ontology enrichment analysis and reduced visualization, and lastly, corresponding genes of DMTs.

2.8 Statistical analyses

Data were analyzed using GraphPad Prism 9 (GraphPad Prism version 9.4.0 for Windows, GraphPad Software, San Diego, California USA) and are presented as mean ± SEM or absolute values, as indicated. Methylation variance at each ICR for bisulfite pyrosequencing and Illumina Mouse Methylation BeadChip results were determined by

averaging individual variances at all CpGs within diet groups. Analysis of variance heatmaps, Tukey's test heatmaps, beta difference/methylation difference histograms and Pearson correlation tests and scatterplots were generated using RStudio (version 4.2.1). $P \leq 0.05$ was considered significant for all analyses.

CHAPTER 3: RESULTS

3.1 Effects of *Mthfr* genotype and folic acid diets on body weights and reproductive organ weights.

Body weight of adult mice was measured as an indicator of general health. Similar weights were observed among all different genotype and diet groups, indicating that neither significantly affected the general health of male mice (Fig. 5A). In a study where they measured enzymatic activity in different tissues, such as liver, brain, kidney, testis and ovary, the testis was shown to possess the highest level of MTHFR expression (Chen et al., 2001). MTHFR deficiency and the folic acid diets used in this study might have an effect in the testes and in other reproductive organs, such as emptied seminal vesicles and epididymides. For this reason, they were all evaluated as indicators of normal testicular function, number of germ cells and testosterone production.

A previous study has shown a significant decrease of approximately 20% in testis weight in *Mthfr*$^{-/-}$ mice compared to *Mthfr*$^{+/+}$ mice on a C57BL/6 background (Chan et al., 2010). In contrast, no testicular weight differences were observed when comparing *Mthfr*$^{+/+}$ and *Mthfr*$^{+/-}$ mice. Consistent with those observations, here, the comparisons revealed no significant changes among all different groups (Fig. 5B, C, D).

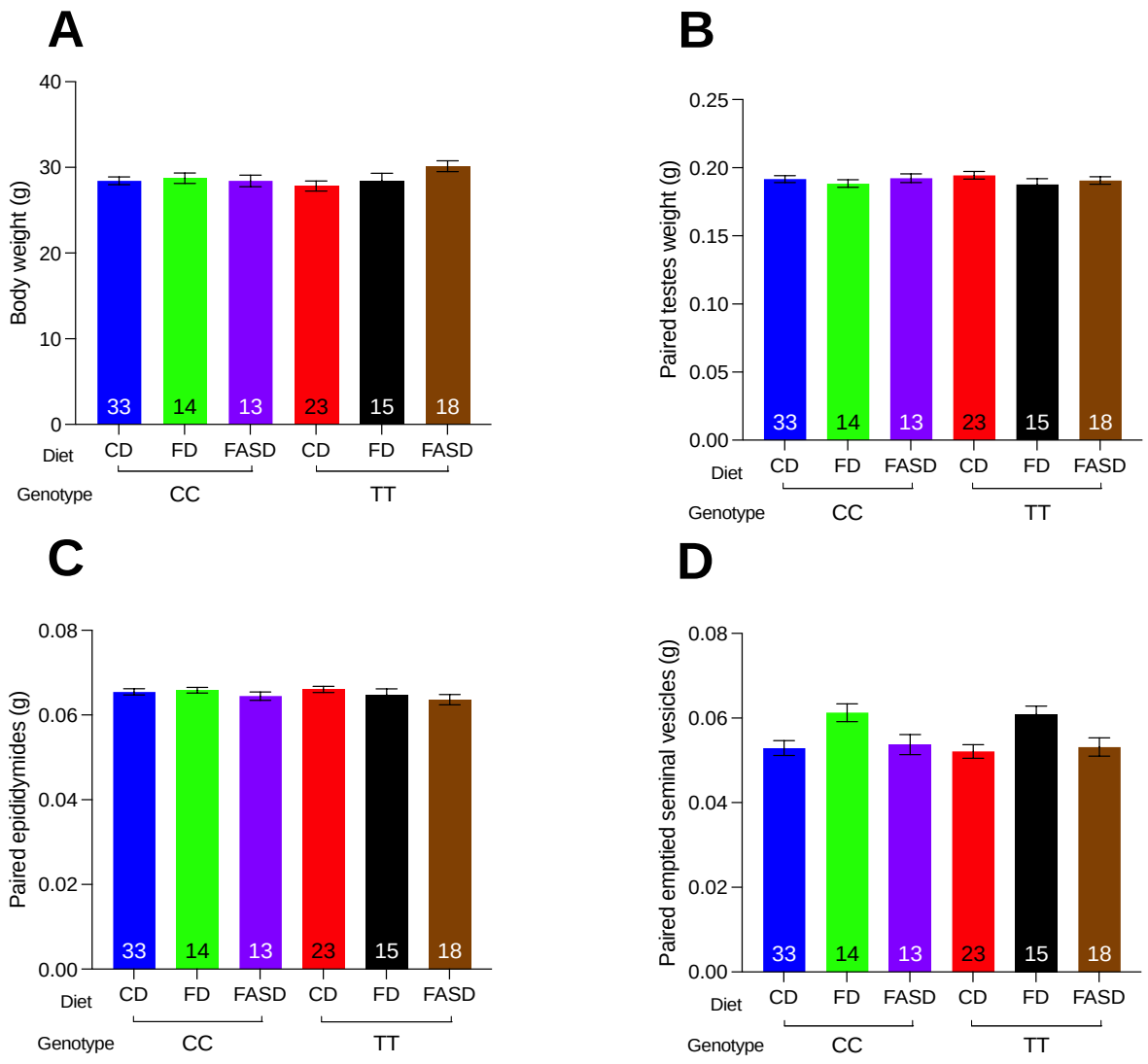


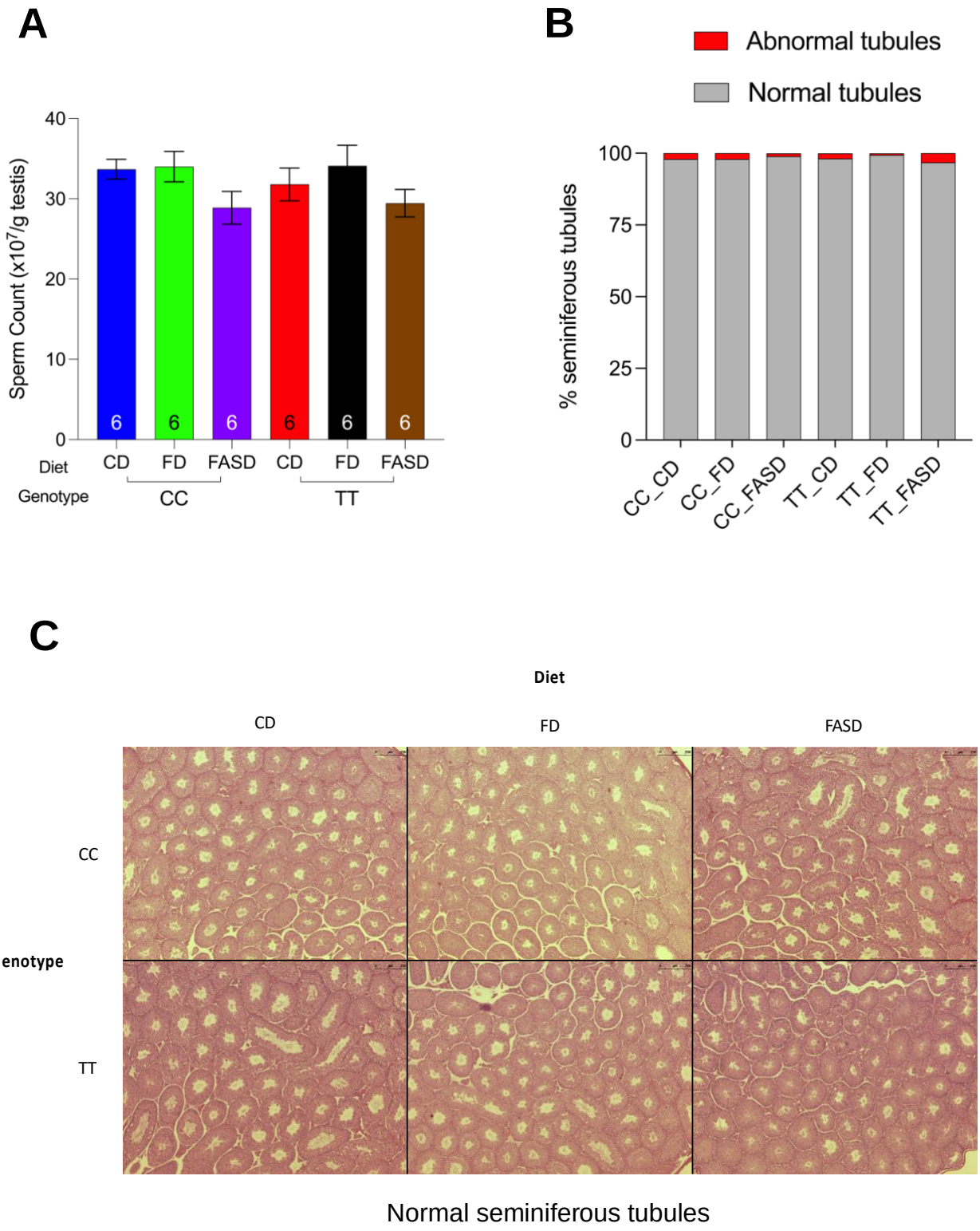
Figure 5. *Mthfr* genotype and folic acid diets do not affect body and reproductive organ weights.

Body weights and reproductive organ weights across all different groups were unaffected by the different genotypes or folic acid diets. Numbers in bars indicate the sample sizes. Values shown as mean +/- SEM. One-way ANOVA with multiple comparisons.

3.2 Effects of *Mthfr* genotype and folic acid diets on sperm counts and testes histology.

MTHFR plays a major role in DNA methylation processes. A reduction of this enzyme can lower SAM levels and potentially affect male germ cells as a consequence. Even though previous studies have found normal sperm counts and normal testicular histology in *Mthfr*$^{+/-}$ mice on a C57BL/6 background (Chan et al., 2010), we sought to determine if any abnormalities existed in this study. Total sperm in each testis was calculated and the results showed a slight variation in sperm counts, with the FASD in both genotypes appearing to be lower than the rest of the groups. The observed differences were not statistically significant (Fig. 6A), to determine this, a one-way ANOVA with multiple testing was performed and showed no significant differences. As well, unpaired t-test revealed a p value of 0.07 for both FASD groups compared to the control group. Next, the proportion of normal seminiferous tubules was determined by counting the number of abnormal tubules in 100 seminiferous tubules examined for each mouse in every group. The vast majority of the observed samples displayed normal histological parameters, such as normal seminiferous tubule shape, normal arrangement of Sertoli cells and visible spermatozoa in the lumen of the tubules (Fig. 6C).

Additionally, complete spermatogenesis and normal staging was observed in all samples by distinguishing all 12 stages of the seminiferous epithelium in the sections analyzed. Minor abnormalities observed included uneven seminiferous tubules, tubules with asymmetric distribution of germ cells and tubules with missing spermatids. The observations revealed that neither genotype or folic acid diet significantly affected testicular histology or sperm counts of these mice (Fig. 6B, C).



Normal seminiferous tubules

Figure 6. *Mthfr* genotype and folic acid diets do not affect total sperm counts or testes histology.

(A) Sperm counts (shown as $x10^7$/g testis) were not significantly impacted by the different genotypes or folic acid diets; data is shown as mean +/- SEM (One-way ANOVA with multiple comparisons). Numbers in bars indicate the sample sizes. (B) Proportion of normal and abnormal testicular tubules across all groups (n=6/group) (One-way ANOVA with multiple comparisons). (C) Representative testicular histological cross-sections from each group of samples. Black scale bar located at the top right represents 750 μm.

3.3 Effects of *Mthfr* genotype and folic acid diets on DNA methylation of imprinting control regions (ICRs).

Normal DNA methylation patterns of imprinted genes during gametogenesis are critical for normal embryonic development. To determine the impact of the *Mthfr* 677C>T genotype and folic acid diets on sperm DNA, first, the methylation status of a paternally methylated ICR (*H19*) and two maternally methylated ICRs (*Snrpn* and *Peg1/Mest*) were assessed using bisulfite pyrosequencing. Simultaneously, this was useful to ensure that the selected samples (n= 6 per group) that would be sent for Illumina Mouse Methylation BeadChip analysis were free of somatic cell contamination. In sperm DNA, only one parental allele is present and methylation of paternally and maternally methylated loci are expected to be close to 100% and 0%, respectively. Bisulfite pyrosequencing analysis of sperm DNA revealed that the methylation status of the selected imprinted loci was within the normal range, independent of *Mthfr* genotype or folic acid diet (Fig. 7).

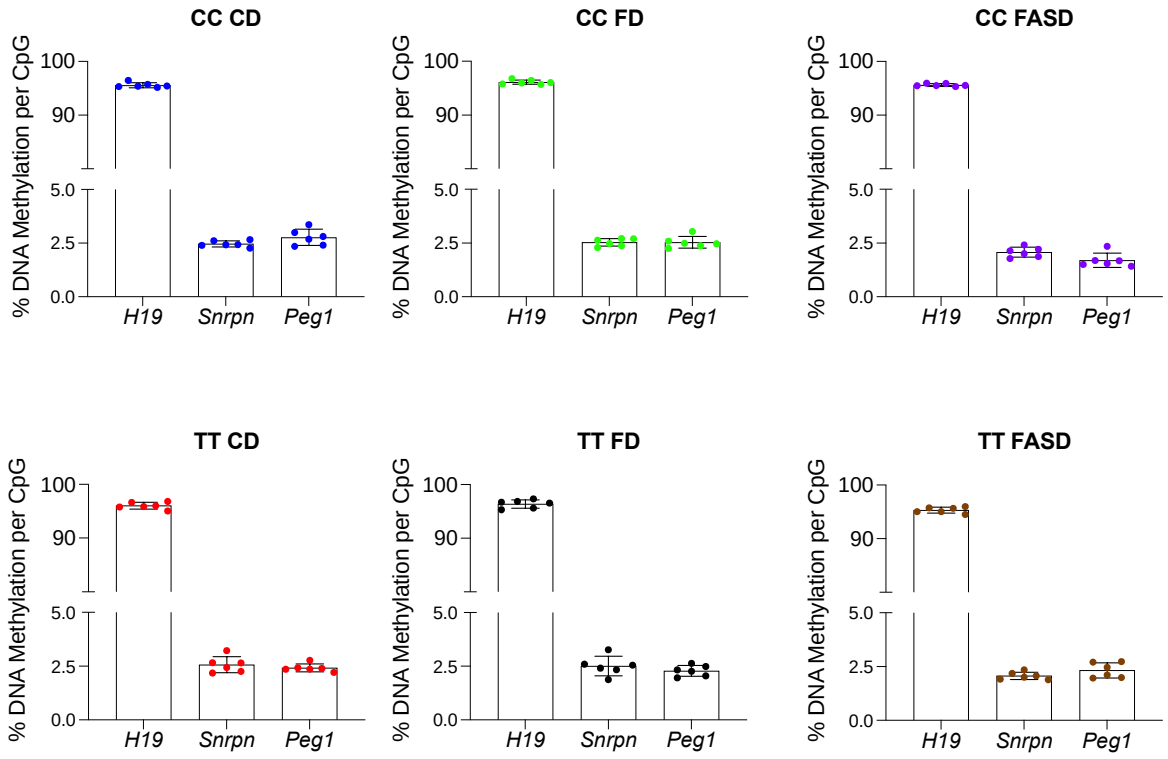


Figure 7. Bisulfite pyrosequencing analysis on imprinting control regions (ICRs) shows correct levels of methylation.

Methylation of germline ICRs was normal and was not significantly affected by TT genotype or folic acid diets. Bars show mean methylation levels (% mean of all samples ± SEM). Dots indicate methylation levels of individual CpGs.

3.4 Results obtained with Illumina Mouse Methylation BeadChip

3.4.1 ICRs coverage of paternally and maternally methylated imprinted genes

Having ensured that the subset of samples was free of somatic cell contamination and given that no alterations in DNA methylation were observed in imprinted genes, a more comprehensive analysis was performed using a recently developed array, Illumina Mouse Methylation BeadChip. The groups of samples analyzed with this approach were *Mthfr* 677CC and *Mthfr* 677TT on a control diet (CD) and folate deficient diet (FD), 6 samples per group were studied, in total 24. First, we looked in greater detail to the same imprinted genes assessed with bisulfite pyrosequencing. The ICR that showed the largest coverage was the paternally methylated imprinted gene *H19* (Fig. 8). With bisulfite pyrosequencing, typically only 6 CpGs are studied in average. For *H19*, the Illumina Array covered up to 40% of the whole ICR, with 41 CpGs covered out of 104. In the case of *Snrpn*, there are 74 CpGs within its ICR, the Illumina Array was able to cover 11 of them (15% coverage) (Fig. 9). For *Peg1/Mest*, 185 CpGs can be found within its ICR, the Illumina Array probes covered 39 CpGs, that is, about 21% of the whole ICR was covered (Fig. 10).

Next, we expanded the analysis to other known ICRs within the mouse genome. In total 24 of them were studied (Fig. 11) and the Illumina Array covered 23 of them (the maternally methylated gene *Fkpb6* was not covered). The genome tracks with the exact location of the remaining 20 ICRs studied, along with the respective coverage by the Illumina Array like in figures 8-10, are not included here but are available upon request. The results demonstrate that this Array is able to cover a much larger number of CpG sites within ICRs when compared to bisulfite pyrosequencing.

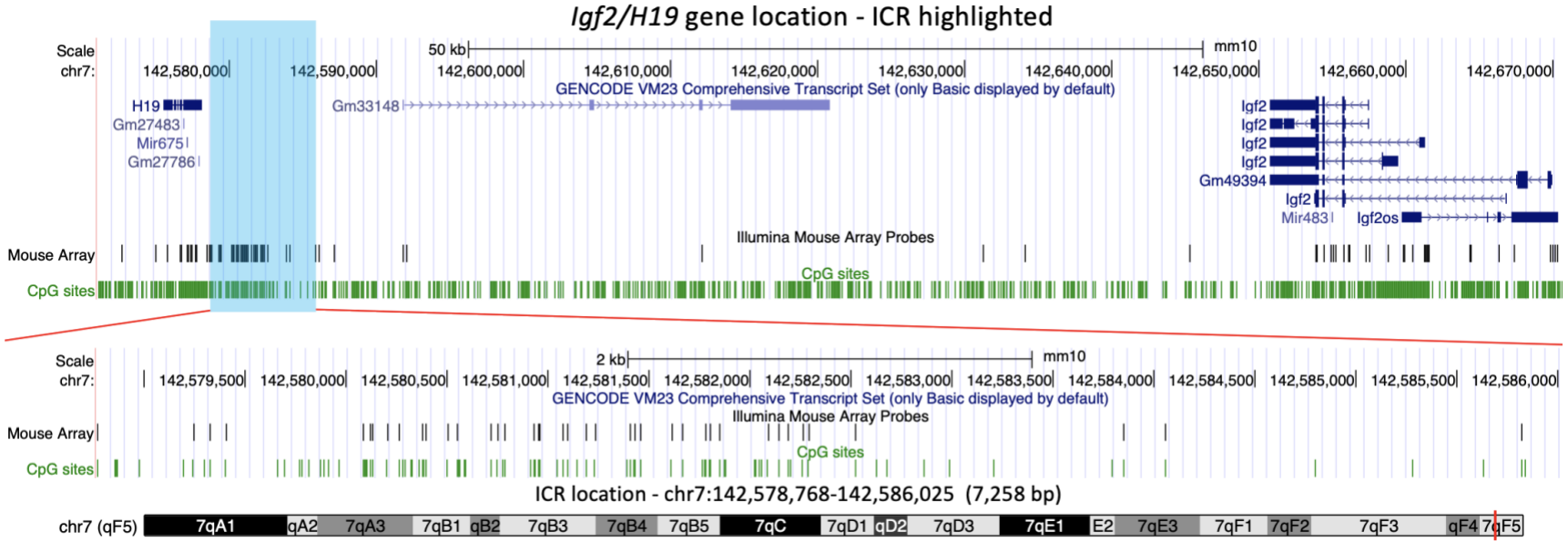


Figure 8. *H19* coverage by Illumina Mouse Methylation BeadChip probes.

USCS Genome Browser view of the paternally methylated gene *H19*. Custom tracks indicate CpG sites (green) and CpGs analyzed by Illumina Mouse Methylation BeadChip (black). Blue box indicates location of paternally methylated ICR. Within the ICR of *H19* there are 104 CpGs in total (Table 2), the Illumina Array probes covered 41 of those (39.4% of the whole ICR). Bottom genome browser shot shows expanded ICRs. ICR, imprinting control region; bp, base pairs; UCSC, University of California, Santa Cruz.

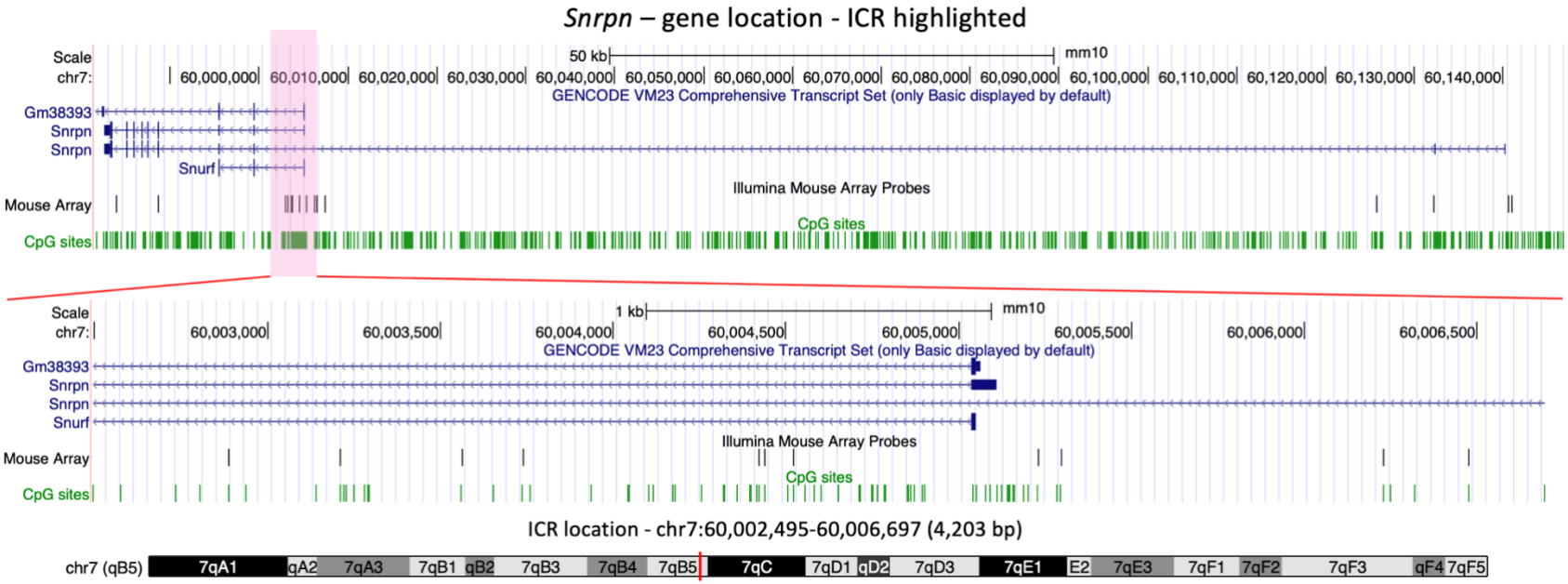


Figure 9. *Snrpn* coverage by Illumina Mouse Methylation Array probes.

USCS Genome Browser view of the maternally methylated gene *Snrpn*. Custom tracks indicate CpG sites (green) and CpGs analyzed by Illumina Mouse Methylation BeadChip (black). Pink box indicates location of maternally methylated ICR. Within the ICR of *Snrpn* there are 74 CpGs in total (Table 2), the Illumina Array probes covered 11 of those (15% of the whole ICR). Bottom genome browser shot shows expanded ICRs. ICR, imprinting control region; bp, base pairs; UCSC, University of California, Santa Cruz.

Figure 10. *Peg1/Mest* coverage by Illumina Mouse Methylation Array probes.

USCS Genome Browser view of the maternally methylated gene *Peg1/Mest*. Custom tracks indicate CpG sites (green) and CpGs analyzed by Illumina Mouse Methylation BeadChip (black). Pink box indicates location of maternally methylated ICR. Within the ICR of *Peg1/Mest* there are 185 CpGs in total (Table 2), the Illumina Array probes covered 39 of those (21% of the whole ICR). Bottom genome browser shot shows expanded ICRs. ICR, imprinting control region; bp, base pairs; UCSC, University of California, Santa Cruz.

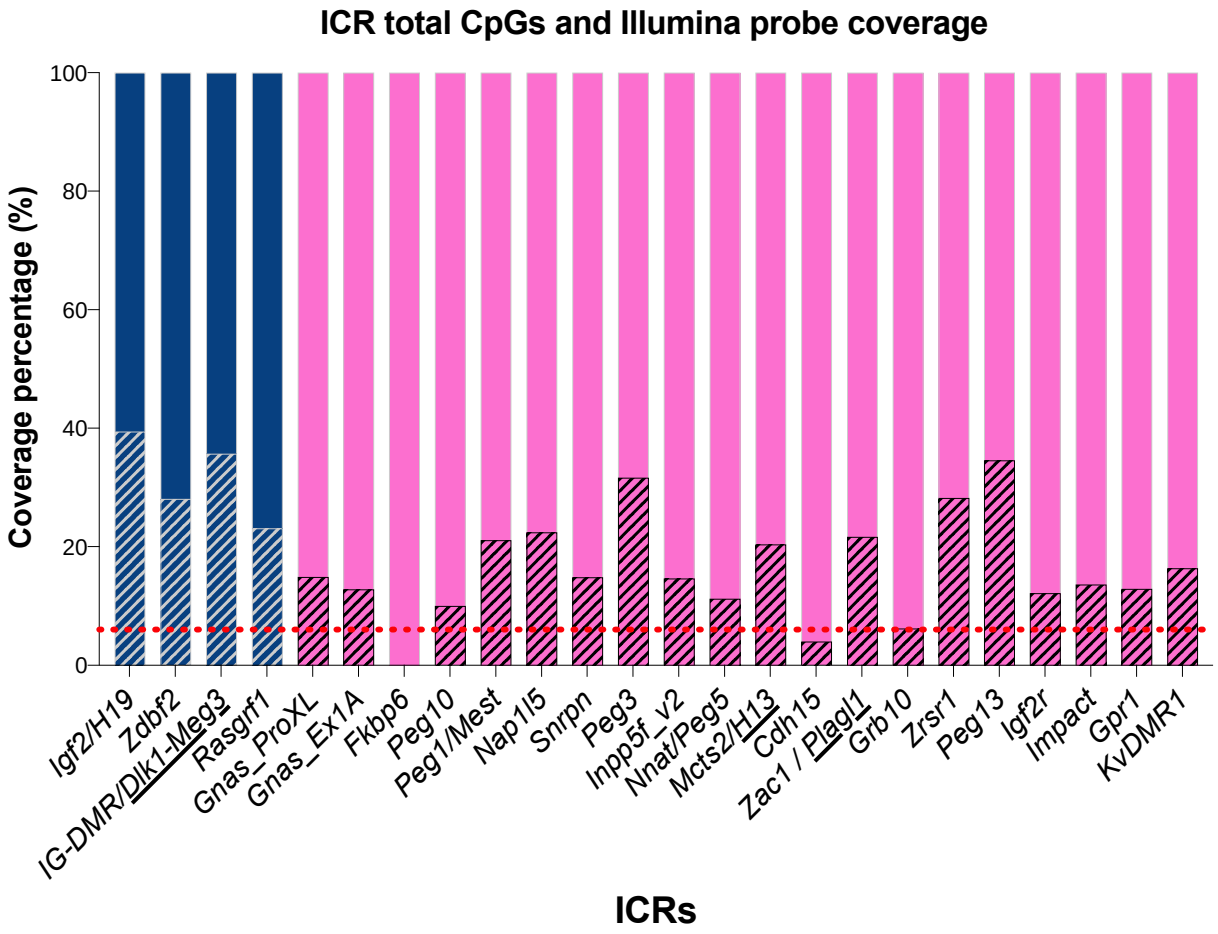


Figure 11. Percentage of ICR CpGs covered by Illumina Mouse Methylation BeadChip.

In total 23 known ICRs within the mouse genome were covered by the Illumina Array. Each bar represents the total amount of CpGs within an ICR shown in percentage, the hatched bars indicate the percentage of coverage by the Illumina Array probes. The number of CpGs located within each ICR and covered by the Illumina probes are indicated in Table 2. The red dotted line showcases the total amount of CpGs analyzed by bisulfite pyrosequencing for *H19*, *Snrpn* and *Peg1/Mest*. Blue bars represent paternally methylated ICRs whereas pink bars indicate maternally methylated ICRs. ICR, imprinting control region.

Imprinted gene	CpGs within the ICR	Illumina probes coverage
Igf2/H19	196	55
IG-DMR/Dlk1-Meg3	196	70
Zdbf2	117	27
Rasgrf1	117	27
Gnas_ProXL	349	52
Gnas_Ex1A	195	25
Peg1/Mest	185	39
Peg10	170	17
Grb10	161	10
Peg3	147	48
Impact	147	20
Zrsr1	131	37
Zac1 / Plagl1	120	26
KvDMR1	104	17
Igf2r	99	12
Nnat/Peg5	98	11
Inpp5f_v2	82	12
Peg13	81	28
Cdh15	75	3
Snrpn	74	11
Nap1l5	58	13
Mcts2/H13	54	11
Gpr1	31	3
Fkbp6	13	0

Table 2. Number of CpGs covered by Illumina Mouse Methylation BeadChip probes.

Paternally methylated imprinted genes are highlighted in blue. Maternally methylated imprinted genes are highlighted in pink. Bold letters represent the imprinted genes also assessed by bisulfite pyrosequencing. Illumina probes coverage indicate the number of single probes that covered single CpGs withing a determined imprinted gene.

3.4.2 ICRs methylation beta values of FD and CD samples.

Methylation levels from all 23 covered ICRs were determined. Paternally methylated ICRs (*Rasgrf1, Igf2/H19, IG-DMR/Dlk1-Meg3)* showed the expected high levels of methylation (Fig 12A). Although mean methylation at imprinted loci was unaffected by genotype or diet, a decrease was seen in *Zdbf2* which displayed areas with low methylation values, possibly due to the location of these probes. Notwithstanding, all maternally methylated ICRs showed the expected low levels of methylation (Fig. 12B) further confirming that neither *Mthfr* genotype nor folic acid diets affected the methylation status of imprinted genes.

3.4.3 Analysis of variance of significant Illumina Array probes.

To find out whether the differences between all different groups were statistically significant we ran an analysis of variance (ANOVA). Before quality control, the Array data included 287,693 probes. After normalization and after excluding control probes, probes that did not pass quality control, and probes without beta values (beta values are the estimate of methylation level using the ratio of intensities between methylated and unmethylated alleles. These are between 0 and 1 with 0 being unmethylated and 1 fully methylated), the total remaining significant probes were 277,466. These were utilized to perform the ANOVA. This analysis revealed 15,037 probes that were statistically significant and clearly separated each different genotype and diet group into their own clusters (Fig. 13).

3.4.4 Post hoc analyses using significant ANOVA probes.

With the ANOVA's significant probes, multiple pairwise comparisons were done by performing Tukey's tests to find out which specific groups' means were different. We first

examined the genotype effect. *Mthfr* 677TT CD (control diet) vs *Mthfr* 677CC CD (Fig 14). On a control diet, this comparison (TT CD vs CC CD) revealed the largest effect out of all groups. It demonstrated a predominance of loss of methylation, with 1731 hypermethylated and 2361 hypomethylated DMCs. When comparing TT vs CC both on a folate deficient diet, the comparison showed 1471 hypermethylated and 1295 hypomethylated DMCs; this group had the lowest effect of all. Next, we evaluated the effect of the folate deficient diet on both genotypes. On the CC genotype, the analysis revealed 1622 hypermethylated and 2159 hypomethylated DMCs (CC FD vs CC CD). On the TT genotype, this diet showed 1746 hypermethylated and 1409 hypomethylated DMCs (TT FD vs TT CD). Between these two groups, the CC genotype showed a predominance of hypomethylation when on FD, whereas the opposite happened on the TT genotype. Lastly, when comparing the extremes, that is TT FD vs CC CD, it emerged as the second most affected group, with 1692 hypermethylated and 2259 hypomethylated DMCs.

We next examined the genotype effect on the CD in greater detail since it contained the most DMCs. A heatmap of the most affected comparison was performed. With a total of 4092 DMCs, the analysis clearly separated both sample groups (TT CD vs CC CD) (Fig. 15A). The largest effect was hypomethylation (1731 hypermethylated probes and 2361 hypomethylated). The majority of probes demonstrated ≤0.10, equivalent to ~10% change in DNA methylation (Fig 15B). All other comparisons showed a similar trend, with the majority of methylation beta differences being ≤10, meaning that these groups all showed small changes in sperm DNA methylation. For all groups, the data separated well when creating a heatmap.

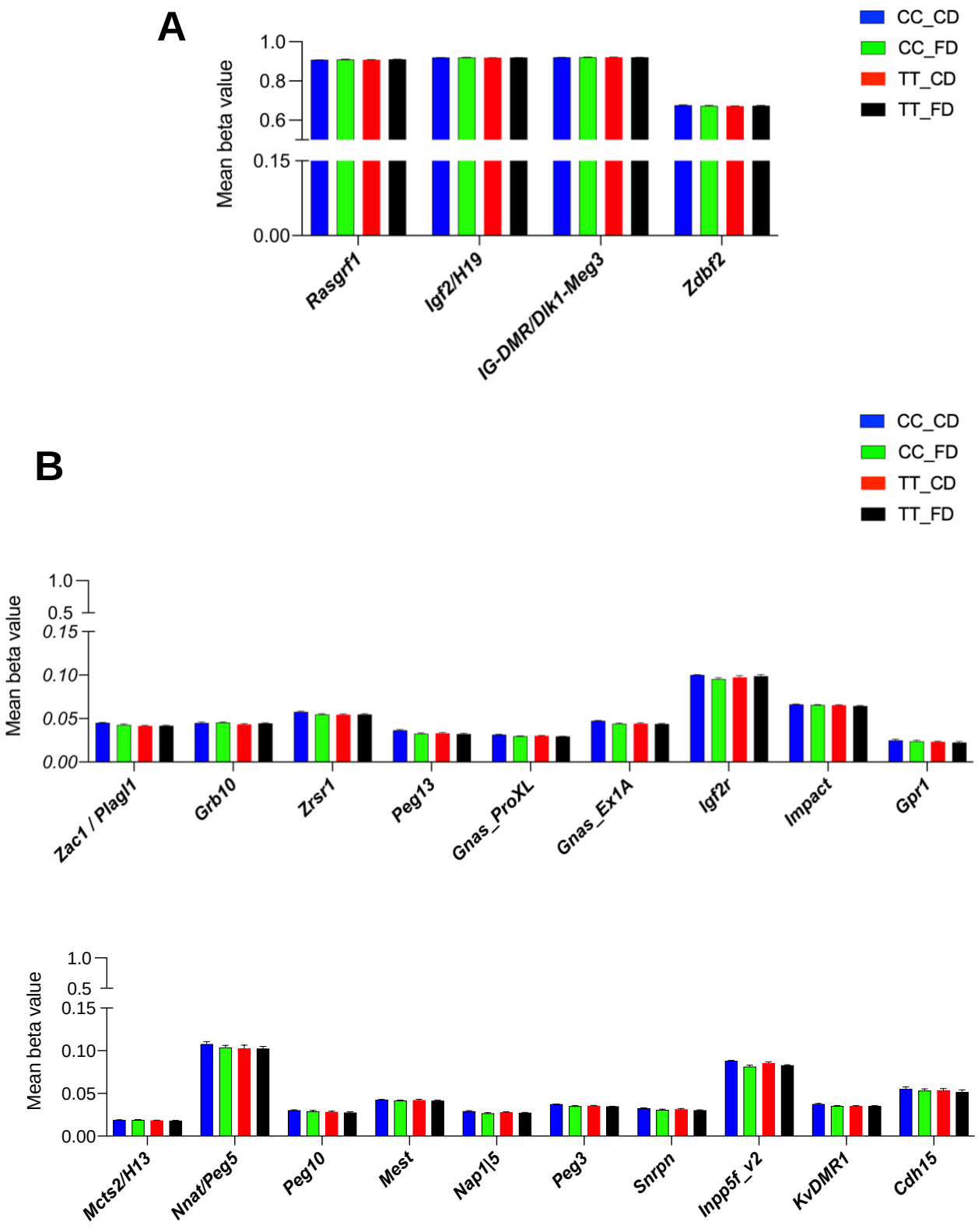

A
CC_CD
CC_FD
TT_CD
TT_FD
Mean beta value
1.0
0.8
0.6
0.15
0.00
Rasgrf1
Igf2/H19
IG-DMR/Dlk1-Meg3
Zdbf2
B
CC_CD
CC_FD
TT_CD
TT_FD
Mean beta value
1.0
0.5
0.15
0.10
0.05
0.00
Zac1 / Plagl1
Grb10
Zrsr1
Peg13
Gnas_ProXL
Gnas_Ex1A
Igf2r
Impact
Gpr1
Mean beta value
1.0
0.5
0.15
0.10
0.05
0.00
Mcts2/H13
Nnat/Peg5
Peg10
Mest
Nap1l5
Peg3
Snrpn
Inpp5f_v2
KvDMR1
Cdh15

Figure 12. ICR methylation results from Illumina Array show correct levels of methylation.

(A) Paternally methylated imprinting control regions (ICRs), *Rasgfr1, Igf2/H19 and IG-DMR/Dlk1-Meg3* showed expected high levels of methylation, *Zdbf2* revealed probes with low values of methylation, **(B)** Maternally methylated ICRs displayed expected low levels of methylation and remained unaffected by different genotypes or folic acid diets. Bars shown as mean +/- SEM.

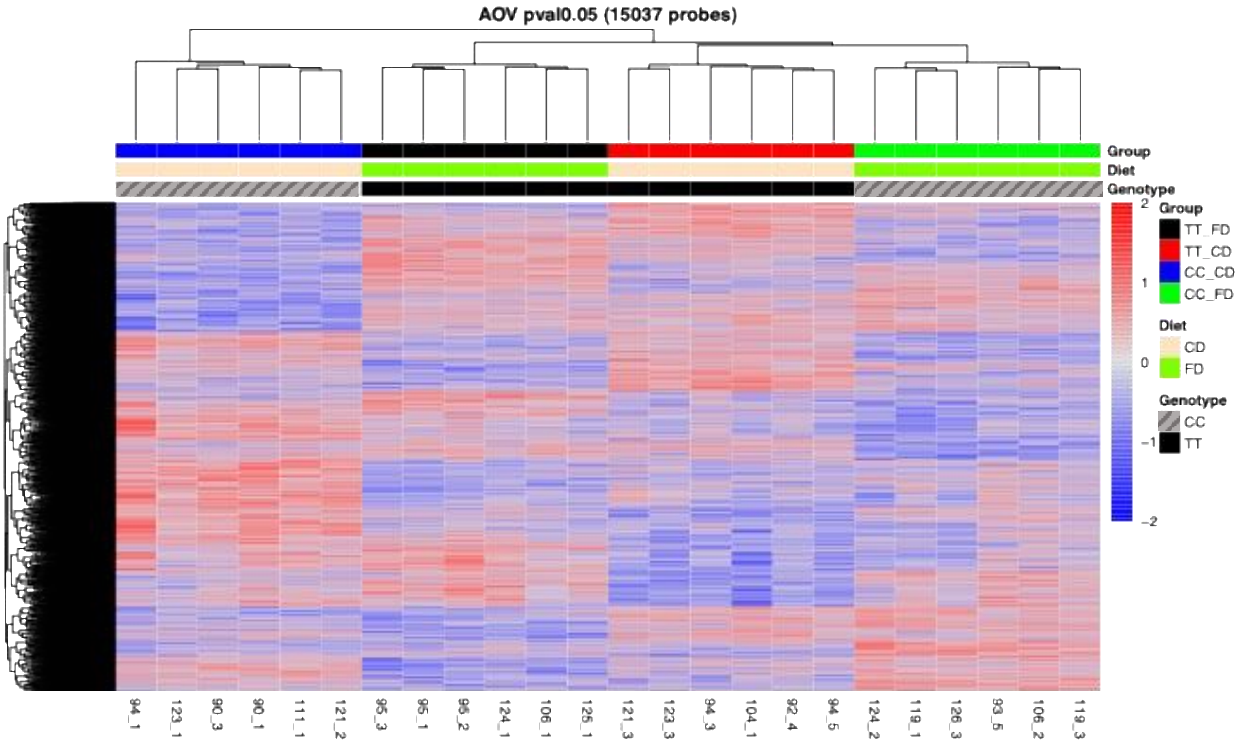


Figure 13. Analysis of variance of significant Illumina Array probes show clear separation of data.

After normalization and quality control analysis of Illumina Mouse Methylation Array probes, 277,466 of them remained significant. With these probes, an analysis of variance (ANOVA) revealed 15,037 significant probes that differentiated four groups according to genotype and diet. The heatmap shows the data clustered into each specific group according to different genotypes and folic acid diets

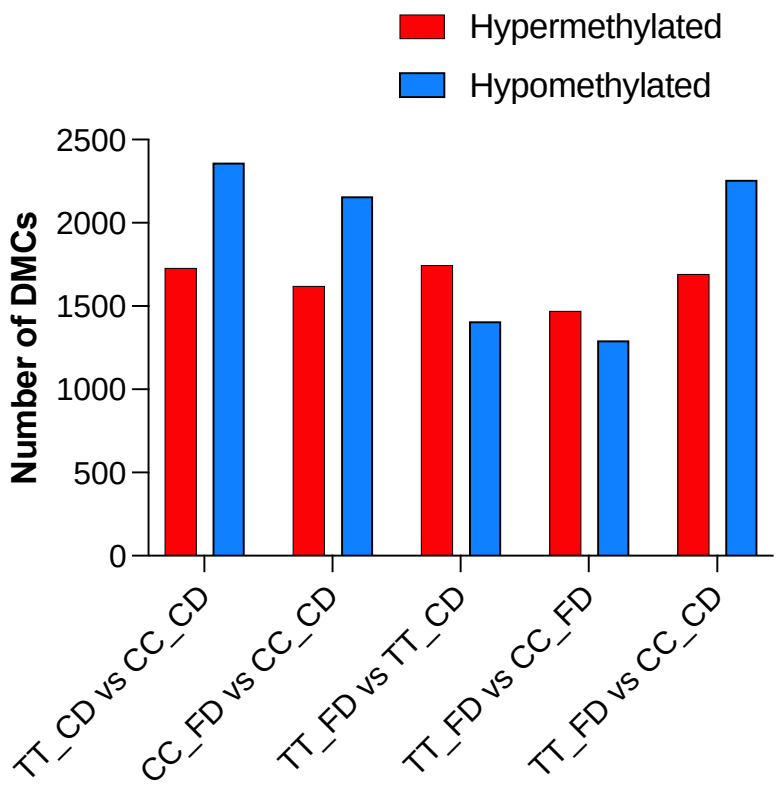
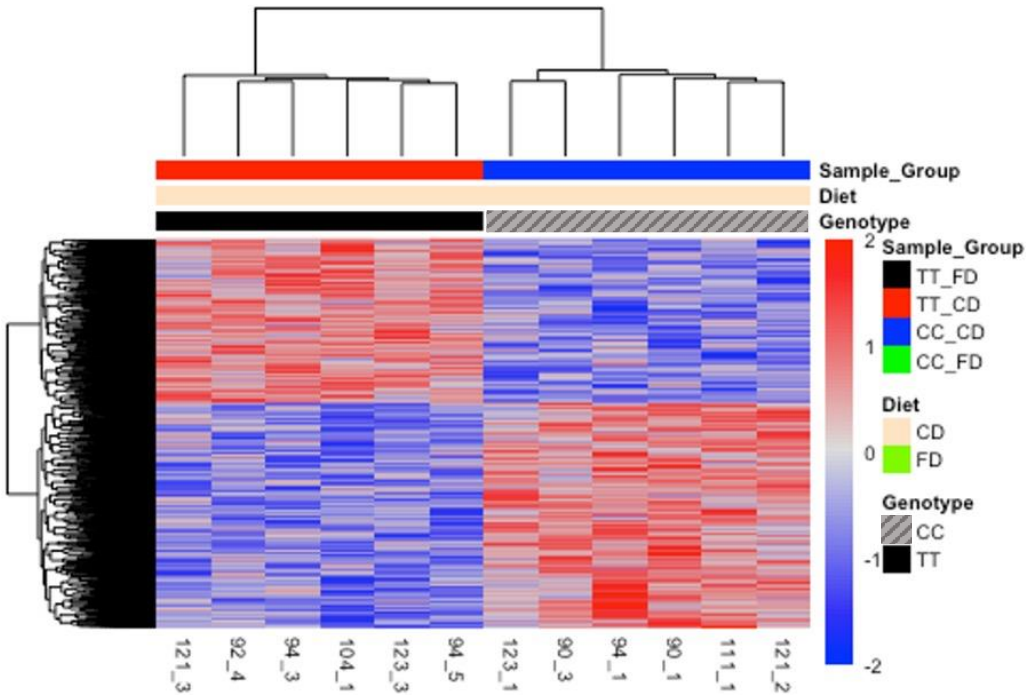


Figure 14. Summary of post hoc analyses.

Multiple Tukey's tests were performed using the significant probes from the ANOVA to look for significant differences in-between all genotype and diet groups. The largest effect was caused by the TT genotype on a control diet (4092 DMCs), with a higher effect on hypomethylation (2361 DMCs). ANOVA, analysis of variance; DMC, differentially methylated site, CD, control diet; FD, folate deficient diet; CC, wild-type genotype; TT, homozygous *Mthfr* 677 C>T polymorphism.

TT genotype effect on control diet (4092 probes)

A



B

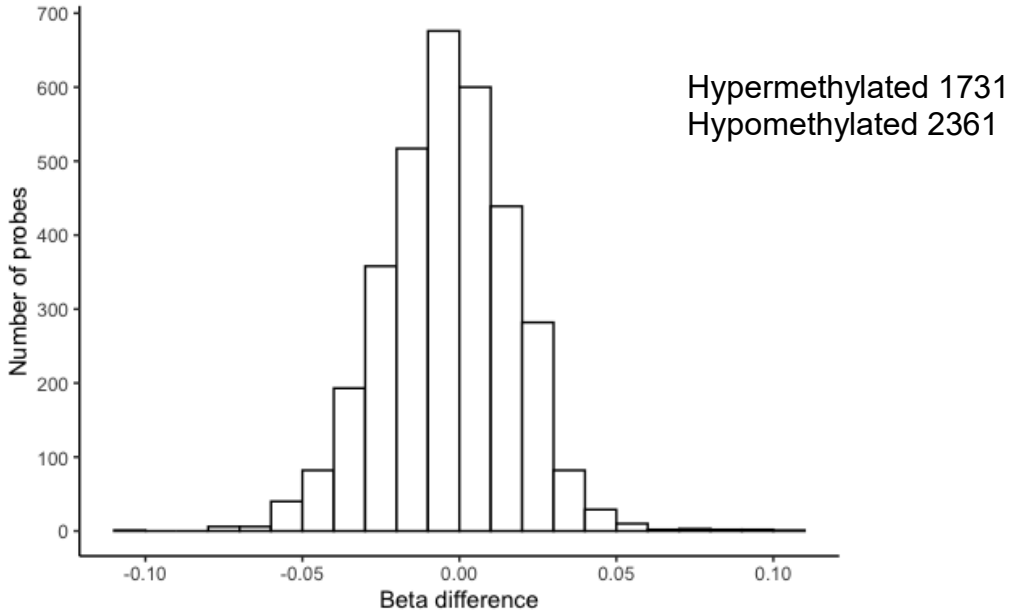


Figure 15. Heatmap of genotype effect on a control diet and beta difference histogram.

(A) Heatmap shows data separated into each different group according to genotype and folic acid diet. **(B)** Histogram showing the magnitude of difference in methylation shown as beta values on the X axis and the total number of probes on the Y axis, with a larger effect of hypomethylation.

3.4.5 Genome distribution of FD and CD samples.

Previous studies have shown that intergenic DNA methylation in sperm is the most susceptible genomic location to the loss of MTHFR, often followed by intronic areas (Aarabi et al., 2015; Aarabi et al., 2018; Chan et al., 2019; Karahan et al., 2021). In this study, significant DMCs were annotated with HOMER software. Background data distribution showed largely intronic areas (45%), followed by intergenic (33%), exons (12%) promoter-TSS (5%) and others (Fig. 16). In all comparisons, promoter areas displayed a larger representation when compared to the background (14-17%). To determine if these changes were significant, Chi squared tests with Yate's correction were performed. In the hypermethylated group, TT FD vs CC FD had a significant increase of promoter-TSS areas ($p=0.0005$) (Fig 16A). In the hypomethylated group the same was true for CC FD vs CC CD ($p=0.0025$) and TT FD vs CC CD ($p=0.0001$) (Fig 16B). Detailed annotation values are shown in Table 3.

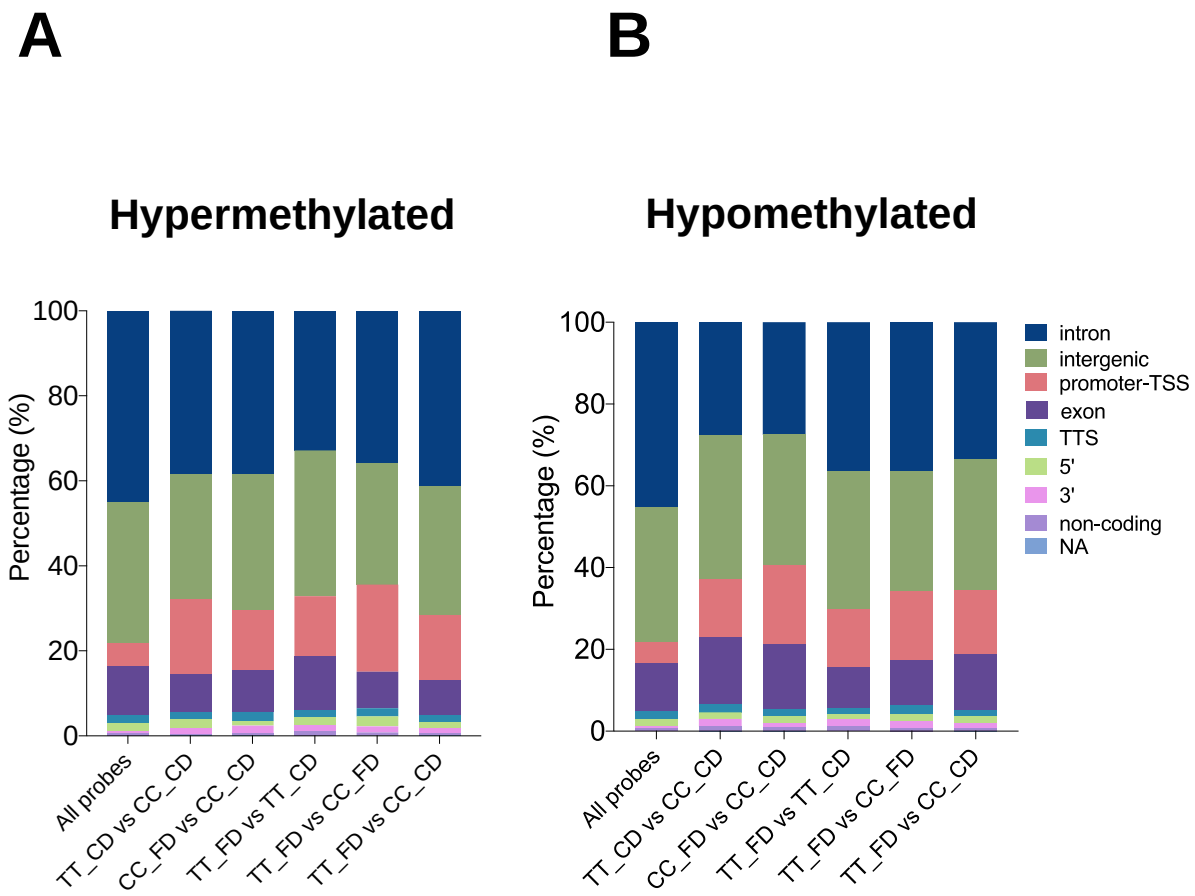


Figure 16. Genomic distribution of all probes and post-hoc comparisons.

The genomic distribution was similar among all different groups, with a larger representation of intronic, intergenic and promoter-TSS areas. **(A)** Hypermethylated DMCs. **(B)** hypomethylated DMCs. DMCs, differentially methylated sites.

Hypermethylated	Intron	Intergenic	Promoter-TSS	Exon	TTS	5'	3'	Non-coding	NA
All probes	45.20	32.97	5.26	11.76	1.79	1.89	0.55	0.58	0.00
TT CD vs CC CD	38.35	29.61	17.52	9.02	1.62	2.02	1.50	0.29	0.06
CC FD vs CC CD	38.53	31.87	14.18	9.80	2.16	1.05	1.85	0.49	0.06
TT FD vs TT CD	32.84	34.26	14.25	12.60	1.77	1.71	1.54	1.03	0.00
TT FD vs CC FD	35.94	28.46	20.52	8.56	1.97	2.31	1.43	0.75	0.07
TT FD vs CC CD	41.19	30.32	15.48	8.04	1.77	1.36	1.24	0.47	0.12
Hypomethylated	**Intron**	**Intergenic**	**Promoter-TSS**	**Exon**	**TTS**	**5'**	**3'**	**Non-coding**	**NA**
All probes	45.20	32.97	5.26	11.76	1.79	1.89	0.55	0.58	0.00
TT CD vs CC CD	27.71	35.08	14.32	16.31	1.95	1.65	1.86	1.06	0.04
CC FD vs CC CD	27.51	31.91	19.31	15.84	1.67	1.67	1.20	0.83	0.05
TT FD vs TT CD	36.31	33.90	14.23	9.91	1.63	1.13	1.70	1.20	0.00
TT FD vs CC FD	36.32	29.49	16.90	11.07	2.03	1.85	1.59	0.72	0.04
TT FD vs CC CD	33.39	32.03	15.80	13.72	1.57	1.52	1.24	0.68	0.05

Table 3. Genome annotation of Illumina Mouse Methylation Array.

Significant DMCs annotated with Homer software, hypermethylated and hypomethylated DMCs shown separately. Values are shown as percentage (%).

3.4.6 Proportion of repeat sequences from Illumina Mouse Methylation Array data.

Transposable elements are mobile repetitive sequences that make up a large portion of the genome, about 45% in humans (Lander et al., 2001) and 37.5% in mice (Waterson et al., 2002). Abnormal methylation at repeat sequences can result in the activation of transposable elements that have the capacity of disrupting the genomic stability. Analysis of repeat sequences by Illumina Array revealed a small proportion of these elements across all groups studied (Fig 17), likely because they are not well covered by this array. The repeat sequences with the highest percentage were LTR areas with a total of 4-5%. LINE and SINE sequences were very similar with a total of 3-4%. Likewise, evolutionary young retrotransposable elements which are known to escape reprogramming in germ cells and pre-implantation had scarce coverage with only about 0.2% of the totality (Fig. 18), specifically L1MdA_I, L1MdTf_I, L1MdTf_II, L1MdA_II, L1MdGf_I, L1MdTf_III, L1MdGf_II and L1MdA_III subfamilies were assessed (Karahan et al., 2021).

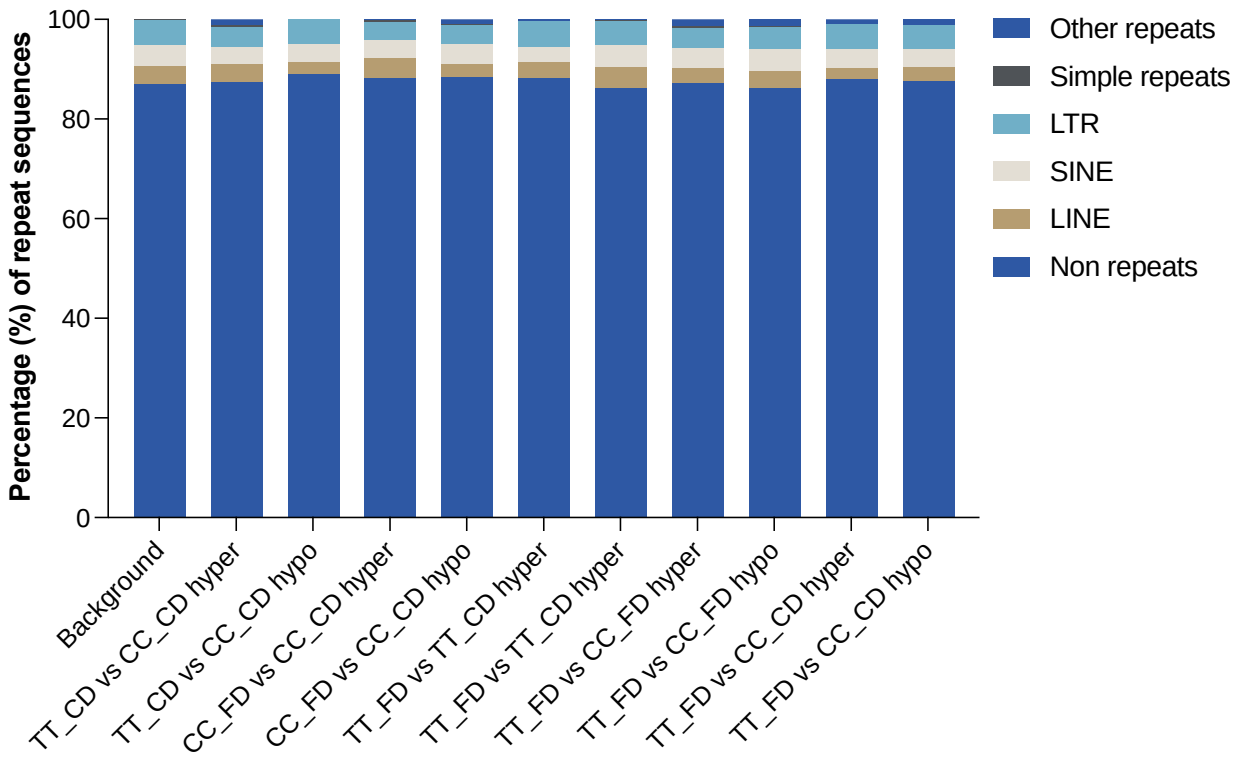


Figure 17. Proportion of repeat sequences from Illumina Mouse Methylation Array.

SINE, LINE and LTR sequences appeared to have nearly equal distribution across all groups but were not broadly covered by the array. LTR, long terminal repeat; LINE, long interspersed nuclear element**;** short interspersed nuclear element; hyper, hypermethylated; hypo, hypomethylated.

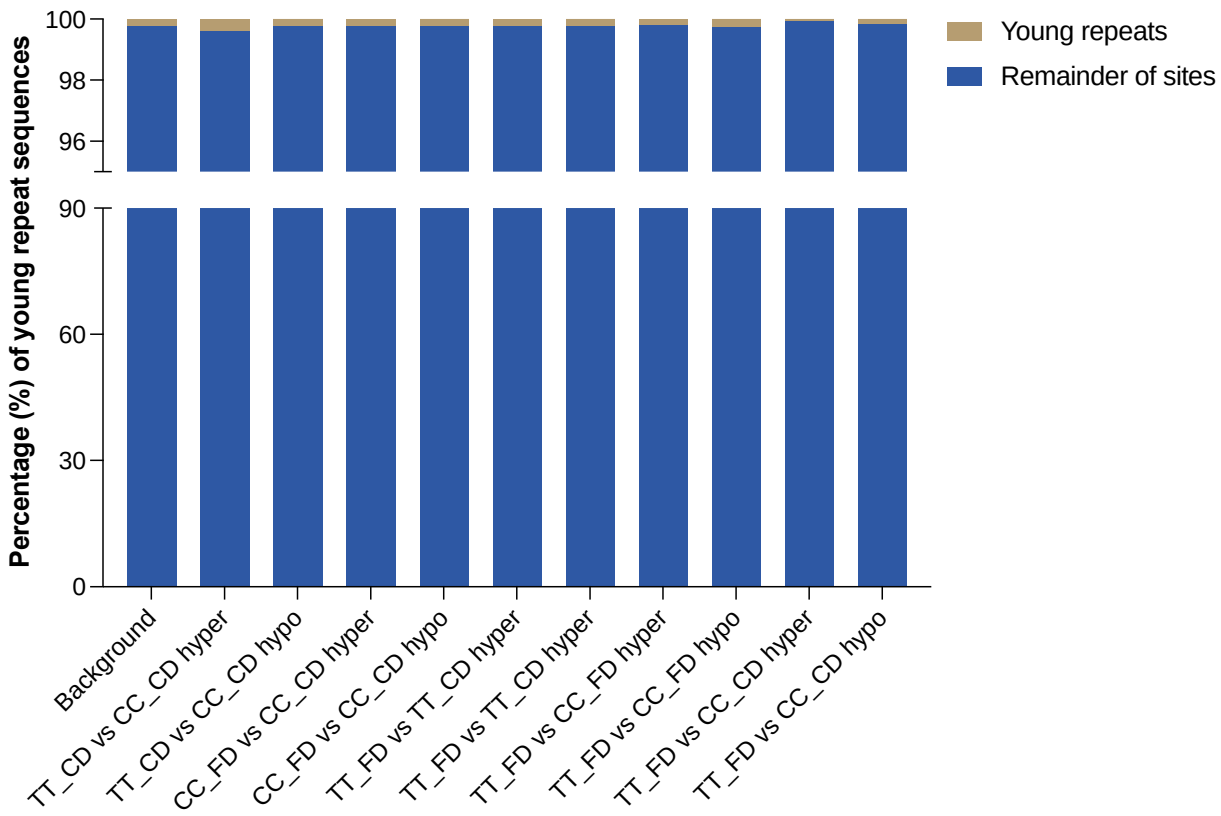


Figure 18. Proportion of evolutionary young retrotransposable elements from Illumina Mouse Methylation Array.

Young retrotransposons, specifically L1MdA_I, L1MdTf_I, L1MdTf_II, L1MdA_II, L1MdGf_I, L1MdTf_III, L1MdGf_II and L1MdA_III subfamilies were interrogated. Nearly none were covered by the array in all the studied groups. Hyper, hypermethylated; hypo, hypomethylated.

3.4.7 Regional analysis of DMCs from Illumina Array data

The results observed thus far have been located in single CpGs, for this reason we sought to determine if any larger-scale effects could be found. To achieve this, neighboring DMCs located within 100bp up to 10,000 bp were merged. Only a small number of DMCs merged into differentially methylated regions (DMRs). In the case of TT CD vs CC CD, the comparison with the largest effect, when merged at 100bp only 42 DMCs out of 2361 DMCs merged into regions (1.7%), and at the maximum distance, 10,000bp, 132 DMCs merged into regions (5.5%), the rest remained as single sites (Fig. 19). This means that only very few regional changes could be observed, possibly due to the location of this array's probes.

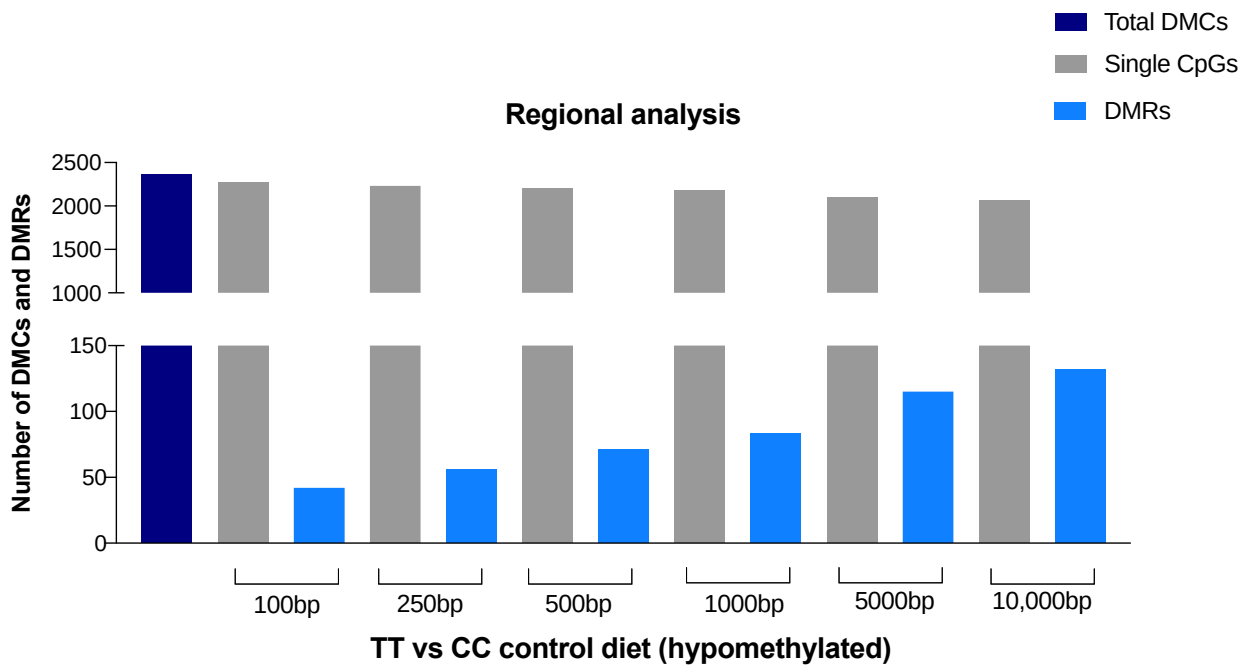


Figure 19. Illumina Array regional analysis of TT CD vs CC CD hypomethylation effect.

To investigate larger-scale regional changes in DNA methylation, neighboring differentially methylated CpGs (DMCs) were merged into regions. Different measurements were tested. The merged DMCs were located within 100 bp, 250bp, 500bp, 1000bp, 5000bp and 10,000 bp of each other. DMCs devoid of any neighboring DMCs were identified as single CpGs. DMCs that merged were identified as differentially methylated regions (DMRs). Here TT CD vs CC CD hypermethylation effect is shown as it had the largest effect out of all the different comparisons.

3.4.8 Gene ontology enrichment analysis of DMCs from Illumina Array data and reduced visualization.

Gene ontology (GO) term pathway enrichment was performed on the three most affected comparisons: *Mthfr* 677TT CD vs *Mthfr* 677CC CD, followed by *Mthfr* 677CC FD vs *Mthfr* 677CC CD and lastly *Mthfr* 677TT FD vs *Mthfr* 677CC CD. Interestingly, three GO terms were consistently shared among all these comparisons: cell migration, cell adhesion and positive regulation of transcription by RNA polymerase II (Fig. 20). As well, various neurological terms were observed, such as nervous system development in TT FD vs CC CD hypermethylated group, memory, and positive regulation of synapse assembly in TT CD vs CC CD hypermethylated group, and negative regulation of neuron apoptotic process in CC FD vs CC CD hypomethylated group (Fig. 20). This supports the idea that MTHFR activity and folic acid status are both important for normal neurological development and functioning.

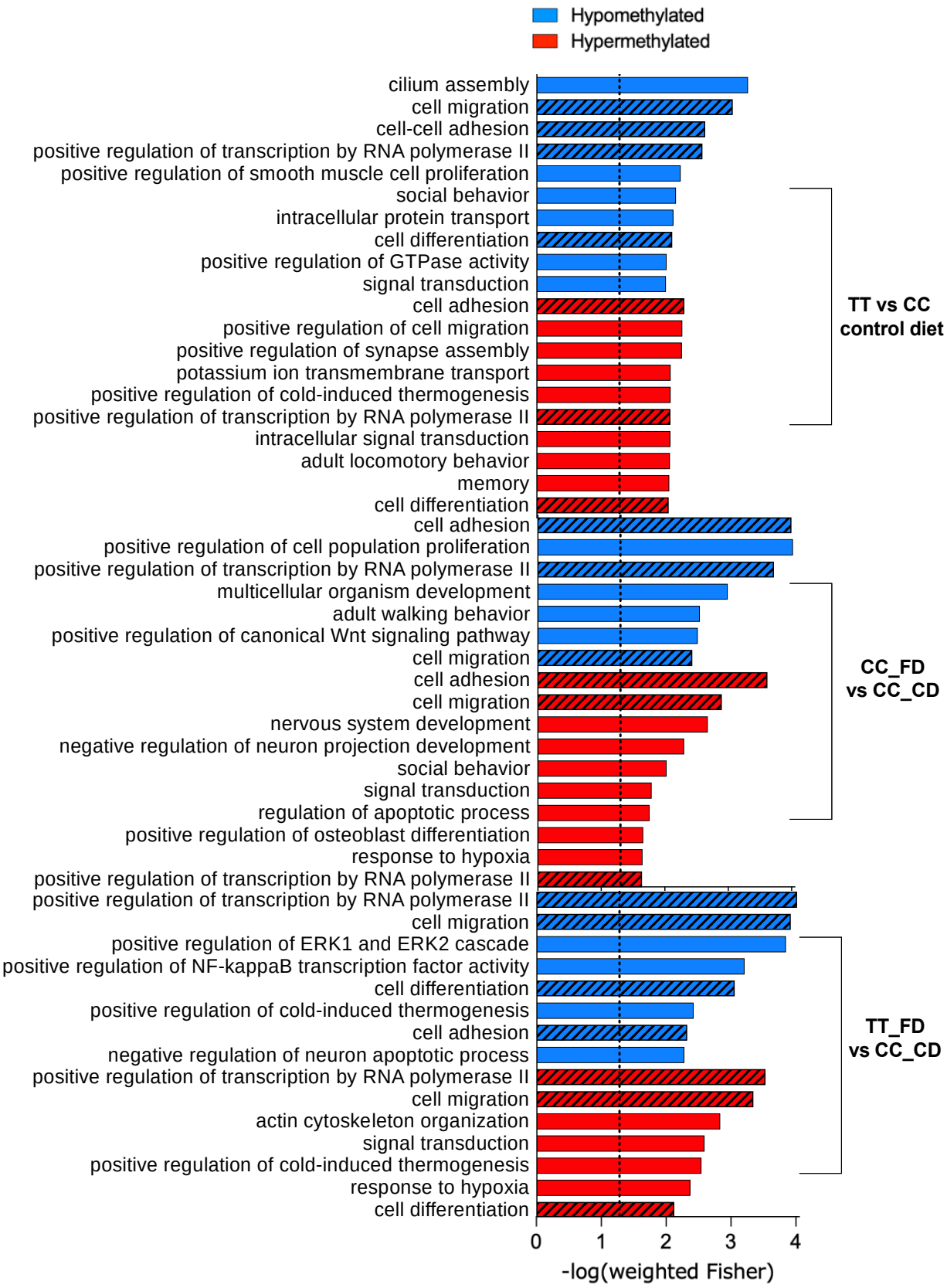

Hypomethylated
Hypermethylated
cilium assembly
cell migration
cell-cell adhesion
positive regulation of transcription by RNA polymerase II
positive regulation of smooth muscle cell proliferation
social behavior
intracellular protein transport
cell differentiation
positive regulation of GTPase activity
signal transduction
cell adhesion
positive regulation of cell migration
positive regulation of synapse assembly
potassium ion transmembrane transport
positive regulation of cold-induced thermogenesis
positive regulation of transcription by RNA polymerase II
intracellular signal transduction
adult locomotory behavior
memory
cell differentiation
cell adhesion
positive regulation of cell population proliferation
positive regulation of transcription by RNA polymerase II
multicellular organism development
adult walking behavior
positive regulation of canonical Wnt signaling pathway
cell migration
cell adhesion
cell migration
nervous system development
negative regulation of neuron projection development
social behavior
signal transduction
regulation of apoptotic process
positive regulation of osteoblast differentiation
response to hypoxia
positive regulation of transcription by RNA polymerase II
positive regulation of transcription by RNA polymerase II
cell migration
positive regulation of ERK1 and ERK2 cascade
positive regulation of NF-kappaB transcription factor activity
cell differentiation
positive regulation of cold-induced thermogenesis
cell adhesion
negative regulation of neuron apoptotic process
positive regulation of transcription by RNA polymerase II
cell migration
actin cytoskeleton organization
signal transduction
positive regulation of cold-induced thermogenesis
response to hypoxia
cell differentiation
TT vs CC
control diet
CC_FD
vs CC_CD
TT_FD
vs CC_CD
0
1
2
3
4
-log(weighted Fisher)

Figure 20. Gene ontology enrichment analysis (TopGO & rrvgo).

GO enrichment analysis of differentially methylated sites (DMCs) using post-hoc analysis of significant probes. All significant terms were reduced using high stringency (threshold = 0.9) using the R package Reduce + Visualize GO (rrvgo). The dotted line indicates the $p \leq 0.05$ threshold for significance for FDR. Hatched bars indicate common enriched pathways shared among different groups.

3.4.9 Corresponding genes/closest genes of DMCs from Illumina Array data.

The list of probes from the Illumina Array was filtered to determine which genes or genes closest to intergenic regions were affected. The genes found for almost all comparisons were annotated as “closest genes” to affected intergenic areas in the genome. The majority of the methylation beta values from the Illumina Array had a beta methylation value difference of ≤ 0.10. For the sake of biological significance, only genes or closest genes to intergenic areas having ≥5% change in methylation (0.05 beta difference) were more closely analyzed. Every group had a larger effect of hypomethylation, with the exception of CC FD vs CC CD. The total number of genes/closest genes in the different groups studied were as follows: TT FD vs CC CD, 16 hypermethylated and 34 hypomethylated; CC FD vs CC CD, 20 hypermethylated and 12 hypomethylated; TT FD vs TT CD, 12 hypermethylated and 22 hypomethylated; TT FD vs CC FD, 2 hypermethylated and 11 hypomethylated; TT FD vs CC CD 4 hypermethylated and 33 hypomethylated (Table 4).

The group with the largest effect (TT CD vs CC CD), revealed certain crucial genes which appeared to be affected. For example, the *Aga* gene, *Rbmy* and numerous olfactory receptor genes were found annotated closest to intergenic areas, and some in promoter-TSS regions. Various other genes or closest genes were affected in more than one group, in some cases following a different direction (hyper or hypomethylated), like *H2al2c*, *Klra1*, *Ctag2*, *Cox7c*. Other commonalities were found, the full list can be found in Table 4. Taken together, these results suggest that MTHFR activity and folic acid status are important not only for normal neurological development, but they are also implicated in numerous other processes.

Gene Name	Gene Description	Annotation	Mean meth diff	Distance to TSS
	TT CD vs CC CD	**Hypermethylation**		
4932411N23Rik	RIKEN cDNA 4932411N23 gene	Intergenic	0.104	437035
Gm20826	predicted gene	Intergenic	0.099	763479
Nap1l3	nucleosome assembly protein 1-like 3	Intergenic	0.079	801624
Ctag2	cancer/testis antigen 2	Intergenic	0.079	667209
4933402E13Rik	RIKEN cDNA 4933402E13 gene	Intergenic	0.077	-245451
Ctag2	cancer/testis antigen 2	Intergenic	0.071	141514
Olfr785	olfactory receptor 785	Intergenic	0.070	-4177
Gm14743	predicted gene 14743	Intergenic	0.068	-53963
Olfr972	olfactory receptor 972	Intergenic	0.066	-6981
Cox7c	cytochrome c oxidase subunit VIIc	Intergenic	0.064	-1187729
H2al2c	H2A histone family member L2C	Intergenic	0.060	-78404
H2al2c	H2A histone family member L2C	Intergenic	0.059	-127108
Slc38a1	solute carrier family 38	Intergenic	0.054	-10378
Pou3f4	POU domain	Intergenic	0.054	48187
Mir1906-1	microRNA 1906-1	Intergenic	0.053	29585
Usp9y	ubiquitin specific peptidase 9	Intergenic	0.051	-310555
	TT CD vs CC CD	**Hypomethylation**		
Mir6342	microRNA 6342	Intergenic	-0.163	-60418
1700017L05Rik	RIKEN cDNA 1700017L05 gene	Intergenic	-0.128	589510
Actr3	ARP3 actin-related protein 3	Intergenic	-0.119	443791
Gm6455	predicted gene 6455	Intergenic	-0.115	-64965
Pla2g4a	phospholipase A2	Intergenic	-0.113	1675327
1700019M22Rik	NFX-like protein	Intergenic	-0.110	-314021
Aga	aspartylglucosaminidase	Intergenic	-0.105	-784117
Gm6455	predicted gene 6455	Intergenic	-0.105	-196929
1700017L05Rik	RIKEN cDNA 1700017L05 gene	Intergenic	-0.102	842774
Actr3	ARP3 actin-related protein 3	Intergenic	-0.098	238872
Gm6455	predicted gene 6455	Intergenic	-0.097	-188659
1700019M22Rik	NFX-like protein	Intergenic	-0.091	-947892
Vmn1r50	vomeronasal 1 receptor 50	Intergenic	-0.088	-6051
Foxg1	forkhead box G1	Intergenic	-0.085	-476751
Gm4736	predicted gene 4736	Intergenic	-0.079	-24911
Nsun3	NOL1/NOP2/Sun domain family member 3	Intergenic	-0.077	414990
1700019M22Rik	NFX-like protein	Intergenic	-0.074	-245169
Nsun3	NOL1/NOP2/Sun domain family member 3	Intergenic	-0.072	403269
Aga	aspartylglucosaminidase	Intergenic	-0.069	-784431
1700019M22Rik	NFX-like protein	Intergenic	-0.068	-264900
Rbmy	RNA binding motif protein	Intergenic	-0.065	170915
Olfr1459	olfactory receptor 1459	Intergenic	-0.060	-3350
Cyb5b	cytochrome b5 type B	Intergenic	-0.057	47685
Arl4a	ADP-ribosylation factor-like 4A	Intergenic	-0.056	593206
Slc35e3	solute carrier family 35	promoter-TSS	-0.056	-577
Klra3	killer cell lectin-like receptor	promoter-TSS	-0.055	78
A330008L17Rik	RIKEN cDNA A330008L17 gene	Intergenic	-0.055	1309769
2010300C02Rik	RIKEN cDNA 2010300C02 gene	Intergenic	-0.055	-6925
1700019M22Rik	NFX-like protein	Intergenic	-0.051	-180376
Jun	jun proto-oncogene	Intergenic	-0.051	-56235
Fam110c	family with sequence similarity 110	Intergenic	-0.051	17095
Gm20815	predicted gene	Intergenic	-0.051	-215379
H2afb1	H2A histone family	Intergenic	-0.050	53254
Sel1l	sel-1 suppressor of lin-12-like (C. elegans)	Intergenic	-0.050	-159919
	CC FD vs CC CD	**Hypermethylation**		
Ctag2	cancer/testis antigen 2	Intergenic	0.086	141514
Olfr938	olfactory receptor 938	promoter-TSS	0.078	-351
Olfr785	olfactory receptor 785	Intergenic	0.073	-4177
Diaph2	diaphanous related formin 2	Intergenic	0.070	1739481
4933402E13Rik	RIKEN cDNA 4933402E13 gene	Intergenic	0.070	-281890
Olfr784	olfactory receptor 784	Intergenic	0.069	-3251
Zfy2	zinc finger protein 2	Intergenic	0.066	203669
Usp9y	ubiquitin specific peptidase 9	Intergenic	0.061	-301429
Pkia	protein kinase inhibitor	Intergenic	0.060	-322180

Olfr101	olfactory receptor 101	Intergenic	0.060	-1399
Ppp1r2-ps9	protein phosphatase 1	Intergenic	0.059	283114
Slitrk4	SLIT and NTRK-like family	Intergenic	0.059	-276510
Gm21708	predicted gene	Intergenic	0.058	1576257
Inhba	inhibin beta-A	Intergenic	0.057	146577
Mrps30	mitochondrial ribosomal protein S30	Intergenic	0.056	117782
Prh1	proline rich protein HaeIII subfamily 1	Intergenic	0.056	11462
Lrrc7	leucine rich repeat containing 7	Intergenic	0.053	-227559
Arl4a	ADP-ribosylation factor-like 4A	Intergenic	0.053	145346
Gm20815	predicted gene	Intergenic	0.052	338837
Rnh1	ribonuclease/angiogenin inhibitor 1	Intergenic	0.052	-5642
	CC FD vs CC CD	**Hypomethylation**		
Gm4736	predicted gene 4736	Intergenic	-0.090	-24911
Nsun3	NOL1/NOP2/Sun domain family member 3	Intergenic	-0.079	771331
Slitrk6	SLIT and NTRK-like family	Intergenic	-0.076	151453
A630073D07Rik	RIKEN cDNA A630073D07 gene	promoter-TSS	-0.063	-750
1700019M22Rik	NFX-like protein	Intergenic	-0.056	-264900
4930572O13Rik	RIKEN cDNA 4930572O13 gene	Intergenic	-0.055	15601
Rfesd	Rieske (Fe-S) domain containing	Intergenic	-0.054	-1452
Zfp365	zinc finger protein 365	Intergenic	-0.053	-7219
Klra1	killer cell lectin-like receptor	Intergenic	-0.052	-103101
1700101O22Rik	RIKEN cDNA 1700101O22 gene	Intergenic	-0.052	412448
Luzp2	leucine zipper protein 2	Intergenic	-0.051	-166597
Trappc11	trafficking protein particle complex 11	Intergenic	-0.050	74068
	TT FD vs TT CD	**Hypermethylation**		
Kcnh8	potassium voltage-gated channel	Intergenic	0.061	-143255
Tagap	T cell activation Rho GTPase activating protein	Intergenic	0.060	-42629
Myf6	myogenic factor 6	TTS	0.058	2791
Cyb5b	cytochrome b5 type B	Intergenic	0.057	47686
Brinp3	morphogenetic protein/retinoic acid inducible neural spe	Intergenic	0.055	1546866
Fam110c	family with sequence similarity 110	Intergenic	0.055	17096
Gm9871	predicted gene 9871	Intergenic	0.054	163574
Gramd3	GRAM domain containing 3	Intergenic	0.053	-459786
Gm6602	predicted gene 6602	Intergenic	0.052	-812755
Pla2g4a	phospholipase A2	Intergenic	0.052	589629
Dsel	dermatan sulfate epimerase-like	Intergenic	0.051	-286264
Pcdhb19	protocadherin beta 19	promoter-TSS	0.050	94
Gm29683	predicted gene	Intergenic	0.050	228646
	TT FD vs TT CD	**Hypomethylation**		
Snora28	small nucleolar RNA	promoter-TSS	-0.143	-753
Tgif2lx2	TGFB-induced factor homeobox 2-like	Intergenic	-0.106	-286687
Pabpc5	poly(A) binding protein	Intergenic	-0.104	-229184
AA545190	EST AA545190	Intergenic	-0.098	151694
4932411N23Rik	RIKEN cDNA 4932411N23 gene	Intergenic	-0.088	437034
Obp1b	odorant binding protein IB	Intergenic	-0.086	-1512
H2al2c	H2A histone family member L2C	Intergenic	-0.075	-57476
Olfr135	olfactory receptor 135	promoter-TSS	-0.074	-449
Gm44	predicted gene 44	Intergenic	-0.073	-122505
Gm16497	predicted gene 16497	Intergenic	-0.066	304872
1700101O22Rik	RIKEN cDNA 1700101O22 gene	Intergenic	-0.062	681039
Cox7c	cytochrome c oxidase subunit VIIc	Intergenic	-0.062	-1187730
C130030K03Rik	proteasome (prosome	Intergenic	-0.060	558687
1700101O22Rik	RIKEN cDNA 1700101O22 gene	Intergenic	-0.060	548441
Sis	sucrase isomaltase (alpha-glucosidase)	Intergenic	-0.058	657942
Scgb2b1	secretoglobin	Intergenic	-0.055	-7426
Gm3376	predicted gene 3376	Intergenic	-0.053	247771
Olfr1154	olfactory receptor 1154	promoter-TSS	-0.053	-546
Kcnv1	potassium channel	Intergenic	-0.052	-375274
1700109I08Rik	RIKEN cDNA 1700109I08 gene	Intergenic	-0.052	83517
Gm6377	predicted gene 6377	Intergenic	-0.051	255456
Gm4981	predicted gene 4981	promoter-TSS	-0.050	-54
	TT FD vs CC FD	**Hypermethylation**		
Cdh12	cadherin 12	Intergenic	0.061	617753

Klra1	killer cell lectin-like receptor	Intergenic	0.055	-103101
TT FD vs CC FD		**Hypomethylation**		
Vmn2r68	vomeronasal 2	Intergenic	-0.135	-56437
Mageb5	melanoma antigen	Intergenic	-0.085	60400
H2al2c	H2A histone family member L2C	Intergenic	-0.084	-175789
Ghitm	growth hormone inducible transmembrane protein	Intergenic	-0.080	-724582
H2al2c	H2A histone family member L2C	Intergenic	-0.077	-207062
Nutm2	NUT family member 2	Intergenic	-0.065	-6373
Esp31	exocrine gland secreted peptide 31	Intergenic	-0.065	7738
1700101O22Rik	RIKEN cDNA 1700101O22 gene	Intergenic	-0.064	548442
Olfr785	olfactory receptor 785	Intergenic	-0.062	-4177
Gm3376	predicted gene 3376	Intergenic	-0.061	225209
Mrps30	mitochondrial ribosomal protein S30	Intergenic	-0.054	117782
TT FD vs CC CD		**Hypermethylation**		
Zfy2	zinc finger protein 2	Intergenic	0.093	254915
Nap1l3	nucleosome assembly protein 1-like 3	Intergenic	0.088	981609
Hnf4g	hepatocyte nuclear factor 4	Intergenic	0.072	-199533
Zfy2	zinc finger protein 2	Intergenic	0.051	299401
TT FD vs CC FD		**Hypomethylation**		
Nsun3	NOL1/NOP2/Sun domain family member 3	Intergenic	-0.148	688076
Gm20268	predicted gene	Intergenic	-0.131	1361109
Cdh20	cadherin 20	Intergenic	-0.116	-1596679
1700101O22Rik	RIKEN cDNA 1700101O22 gene	Intergenic	-0.109	501630
Rbmy	RNA binding motif protein	Intergenic	-0.098	225054
Cdh20	cadherin 20	Intergenic	-0.096	-1724116
Cyp2c39	cytochrome P450	promoter-TSS	-0.096	-35
1700101O22Rik	RIKEN cDNA 1700101O22 gene	Intergenic	-0.094	548442
A330008L17Rik	RIKEN cDNA A330008L17 gene	Intergenic	-0.088	1634965
Gm20268	predicted gene	Intergenic	-0.088	1435683
1700101O22Rik	RIKEN cDNA 1700101O22 gene	Intergenic	-0.084	412448
Slc22a26	solute carrier family 22 (organic cation transporter)	Intergenic	-0.081	-2285
Nsun3	NOL1/NOP2/Sun domain family member 3	Intergenic	-0.079	771331
1700101O22Rik	RIKEN cDNA 1700101O22 gene	Intergenic	-0.078	-7371
Nsun3	NOL1/NOP2/Sun domain family member 3	Intergenic	-0.077	467537
Gm4736	predicted gene 4736	Intergenic	-0.077	-24911
Hs3st1	heparan sulfate (glucosamine) 3-O-sulfotransferase 1	Intergenic	-0.075	-926019
Gm20268	predicted gene	Intergenic	-0.074	1303354
Rbmy	RNA binding motif protein	Intergenic	-0.073	225209
Cdh6	cadherin 6	Intergenic	-0.070	-1105483
Cdh11	cadherin 11	Intergenic	-0.067	1678244
Gm4736	predicted gene 4736	Intergenic	-0.067	-99500
1700101O22Rik	RIKEN cDNA 1700101O22 gene	Intergenic	-0.066	681040
Rbmy	RNA binding motif protein	Intergenic	-0.064	247770
Esp31	exocrine gland secreted peptide 31	Intergenic	-0.062	7738
Tas2r114	taste receptor	Intergenic	-0.059	-114213
Edem1	ER degradation enhancer	Intergenic	-0.059	873270
Obp1a	odorant binding protein IA	Intergenic	-0.058	69997
A430089I19Rik	RIKEN cDNA A430089I19 gene	Intergenic	-0.056	87799
A430089I19Rik	RIKEN cDNA A430089I19 gene	Intergenic	-0.056	87799
Slc38a5	solute carrier family 38	promoter-TSS	-0.055	-371
Rab14	RAB14	Intergenic	-0.055	-1128
Cdh6	cadherin 6	Intergenic	-0.053	-1114807

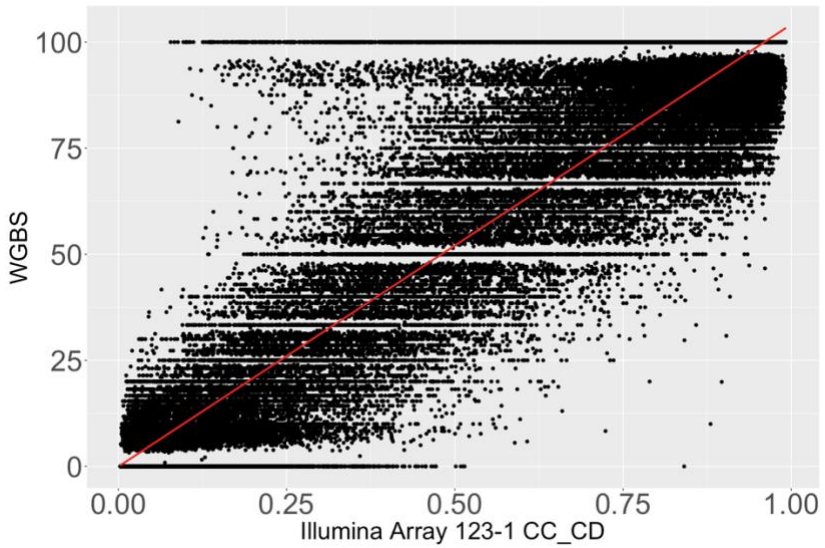
Table 4. Complete list of genes/closest genes of Illumina Array DMCs from each group with ≥5% change in methylation.

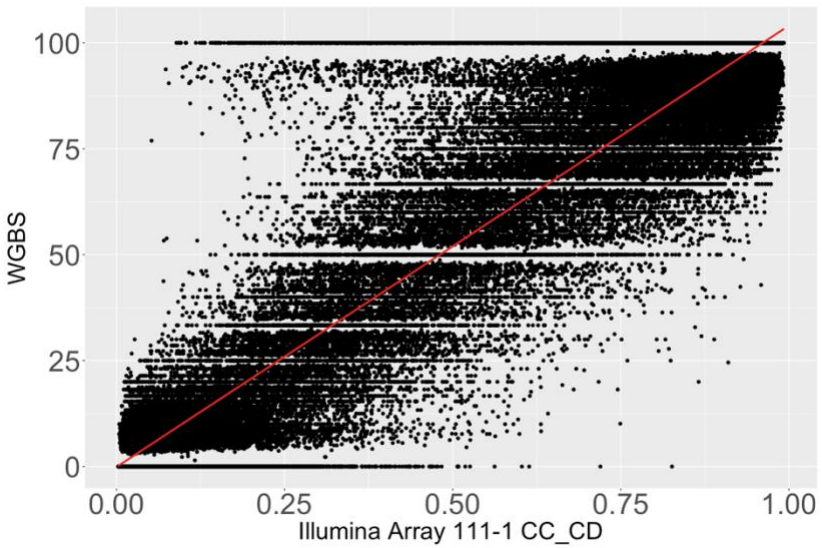
Mean meth diff, mean difference in methylation form beta values of Illumina Mouse Methylation Array; TSS, transcription start site.

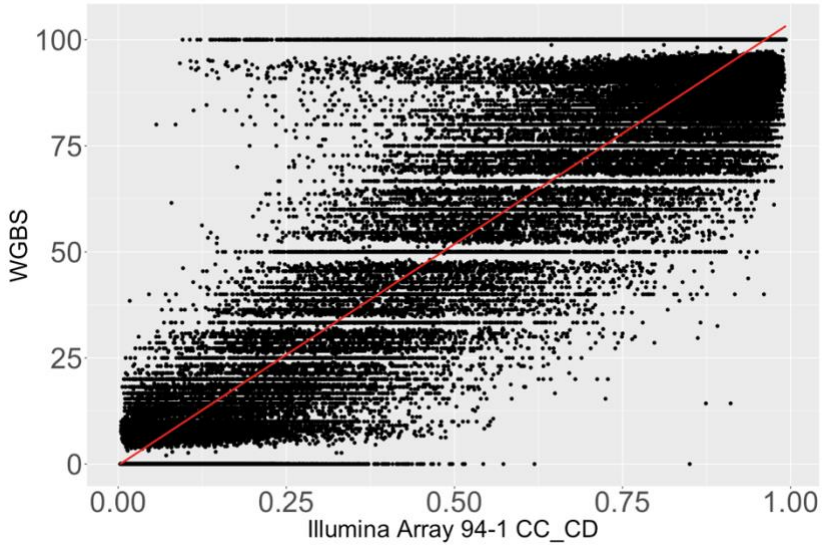
3.5 Results obtained with WGBS

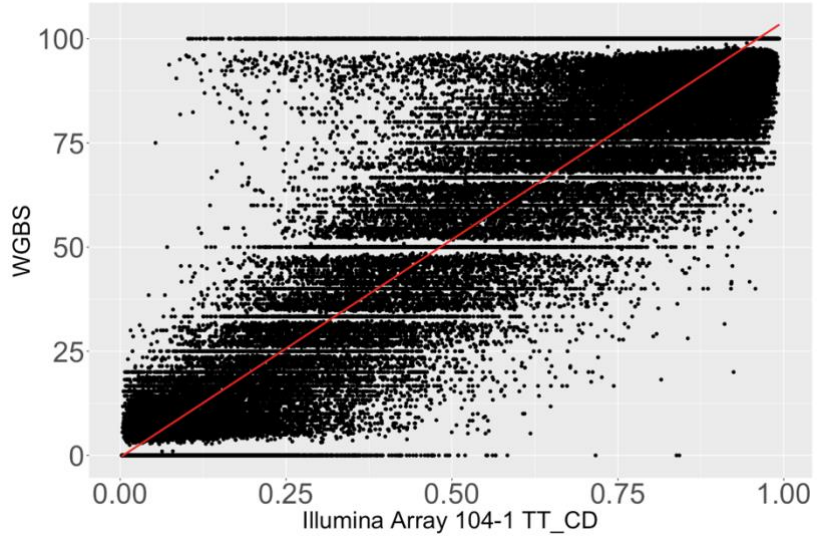
3.5.1 Comparing control samples analyzed by both Illumina Array and WGBS

As described previously, FASD samples were analyzed by WGBS along with control samples which included 6 previously analyzed by Illumina Mouse Methylation BeadChip (3 CC and 3 TT). Pearson correlation analyses between Illumina Array data and WGBS at 10x coverage, showed a positive correlation with an $R^2 \geq 0.94$ and p value of < 2.2e-16 for all six samples (Fig 21). The values obtained by both technologies were better correlated for high and low values of methylation. In the case of intermediate values, however, the results showed higher variability. To take a closer look, genes with high, intermediate, and low methylation values were chosen, 2 per range (Fig 22). Again, highly methylated genes, in this case *H19* and *Fgfr12* corroborated well. The same was true for lowly methylated genes, in this case *Foxg1* and *Snrpn*. For intermediate methylation values, the results varied slightly, with the Illumina Array *Elmod1* showing beta values between 0.53 and 0.54, whereas as with WGBS it displayed between 43% and 53% methylation. Likewise, with the Illumina Array *Pcdhb19* showed beta values between 0.56 and 0.57, whereas WGBS displayed methylation values between 52 and 58%.


WGBS
100
75
50
25
0
0.00
0.25
0.50
0.75
1.00
Illumina Array 123-1 CC_CD


WGBS
100
75
50
25
0
0.00
0.25
0.50
0.75
1.00
Illumina Array 111-1 CC_CD


WGBS
100
75
50
25
0
0.00
0.25
0.50
0.75
1.00
Illumina Array 94-1 CC_CD


WGBS
100
75
50
25
0
0.00
0.25
0.50
0.75
1.00
Illumina Array 104-1 TT_CD

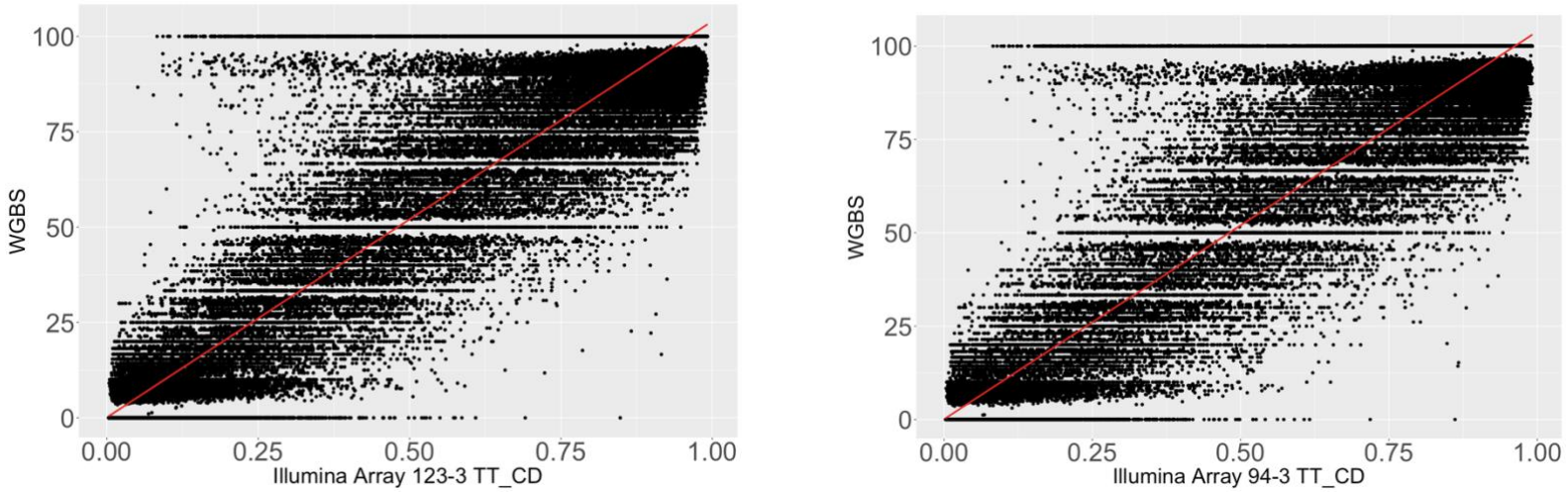


Figure 21. Pearson correlation scatterplots of common CD samples analyzed by Illumina Mouse Methylation Array and WGBS.

In all samples (3 CC CD and 3 TT CD), the Pearson correlation R^2 value was ≥0.94 and the p value was < 2.2e-16 for all samples indicating a strong positive correlation between the datasets. The correlation was stronger for low and high methylation values, whereas intermediate methylation values displayed more variability.

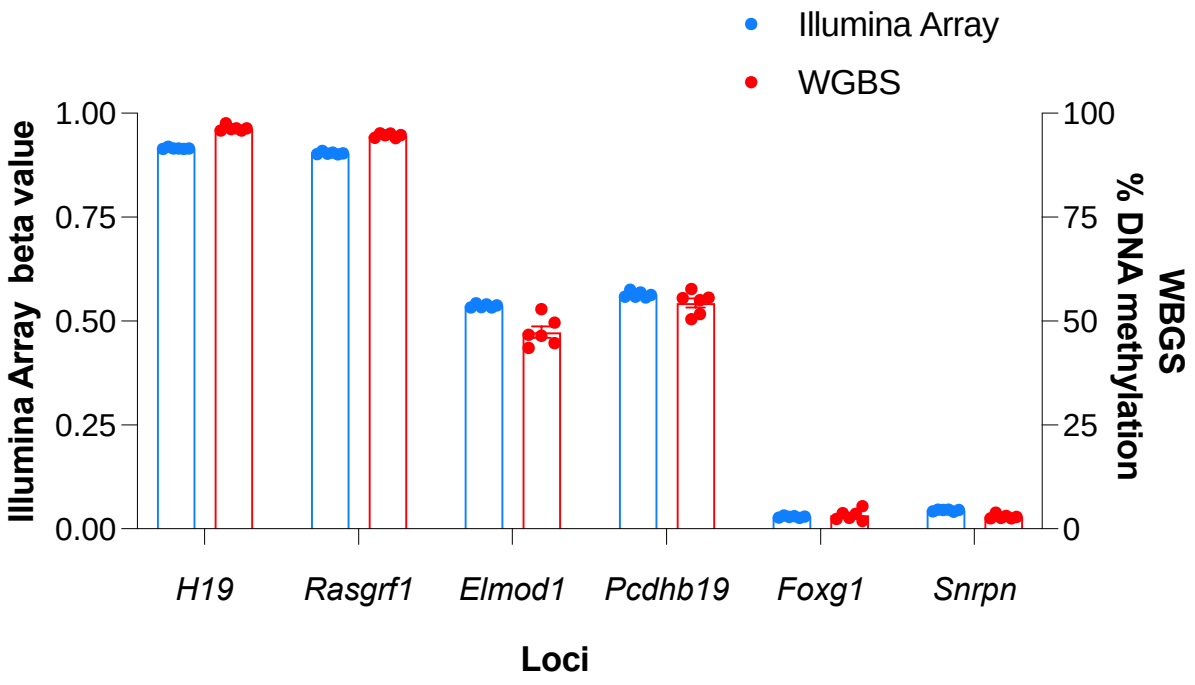


Figure 22. Methylation values for control samples analyzed by both technologies.

In total 6 control samples (3 CC CD and 3 TT CD) were analyzed by both the Illumina Mouse Methylation Array and whole-genome bisulfite sequencing. To showcase the methylation values, genes with high, medium, and low mean methylation values were graphed. Individual dots indicate mean methylation values per sample (6 total/gene). Bars shown as mean +/- SEM.

3.5.2 Effects of *Mthfr* genotype and FASD and CD samples on imprinted genes of WGBS data at 10x coverage.

As seen with the Illumina Array data, paternally methylated genes showed the expected high levels of methylation, however, *Zdbf2* once again displayed areas with lower methylation values (Fig 23A). Maternally methylated genes all showed the expected low levels of methylation (Fig 23B), confirming that these were unaffected by a moderate dose of folic acid supplementation. Overall, WGBS showed a larger coverage of these genes.

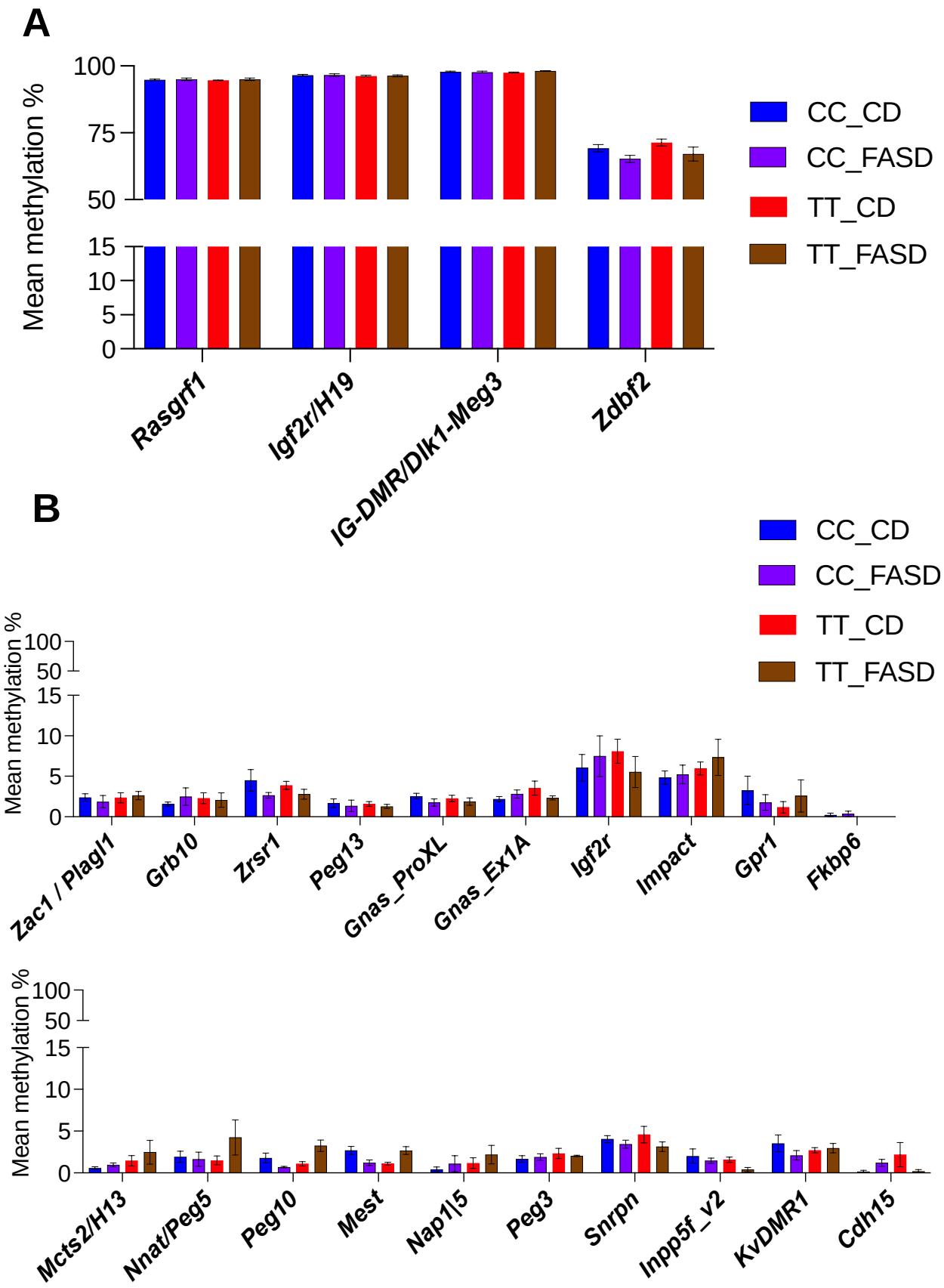

A
Mean methylation %
100
75
50
15
10
5
0
CC_CD
CC_FASD
TT_CD
TT_FASD
Rasgrf1
Igf2r/H19
IG-DMR/Dlk1-Meg3
Zdbf2
B
CC_CD
CC_FASD
TT_CD
TT_FASD
Mean methylation %
100
50
15
10
5
0
Zac1 / Plagl1
Grb10
Zrsr1
Peg13
Gnas_ProXL
Gnas_Ex1A
Igf2r
Impact
Gpr1
Fkbp6
Mean methylation %
100
50
15
10
5
0
Mcts2/H13
Nnat/Peg5
Peg10
Mest
Nap1l5
Peg3
Snrpn
Inpp5f_v2
KvDMR1
Cdh15

Figure 23. ICR methylation values from WGBS.

The results are similar to those obtained with the Illumina Array. **(A)** Paternally methylated imprinting control regions (ICRs), *Rasgfr1, Igf2/H19 and IG-DMR/Dlk1-Meg3* showed expected high levels of methylation, *Zdbf2* revealed areas with lower values of methylation, **(B)** Maternally methylated ICRs displayed expected low levels of methylation and remained unaffected by different genotypes or folic acid diets. Bars shown as mean +/- SEM.

3.5.3 Analysis of differentially methylated tiles (DMTs) of FASD and CD samples

Folic acid supplemented samples along with control samples, were analyzed with WGBS. This was conducted using methylKit where differentially methylated tiles were detected using a minimal of 10x coverage and 5% change. First, we examined the genotype effect, TT CD vs CC CD and TT FASD vs CC FASD (Fig 24). Once again, the results obtained showed the largest effect was caused by the TT genotype on a control diet, with 141 hypermethylated and 219 hypomethylated DMTs. When comparing both genotypes on the FASD diet, they revealed 121 hypermethylated DMTs and 139 hypomethylated DMTs. Next, we examined the effects of the FASD by comparing CC FASD vs CC CD and TT FASD vs TT FASD. The first group showed 137 hypermethylated DMTs and 108 hypomethylated DMTs, the second group displayed a larger number with 194 hypermethylated and 131 hypomethylated DMTs. Lastly, when comparing the extremes (TT FASD vs CC CD), we observed the least number of changes as this group showed 110 hypermethylated and 106 hypomethylated DMTs.

For all groups compared, the majority of methylation changes was in average between 5 and 10% (Fig 25). Additionally, the distribution of methylation differences suggests that a moderate dose of folic acid supplementation causes an increase of DMTs with ≥10% methylation difference on the TT genotype. To determine if this was true, a Fisher's exact test was performed between TT FASD vs TT CD and TT CD vs CC CD. The result shows that, when adding folic acid to the TT genotype there is no significant increase in hypermethylated DMTs that have a ≥10% methylation change ($p= 0.108$).

Interestingly, the proportion of DMTs, particularly hypomethylated tiles diminished in TT FASD vs CC CD, suggesting a partial correction from the effects caused by the TT

genotype on a control diet. To determine if any folic acid supplementation was able to correct any DMTs between these groups, the methylation value was compared for CC CD, TT CD and TT FASD tiles that overlapped with the hypomethylated DMTs encountered in the TT CD vs CC CD group. Of 141 hypermethylated DMTs, 49 of them overlapped with the TT FASD group. Of those 49 DMTs, 44 showed some level of correction. In the hypomethylated group, out of 219 hypomethylated DMTs, 84 of them overlapped with the TT FASD group. The results observed determined that 82 of these tiles gained methylation after supplementing the TT genotype with folic acid (Fig 26).

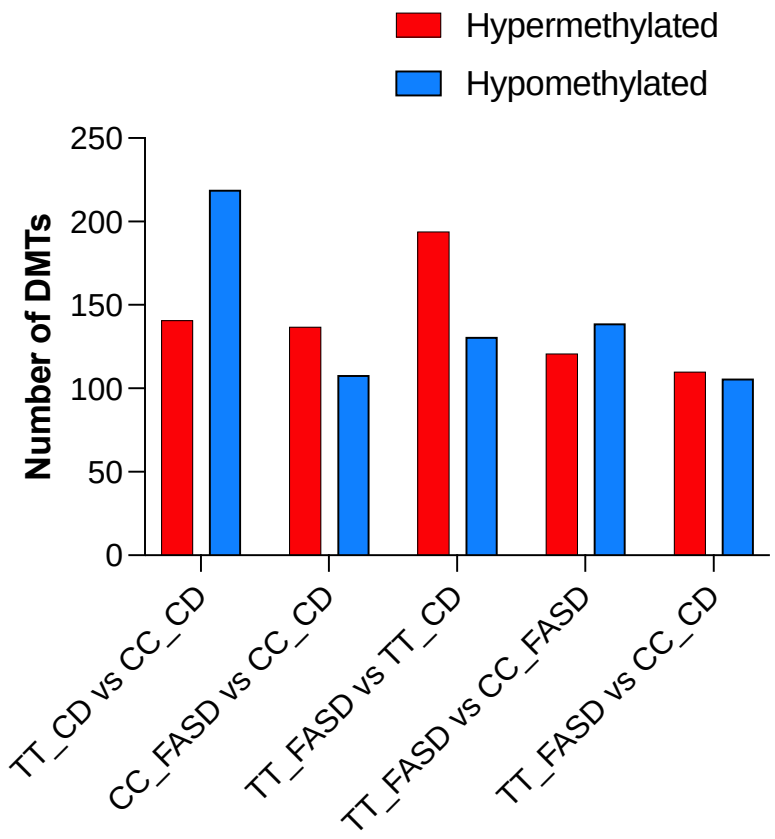


Figure 24. Summary of WGBS DMTs.

WGBS analysis was performed at 10x coverage and DMTs with at least 5% difference in methylation were counted. WGBS, whole-genome bisulfite sequencing; DMTs, differentially methylated tiles (100bp/tile)

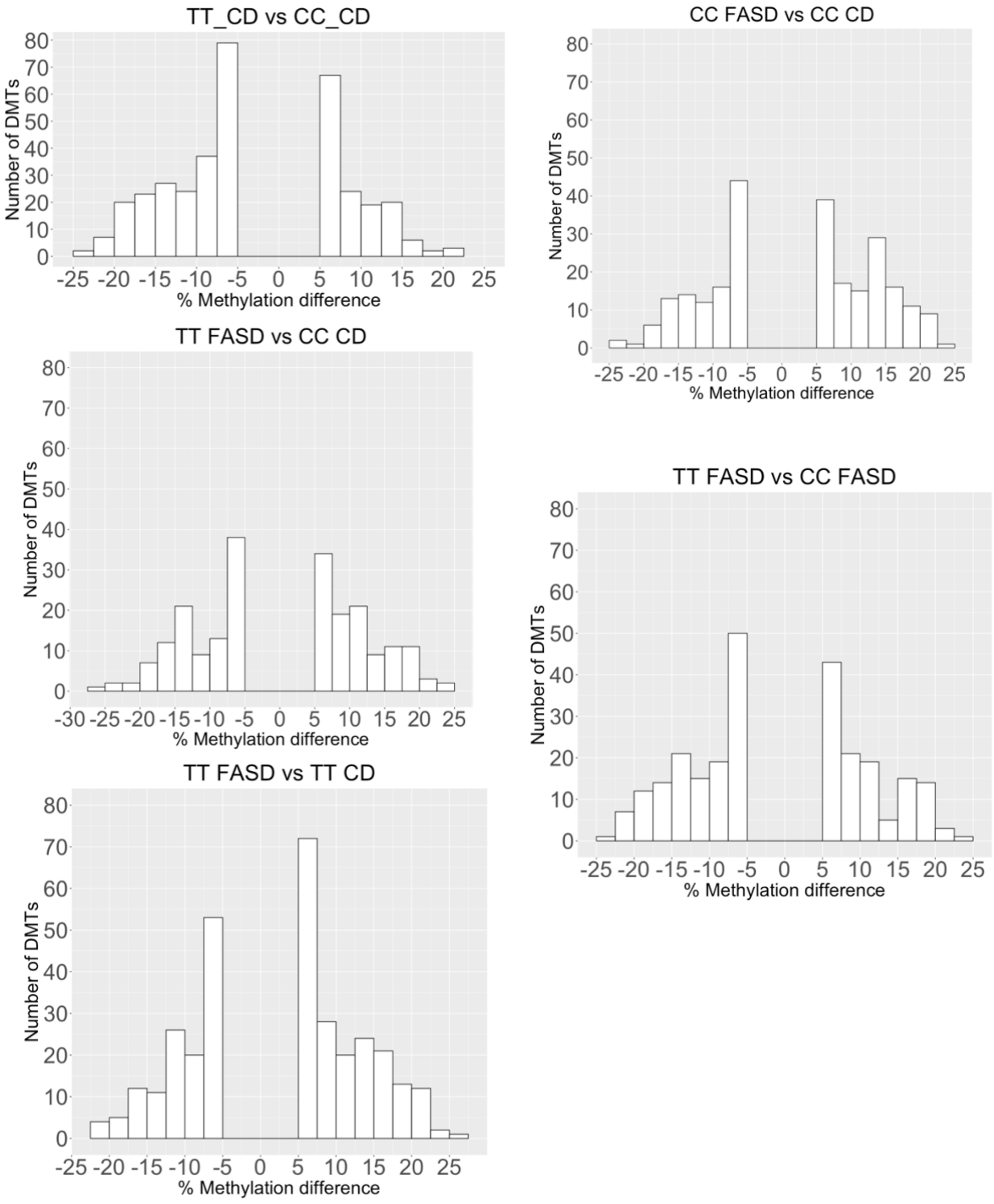


Figure 25. Percentage of methylation difference of WGBS DMTs.

Difference in methylation of DMTs at ≥5% change for all comparisons. WGBS, whole-genome bisulfite sequencing; DMTs, differentially methylated tiles (100bp).

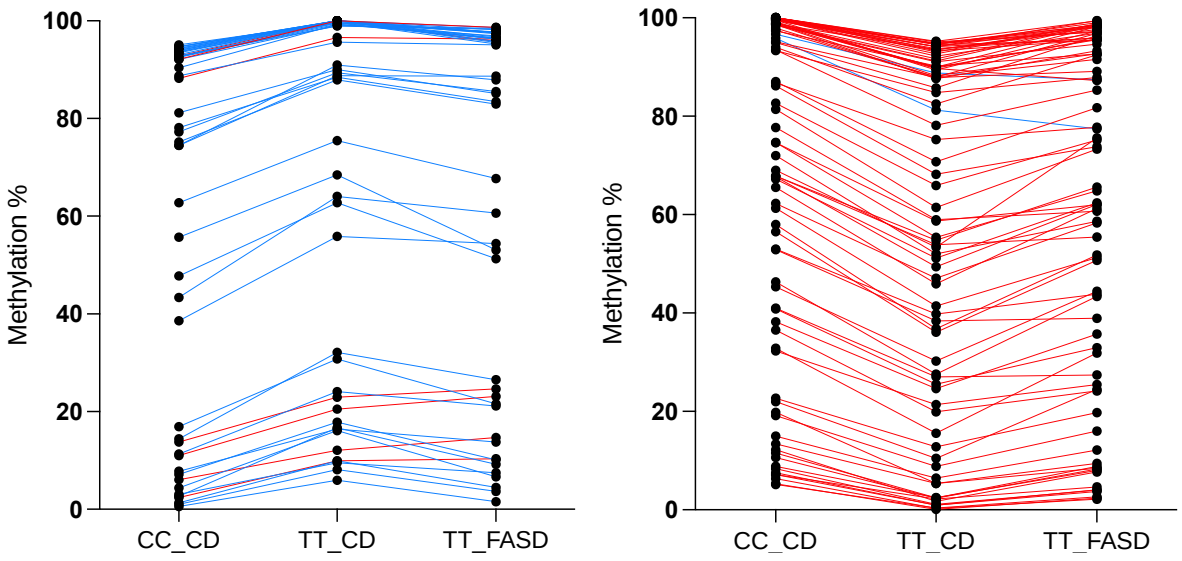


Figure 26. Correction of TT genotype with FASD

Values from CC CD, TT CD and TT FASD were intersected with DMTs from TT CD vs CC CD. Left hypermethylated DMTs; right hypomethylated DMTs. Each dot represents the average of all six samples located at each hyper- and hypomethylated DMT from the TT CD vs CC CD comparison (49 hyper- and 84 hypomethylated). On the left, blue lines represent tiles that decreased methylation after supplementing the TT genotype with folic acid (44 lines) and red lines indicate further increase in methylation (5 lines). On the right, red lines represent tiles and had increased methylation after supplementing the TT genotype with folic acid (82 lines) and blue lines indicate further decrease in methylation (2 lines).

3.5.4 Genomic distribution of FASD and CD samples

As mentioned previously, studies have shown that intergenic areas are the most affected by MTHFR deficiency and folic acid status. With WGBS, as expected, the most represented areas were mostly intergenic, followed by intronic zones in the background data. Chi squared with Yate's correction tests were performed to determine significant differences across all groups compared to the background. In the hypermethylated groups, exons were more affected in TT CD vs CC CD (p=<0.0001), CC FASD vs CC CD (p=<0.0001) and TT FASD vs TT CD (p=<0.0001). Likewise, exons in the hypomethylated group were more affected in TT CD vs CC CD (p=<0.0001) and CC FASD vs CC CD (p=0.0003) (Fig 27A). Promoter-TSS areas in the hypermethylated group were more affected TT CD vs CC CD (p=<0.0001) (Fig 27B).

3.5.5 Proportion of repeat sequences of FASD and CD samples

Analysis of repeat sequences by WGBS revealed a slightly larger proportion of LINE elements (14%) compared to SINE (9%) and LTR (10%) sequences in the background data (Fig 28). Like before, Chi squared with Yate's correction tests were performed to determine significant differences across all groups compared to the background. LTR elements in TT CD vs CC CD were increased in the hypermethylated group (p=0.0009), and CC FASD vs CC CD hypomethylated group (p=0.0023).

While evolutionary young retrotransposable elements were better covered by WGBS, the proportion of these elements remained minimal (≤ 3%) (Fig. 29). As described before, L1MdA_I, L1MdTf_I, L1MdTf_II, L1MdA_II, L1MdGf_I, L1MdTf_III, L1MdGf_II and L1MdA_III subfamilies were assessed (Karahan et al., 2021). Fisher's exact tests from all comparisons revealed no significant differences.

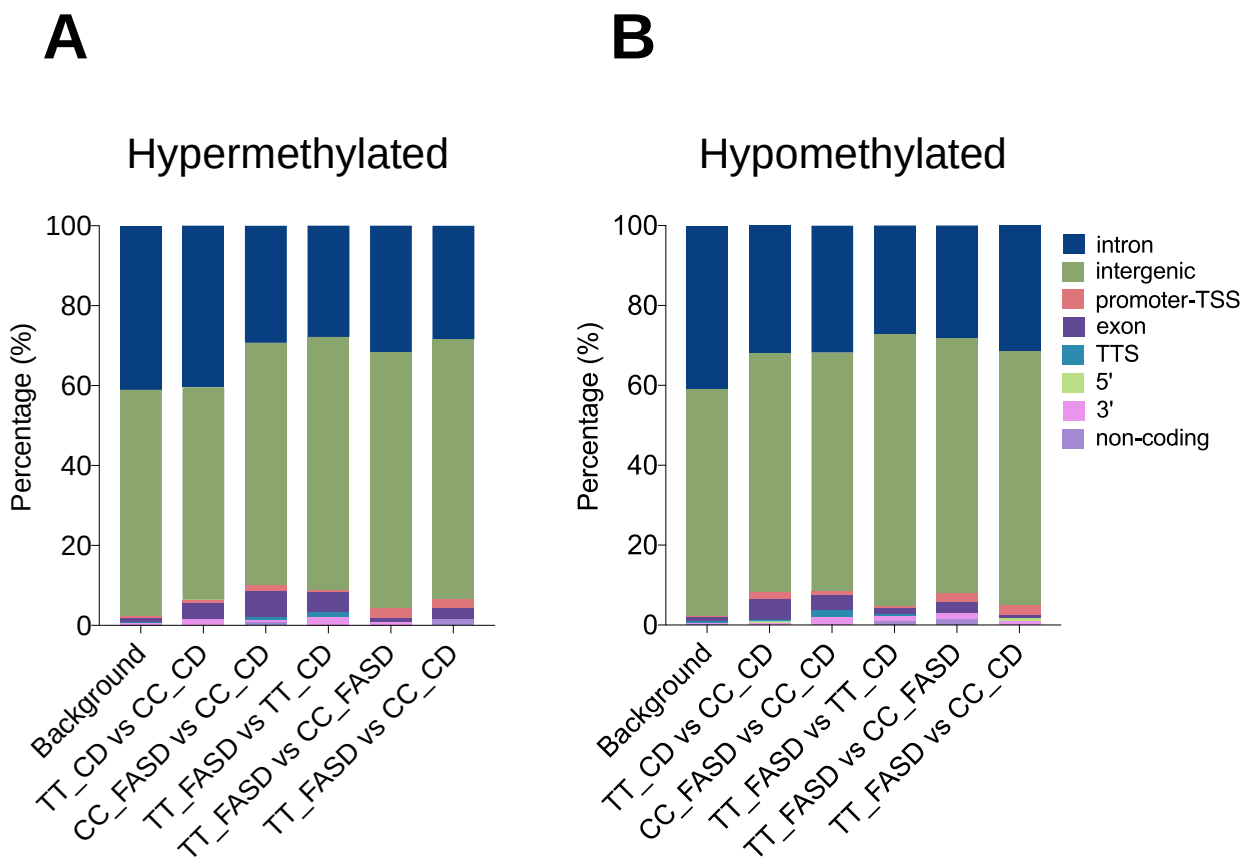


Figure 27. Genomic distribution of WGBS data.

The genomic distribution was similar among all different groups. Intergenic and intronic areas were the most represented. Exons and promoter-TSS areas were more affected in certain groups. **(A)** Hypermethylated DMTs. **(B)** hypomethylated DMTs. WGBS, whole-genome bisulfite sequencing; DMTs, differentially methylated tiles.

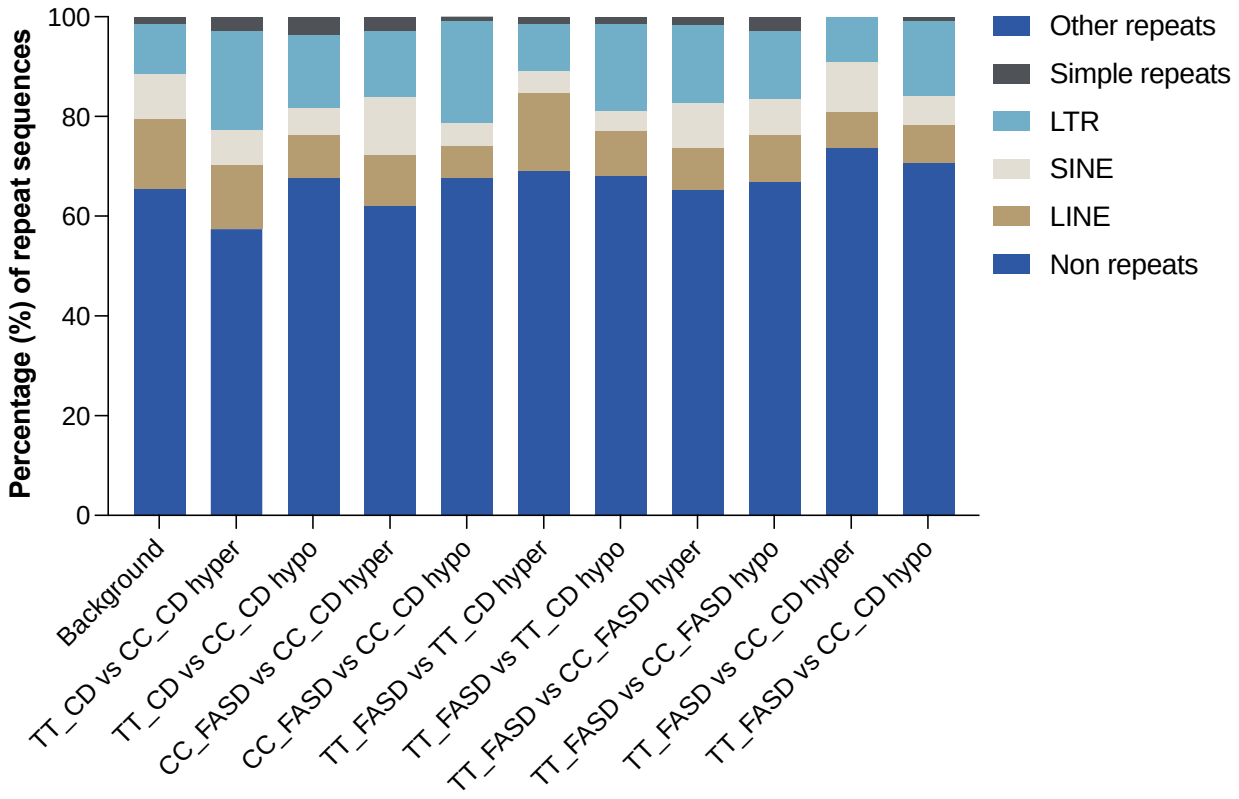


Figure 28. Proportion of repeat sequences of WGBS data.

Repeat sequences from WGBS at 10x coverage and ≥ 5% methylation difference. LTR, long terminal repeat; LINE, long interspersed nuclear element**;** short interspersed nuclear element; hyper, hypermethylated; hypo, hypomethylated

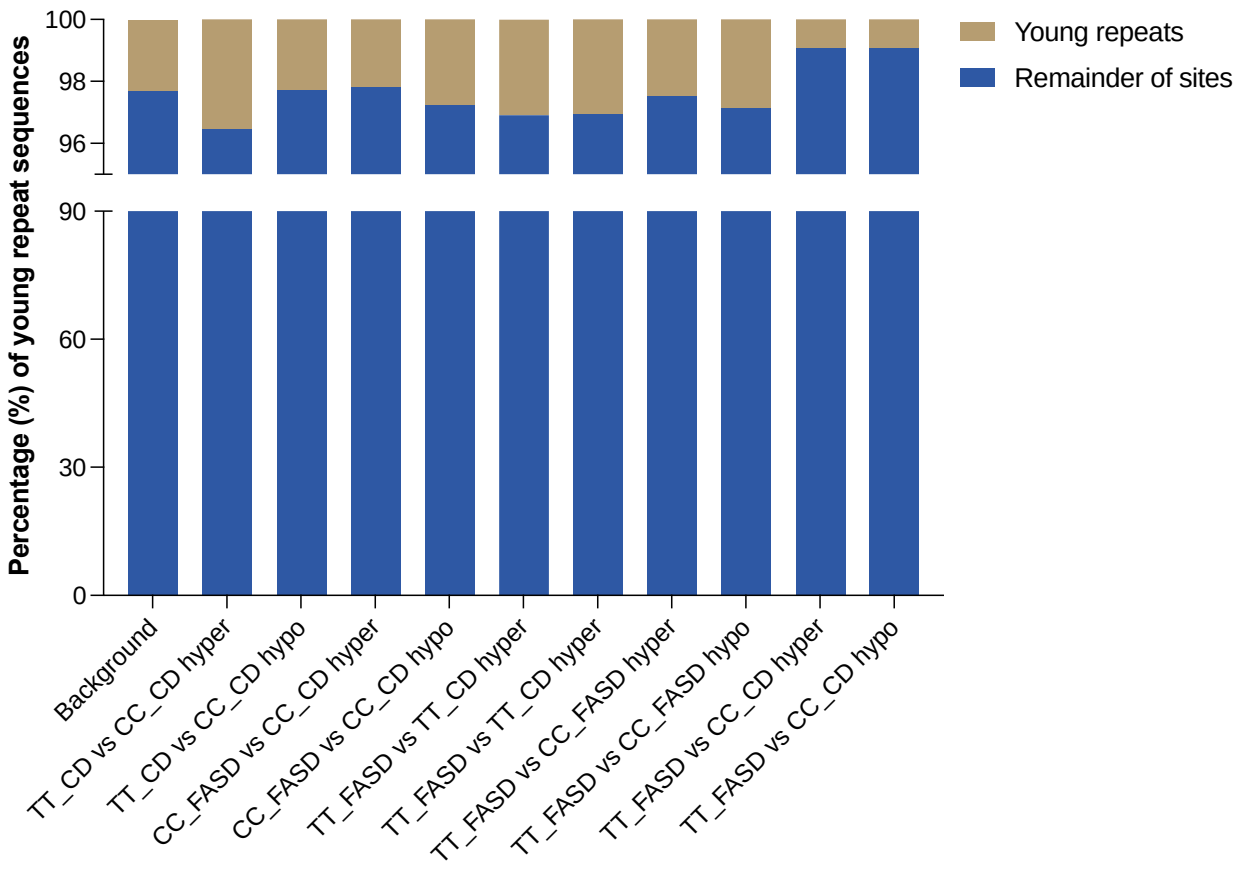


Figure 29. Proportion of evolutionary young retrotransposable elements of WGBS data.

Young retrotransposons from WGBS at 10x coverage and ≥ 5% methylation difference. Only 2-3% young retrotransposons were affected overall. Specifically, L1MdA_I, L1MdTf_I, L1MdTf_II, L1MdA_II, L1MdGf_I, L1MdTf_III, L1MdGf_II and L1MdA_III subfamilies were interrogated. Nearly none were covered by WGBS in all the groups studied. Hyper, hypermethylated; hypo, hypomethylated.

3.5.6 Regional analysis of DMTs of FASD and CD samples

As with the Illumina Array data, we wished to ascertain if any larger-scale effects could be found by merging DMT's into regions. To achieve this, neighboring DMTs located within 100bp up to 10,000 bp were merged as before. Again, only few DMTs merged into differentially methylated regions (DMRs). Using the comparison with the most DMTs as an example (TT CD vs CC CD), when merged at 100bp only 2 DMTs out of 131 DMTs merged into regions (1.5%). From that point forward, for every distance attempted up to the maximum (10,000bp), only 4 DMTs merged into regions (3%), the rest remained as unmerged tiles (Fig. 30). This suggests that, in these mice, the effects caused by the TT genotype and different folic acid diets do not significantly affect the sperm DNA methylation values at a regional level.

3.5.7 Gene ontology enrichment analysis and reduced visualization of FASD and CD samples

As with the Illumina Array, gene ontology (GO) term pathway enrichment was performed on the three most affected comparisons: *Mthfr* 677TT CD vs *Mthfr* 677CC CD, followed by *Mthfr* 677CC FASD vs *Mthfr* 677CC FD and lastly *Mthfr* 677TT FASD vs *Mthfr* 677TT CD. Once again, cell adhesion was a common term, found in TT CD cs CC CD, CC FASD vs CC CD and TT FASD vs TT CD all hypomethylated. Another term shared in three groups was negative regulation of canonical *Wnt* signaling pathway, found in CC FASD vs CC CD hypomethylated and TT FASD vs TT CD hyper- and hypomethylated (Fig 31). Quite interestingly, more neurological terms were found with this analysis, such as axonogenesis (de novo generation of a long process of a neuron) in TT CD vs CC CD hypermethylated, CC FASD vs CC CD and TT FASD vs TT CD both hypomethylated;

positive regulation of excitatory postsynaptic potential in CC FASD vs CC CD and TT FASD vs TT CD both hypomethylated; chemical synaptic transmission in CC FASD vs CC CD and TT FASD vs TT CD both hypomethylated; positive regulation of neuron differentiation in TT FASD vs TT CD hypermethylated and lastly, regulation of synaptic vesicle exocytosis in CC FASD vs CC CD and TT FASD vs TT CD both hypomethylated (Fig 31). These results further suggest that MTHFR activity and folic acid status are both critical for normal neurological development and proper functioning.

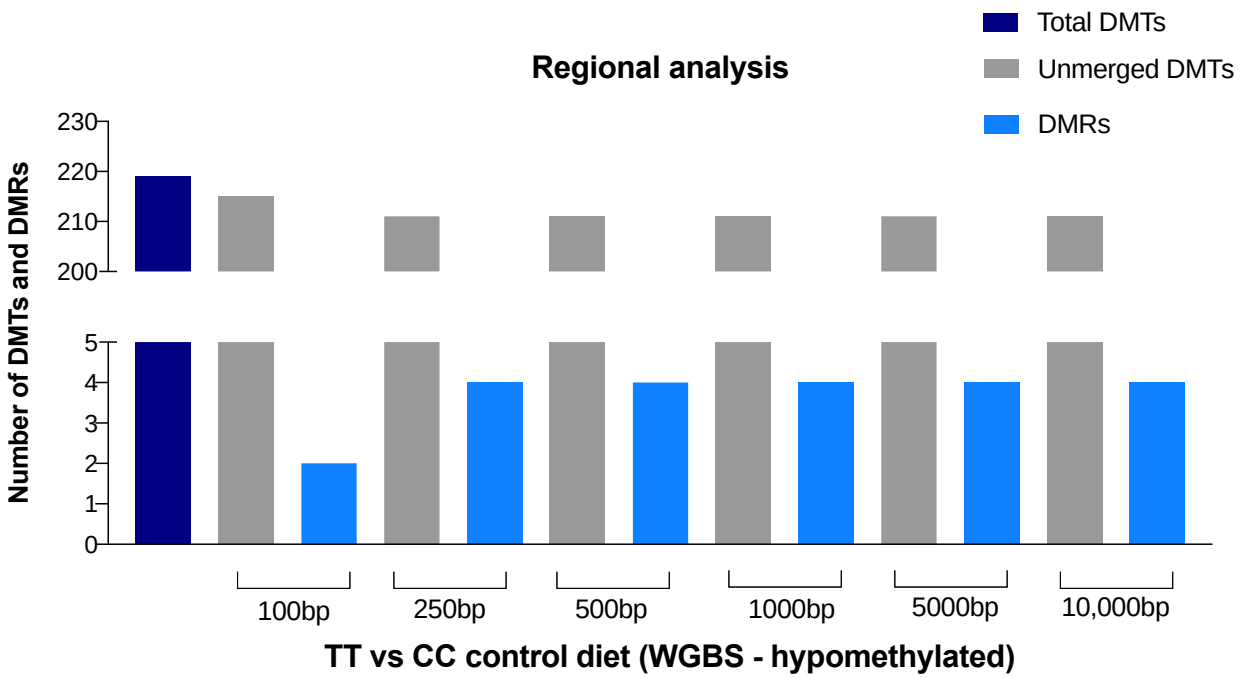


Figure 30. WGBS regional analysis of TT CD vs CC CD hypomethylation effect.

To investigate larger-scale regional changes in DNA methylation, neighboring differentially methylated tiles (DMTs) were merged into regions. Various distances were tested. The merged DMCs were located within 100 bp, 250bp, 500bp, 1000bp, 5000bp and 10,000bp of each other. DMTs devoid of any neighboring tiles were identified as unmerged DMTs. DMTs that merged were identified as differentially methylated regions (DMRs). Here TT CD vs CC CD hypermethylation effect is shown as it had the largest effect out of all the different comparisons.

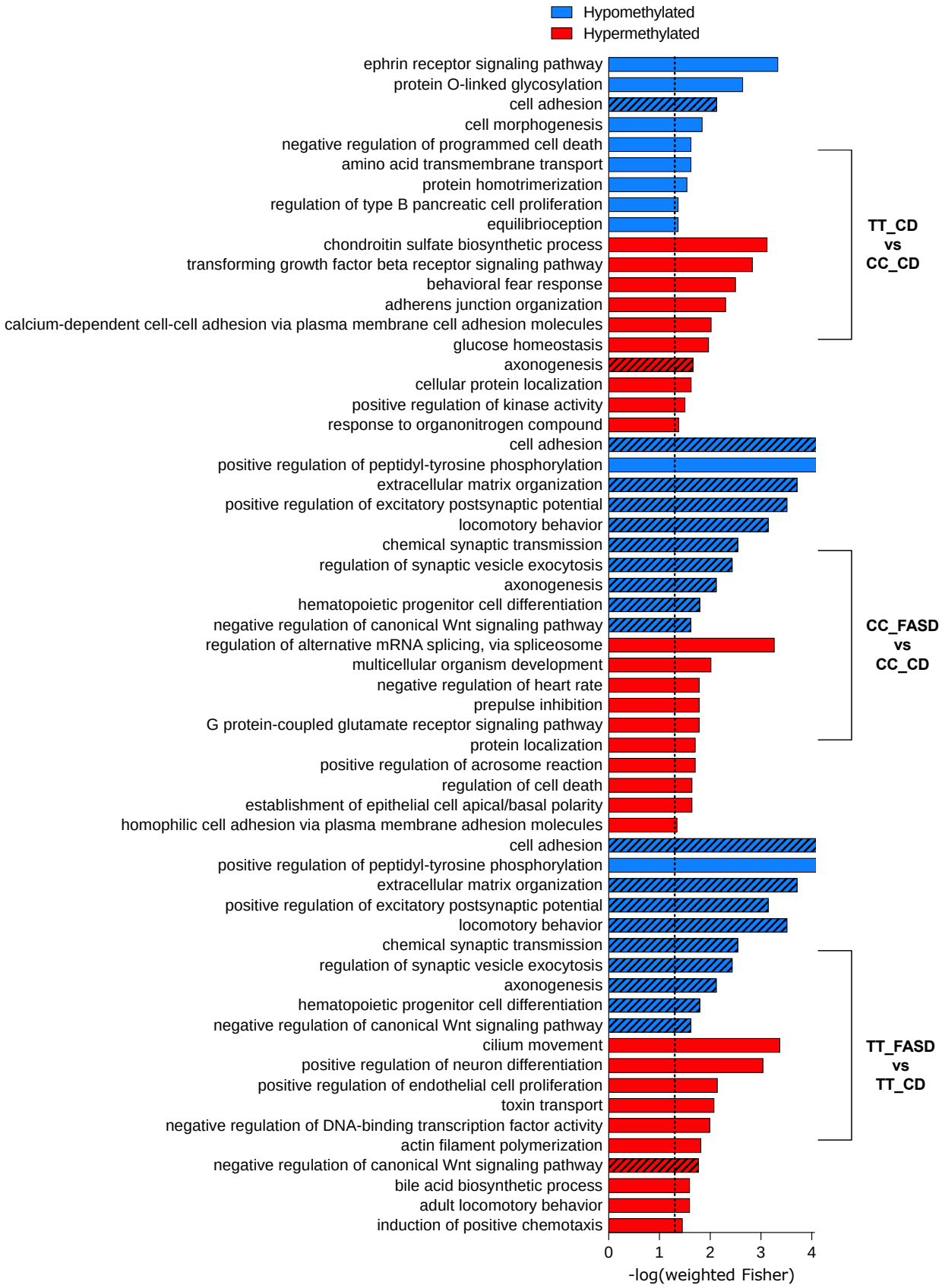

Hypomethylated
Hypermethylated
ephrin receptor signaling pathway
protein O-linked glycosylation
cell adhesion
cell morphogenesis
negative regulation of programmed cell death
amino acid transmembrane transport
protein homotrimerization
regulation of type B pancreatic cell proliferation
equilibrioception
chondroitin sulfate biosynthetic process
transforming growth factor beta receptor signaling pathway
behavioral fear response
adherens junction organization
calcium-dependent cell-cell adhesion via plasma membrane cell adhesion molecules
glucose homeostasis
axonogenesis
cellular protein localization
positive regulation of kinase activity
response to organonitrogen compound
cell adhesion
positive regulation of peptidyl-tyrosine phosphorylation
extracellular matrix organization
positive regulation of excitatory postsynaptic potential
locomotory behavior
chemical synaptic transmission
regulation of synaptic vesicle exocytosis
axonogenesis
hematopoietic progenitor cell differentiation
negative regulation of canonical Wnt signaling pathway
regulation of alternative mRNA splicing, via spliceosome
multicellular organism development
negative regulation of heart rate
prepulse inhibition
G protein-coupled glutamate receptor signaling pathway
protein localization
positive regulation of acrosome reaction
regulation of cell death
establishment of epithelial cell apical/basal polarity
homophilic cell adhesion via plasma membrane adhesion molecules
cell adhesion
positive regulation of peptidyl-tyrosine phosphorylation
extracellular matrix organization
positive regulation of excitatory postsynaptic potential
locomotory behavior
chemical synaptic transmission
regulation of synaptic vesicle exocytosis
axonogenesis
hematopoietic progenitor cell differentiation
negative regulation of canonical Wnt signaling pathway
cilium movement
positive regulation of neuron differentiation
positive regulation of endothelial cell proliferation
toxin transport
negative regulation of DNA-binding transcription factor activity
actin filament polymerization
negative regulation of canonical Wnt signaling pathway
bile acid biosynthetic process
adult locomotory behavior
induction of positive chemotaxis
TT_CD vs CC_CD
CC_FASD vs CC_CD
TT_FASD vs TT_CD
0
1
2
3
4
-log(weighted Fisher)

Figure 31. Gene ontology enrichment analysis (TopGO & rrvgo) with WGBS.

GO enrichment analysis of DMTs (100bp/tile) analyzed by WGBS at 10x coverage and ≥5% methylation difference. All significant terms were reduced using high stringency (threshold= 0.9) using the R package Reduce + Visualize GO (rrvgo). The dotted line indicates the p≤0.05 threshold for significance for FDR. Hatched bars indicate common enriched pathways shared among different groups.

3.5.8 Corresponding genes of DMTs of FASD and CD samples

The top 20 most affected genes were listed, the highest difference in methylation being ~20% change and lowest ~8% for the majority of the groups (Table 5). The most affected group, TT CD vs CD, displayed interesting genes including *Fgf12*, *Zfp608*, *Lmo3*, *Lsamp, Pcdhb19* and *Snurf*. In the CC FASD vs CC CD hypermethylated group *Pde4b, Ntm* and *Nrxn3* were found.

Certain genes of interest were affected in more than one group. For example, *Lmo3* which appeared in TT CD vs CC CD hypermethylated was also affected in TT FASD vs TT CD hypermethylated. *Pcdhb19* was altered in TT CD vs CC CD, TT FASD vs TT CD hypo- and hypermethylated, respectively. *Zfp608* in TT CD vs CC CD hyper- and TT FASD vs TT CD hypomethylated, respectively. *Fgf12* found in TT CD vs CC CD hypermethylated was also affected in CC FASD vs CC CD hypomethylated. A comparison of all affected genes, including those closest to an intergenic area showed commonalities between some groups (Table 6).

Gene Name	Gene Description	Annotation	Mean difference in methylation
	TT CD vs CC CD	**Hypermethylation**	
Fgf12	fibroblast growth factor 12	intron	18.07
A930012L18Rik	RIKEN cDNA A930012L18 gene	intron	18.00
Dach1	dachshund 1 (Drosophila)	intron	15.97
Zfp608	zinc finger protein 608	intron	15.75
Plekhf1	pleckstrin homology domain containing	intron	14.98
Tmprss11d	transmembrane protease, serine 11d	exon	14.26
Lmo3	LIM domain only 3	intron	14.16
Magi2	membrane associated guanylate kinase,	intron	13.59
D6Ertd474e	DNA segment, Chr 6, ERATO Doi 474, expressed	intron	13.46
Erich1	glutamate rich 1	intron	13.34
Tctex1d1	Tctex1 domain containing 1	exon	13.15
Fndc1	fibronectin type III domain containing 1	intron	12.63
Pcdh9	protocadherin 9	intron	12.40
Aldh1l2	aldehyde dehydrogenase 1 family, member L2	intron	11.91
Dpp4	dipeptidylpeptidase 4	3' UTR	11.39
Lsamp	limbic system-associated membrane protein	intron	10.78
Dclk1	doublecortin-like kinase 1	exon	10.70
Ank1	ankyrin 1, erythroid	intron	10.48
3110039I08Rik	RIKEN cDNA 3110039I08 gene	intron	10.39
Gm20745	predicted gene, 20745	exon	10.25
	TT CD vs CC CD	**Hypomethylation**	
Pth2r	parathyroid hormone 2 receptor	intron	-22.37
Pcdhb19	protocadherin beta 19	exon	-21.63
Abcd3	ATP-binding cassette, sub-family D (ALD), member 3	intron	-20.51
1700112J05Rik	RIKEN cDNA 1700112J05 gene	intron	-19.29
Kif6	kinesin family member 6	intron	-18.40
Rtkn2	rhotekin 2	intron	-18.15
Pdcd1lg2	programmed cell death 1 ligand 2	intron	-17.96
Sema3e	sema domain, immunoglobulin domain (Ig)	intron	-17.71
9430014N10Rik	RIKEN cDNA 9430014N10 gene	intron	-17.65
Zbbx	zinc finger, B-box domain containing	intron	-17.48
Rbms3	RNA binding motif, single stranded interacting protein	intron	-17.39
Snurf	SNRPN upstream reading frame	intron	-16.97
Plcb4	phospholipase C, beta 4	intron	-16.78
Sema3e	sema domain, immunoglobulin domain (Ig)	intron	-16.46
Klrd1	killer cell lectin-like receptor, subfamily D, member 1	intron	-16.15
1700029J03Rik	RIKEN cDNA 1700029J03 gene	intron	-15.83
Kif6	kinesin family member 6	intron	-15.22
Tat	tyrosine aminotransferase	promoter-TSS	-14.61
Grip1os2	glutamate receptor interacting protein 1, opposite strand 2	intron	-14.43
Snx11	sorting nexin 11	TTS	-14.43
	CC FASD vs CC CD	**Hypermethylation**	
Slc15a5	solute carrier family 15, member 5	intron	21.90
Gsdmcl2	gasdermin C-like 2	intron	21.20
4930567K20Rik	RIKEN cDNA 4930567K20 gene	intron	21.07
Lect2	leukocyte cell-derived chemotaxin 2	exon	20.86
Tmem14c	transmembrane protein 14C	promoter-TSS	18.07
Tspan12	tetraspanin 12	intron	17.28
Akap7	A kinase (PRKA) anchor protein 7	intron	16.59
4930515L03Rik	RIKEN cDNA 4930515L03 gene	intron	15.68
Asb14	ankyrin repeat and SOCS box-containing 14	exon	15.50
Adra1a	adrenergic receptor, alpha 1a	intron	15.44
Pcdh9	protocadherin 9	intron	15.09
Csrnp3	cysteine-serine-rich nuclear protein 3	intron	14.87
Cdh6	cadherin 6	intron	14.62
Zfp518b	zinc finger protein 518B	intron	14.52
Tspan12	tetraspanin 12	intron	14.29
Pde4b	phosphodiesterase 4B, cAMP specific	intron	12.88
Srrm4	serine/arginine repetitive matrix 4	intron	12.74
Trhde	TRH-degrading enzyme	intron	11.85

Mbtps1	membrane-bound transcription factor peptidase, site 1	intron	11.33
BC048546	cDNA sequence BC048546	intron	11.28
	CC FASD vs CC CD	**Hypomethylation**	
Cdca7l	cell division cycle associated 7 like	intron	-17.13
Acvr1	activin A receptor, type 1	intron	-16.40
Nxpe3	neurexophilin and PC-esterase domain family, member 3	intron	-16.09
1700022A22Rik	RIKEN cDNA 1700022A22 gene	TTS	-14.84
1700112J05Rik	RIKEN cDNA 1700112J05 gene	intron	-14.37
Fgf12	fibroblast growth factor 12	intron	-14.32
Acvr1	activin A receptor, type 1	intron	-13.97
Ntm	neurotrimin	intron	-13.86
Prg4	proteoglycan 4 (megakaryocyte stimulating factor)	intron	-11.59
Tmem117	transmembrane protein 117	intron	-11.59
Ptprz1	protein tyrosine phosphatase, receptor type Z, polypeptide 1	intron	-11.46
App	amyloid beta (A4) precursor protein	intron	-10.96
Nrxn3	neurexin III	intron	-9.54
Gm13490	predicted gene 13490	intron	-9.20
Ppid	peptidylprolyl isomerase D (cyclophilin D)	TTS	-8.53
Armc3	armadillo repeat containing 3	intron	-7.93
Wrnip1	Werner helicase interacting protein 1	intron	-7.79
Rims1	regulating synaptic membrane exocytosis 1	intron	-7.63
Dscam	Down syndrome cell adhesion molecule	intron	-7.43
A430033K04Rik	RIKEN cDNA A430033K04 gene	intron	-7.30
	TT FASD vs TT CD	**Hypermethylation**	
Tmem39a	transmembrane protein 39a	intron	24.98
Exoc4	exocyst complex component 4	intron	22.20
Pde1a	phosphodiesterase 1A, calmodulin-dependent	TTS	22.02
Arl6	ADP-ribosylation factor-like 6	intron	21.45
Pcdhb19	protocadherin beta 19	exon	21.18
Otogl	otogelin-like	intron	20.51
Prb1	proline-rich protein BstNI subfamily 1	intron	19.63
Abca12	ATP-binding cassette, sub-family A (ABC1), member 12	intron	18.52
Cntnap2	contactin associated protein-like 2	intron	17.91
1700006F04Rik	RIKEN cDNA 1700006F04 gene	intron	17.68
Inpp4b	inositol polyphosphate-4-phosphatase, type II	intron	16.23
Lipk	lipase, family member K	intron	15.95
Hmgcll1	3-hydroxymethyl-3-methylglutaryl-Coenzyme A lyase-like 1	intron	15.13
Pcdhb19	protocadherin beta 19	exon	14.47
Gm13090	predicted gene 13090	intron	13.98
Cdh6	cadherin 6	3'UTR	13.78
Sema6a	sema domain, transmembrane domain ™	intron	13.42
Kif6	kinesin family member 6	intron	12.96
Mir6239	microRNA 6239	intron	12.58
Vwa3a	von Willebrand factor A domain containing 3A	intron	12.43
	TT FASD vs TT CD	**Hypomethylation**	
Slc7a15	solute carrier family 7, member 15	intron	-16.42
Kctd15	potassium channel tetramerisation domain containing 15	intron	-16.19
Slit3	slit homolog 3 (Drosophila)	intron	-16.04
Spata16	spermatogenesis associated 16	intron	-16.03
Pcdh9	protocadherin 9	intron	-15.84
Atad2b	ATPase family, AAA domain containing 2B	intron	-15.74
Zfp608	zinc finger protein 608	intron	-15.35
Mecom	MDS1 and EVI1 complex locus	intron	-15.22
4930542C21Rik	RIKEN cDNA 4930542C21 gene	intron	-14.63
Etl4	enhancer trap locus 4	intron	-13.62
Ttc30a1	tetratricopeptide repeat domain 30A1	intron	-13.27
Nron	non-protein coding RNA, repressor of NFAT	intron	-13.00
Galc	galactosylceramidase	intron	-12.56
1700108J01Rik	RIKEN cDNA 1700108J01 gene	intron	-12.09
Sox6	SRY (sex determining region Y)-box 6	intron	-12.01
Gm20754	predicted gene, 20754	intron	-11.30
Lect2	leukocyte cell-derived chemotaxin 2	intron	-11.14

Anks1b	ankyrin repeat and sterile alpha motif domain	intron	-11.11
Papss2	3'-phosphoadenosine 5'-phosphosulfate synthase 2	intron	-10.91
Acss3	acyl-CoA synthetase short-chain family member 3	intron	-10.77
	TT FASD vs CC FASD	**Hypermethylation**	
Otogl	otogelin-like	intron	19.91
Endou	endonuclease, polyU-specific	intron	18.79
Csrnp3	cysteine-serine-rich nuclear protein 3	5' UTR	18.16
Sdpr	serum deprivation response	intron	17.63
Inpp4b	inositol polyphosphate-4-phosphatase, type II	intron	17.28
Tekt3	tektin 3	intron	16.56
Nxpe3	neurexophilin and PC-esterase domain family, member 3	intron	16.37
Lrrc66	leucine rich repeat containing 66	3' UTR	15.12
Epha3	Eph receptor A3	intron	14.29
4931408D14Rik	RIKEN cDNA 4931408D14 gene	promoter-TSS	14.27
2810442N19Rik	RIKEN cDNA 2810442N19 gene	intron	13.89
Ccdc110	coiled-coil domain containing 110	intron	12.42
Ptprg	protein tyrosine phosphatase, receptor type, G	intron	11.84
Ust	uronyl-2-sulfotransferase	intron	11.02
4933413L06Rik	RIKEN cDNA 4933413L06 gene	intron	10.74
Polb	polymerase (DNA directed), beta	intron	10.05
Gm19303	predicted gene, 19303	intron	9.54
Skor2	SKI family transcriptional corepressor 2	intron	9.37
Erbb4	erb-b2 receptor tyrosine kinase 4	intron	9.28
Snhg3	small nucleolar RNA host gene 3	intron	8.88
	TT FASD vs CC FASD	**Hypomethylation**	
Ppp2r2b	protein phosphatase 2, regulatory subunit B, beta	promoter-TSS	-22.09
Spag16	sperm associated antigen 16	intron	-20.24
Gm7056	predicted gene 7056	intron	-19.56
Rpl31	ribosomal protein L31	intron	-17.64
Mir7223	microRNA 7223	intron	-17.10
Pde11a	phosphodiesterase 11A	intron	-16.86
Nav3	neuron navigator 3	intron	-15.70
Slit3	slit homolog 3 (Drosophila)	intron	-15.57
Atg3	autophagy related 3	intron	-15.42
3110099E03Rik	RIKEN cDNA 3110099E03 gene	promoter-TSS	-14.74
Lox	lysyl oxidase	3' UTR	-14.29
Carlr	cardiac and apoptosis-related long non-coding RNA	intron	-14.25
Loxl4	lysyl oxidase-like 4	intron	-13.30
4930426D05Rik	RIKEN cDNA 4930426D05 gene	exon	-12.79
Mir3106	microRNA 3106	intron	-12.63
Mdm2	transformed mouse 3T3 cell double minute 2	intron	-11.65
Ifih1	interferon induced with helicase C domain 1	intron	-11.60
Rfx6	regulatory factor X, 6	intron	-11.07
Tcf7l2	transcription factor 7 like 2, T cell specific, HMG box	promoter-TSS	-10.89
Gm20754	predicted gene, 20754	intron	-10.67
	TT FASD vs CC CD	**Hypermethylation**	
Abca12	ATP-binding cassette, sub-family A (ABC1), member 12	intron	20.07
Gm13090	predicted gene 13090	intron	19.82
Endou	endonuclease, polyU-specific	intron	19.54
Skint6	selection and upkeep of intraepithelial T cells 6	intron	19.02
A930011G23Rik	RIKEN cDNA A930011G23 gene	intron	18.88
Pcdh7	protocadherin 7	intron	18.76
Lpar6	lysophosphatidic acid receptor 6	exon	18.44
Mucl1	mucin-like 1	intron	16.72
Edil3	EGF-like repeats and discoidin I-like domains 3	intron	15.28
Epha3	Eph receptor A3	intron	14.57
Nedd9	neural precursor cell expressed, down-regulated gene 9	exon	13.25
Lmo3	LIM domain only 3	intron	13.21
Caln1	calneuron 1	intron	11.88
1700010K23Rik	RIKEN cDNA 1700010K23 gene	intron	11.80
Ror2	receptor tyrosine kinase-like orphan receptor 2	intron	11.51
Cfap43	cilia and flagella associated protein 43	intron	11.29

2810442N19Rik	RIKEN cDNA 2810442N19 gene	intron	11.28
8430437L04Rik	RIKEN cDNA 8430437L04 gene	promoter-TSS	9.57
Gm19303	predicted gene, 19303	intron	9.24
Zfp985	zinc finger protein 985	intron	8.84
	TT FD vs CC FD	**Hypomethylation**	
C130083M11Rik	RIKEN cDNA C130083M11 gene	promoter-TSS	-25.79
4930518J21Rik	RIKEN cDNA 4930518J21 gene	promoter-TSS	-21.04
Spert	spermatid associated	intron	-19.79
Pex3	peroxisomal biogenesis factor 3	TTS	-17.61
Ntm	neurotrimin	intron	-17.22
Spock1	sparc/osteonectin, cwcv and kazal-like domains	intron	-15.62
Acvr1	activin A receptor, type 1	intron	-15.15
Loxl4	lysyl oxidase-like 4	intron	-14.66
Rmst	rhabdomyosarcoma 2 associated transcript	non-coding	-14.57
Grip1os2	glutamate receptor interacting protein 1	intron	-14.55
Opcml	opioid binding protein/cell adhesion molecule-like	intron	-13.56
Gm17660	predicted gene, 17660	exon	-12.85
Klhl41	kelch-like 41	exon	-12.37
Zfp608	zinc finger protein 608	intron	-12.24
Mir3106	microRNA 3106	intron	-11.81
Nav3	neuron navigator 3	intron	-11.25
Dock10	dedicator of cytokinesis 10	intron	-8.54
Pitpnb	phosphatidylinositol transfer protein, beta	intron	-8.53
G530011O06Rik	RIKEN cDNA G530011O06 gene	intron	-8.42
Rps23	ribosomal protein S23	intron	-8.07

Table 5. Top 20 most affected genes from WGBS DMTs.

Mean difference in methylation corresponds to methylation values from whole genome bisulfite sequencing data.

Gene Name	Gene Description	Annotation	Mean meth diff	Distance to TSS
	(TT FASD vs CC CD) & (TT CD vs CC CD)	**Hypermethylation**		
4930405D11Rik	RIKEN cDNA 4930405D11 gene	Intergenic	21.19	252559
2510009E07Rik	RIKEN cDNA 2510009E07 gene	Intergenic	9.00	-33585
4930548K13Rik	RIKEN cDNA 4930548K13 gene	Intergenic	7.41	-565801
Lmo3	LIM domain only 3	Intron	14.16	-1282
AW046200	expressed sequence AW046200	Intron	7.49	-67403
1700007B14Rik	RIKEN cDNA 1700007B14 gene	Intergenic	6.68	-50144
Impg1	interphotoreceptor matrix proteoglycan 1	Intergenic	16.17	-143012
	(TT FASD vs CC CD) & (TT CD vs CC CD)	**Hypomethylation**		
Grip1os2	glutamate receptor interacting protein 1, opposite strand 2	Intron	-14.55	114849
Slc16a7	solute carrier family 16 (monocarboxylic acid transporters)	Intergenic	-6.81	265085
Adcy2	adenylate cyclase 2	Intergenic	-8.64	-58809
Col11a1	collagen, type XI, alpha 1	Intergenic	-19.78	-72690
4930433B08Rik	RIKEN cDNA 4930433B08 gene	Intergenic	-14.33	-234373
Unc5c	unc-5 netrin receptor C	Intergenic	-11.26	-78214
Ntm	neurotrimin	Intron	-17.22	816079
	(TT FASD vs CC CD) & (CC FASD vs CC CD)	**Hypermethylation**		
Ufm1	ubiquitin-fold modifier 1	Intergenic	15.01	-51675
Slc35b3	solute carrier family 35	Intergenic	8.98	-15343

Table 6. Shared genes/closest genes from WGBS compared groups

Mean meth diff, mean difference in methylation; TSS, transcription start site.

CHAPTER 4: DISCUSSION

The folate metabolism is crucial for the synthesis of nucleic acid precursors, amino acids and for the production of S-adenosylmethionine (SAM), a universal methyl donor (Clare et al., 2019)**.** MTHFR is a crucial enzyme which has been under investigation for many years due to its key role in the folate metabolic pathway, which allows the generation of methyl groups. It is also implicated in an increased risk in numerous health disorders when not functioning properly, including a negative impact in male fertility and sperm DNA methylation. In many cases, the exact relationship between MTHFR deficiency and these health risks are not well understood, warranting further investigation.

A common polymorphism of this gene in humans is the *MTHFR* 677C>T variant which results in a thermolabile enzyme with ~50% reduced activity. Previous studies have made use of a knockout mouse model (*Mthfr*$^{+/-}$) to investigate the underlying mechanisms of these disorders. These mice reproduce certain parameters encountered in humans, such as hyperhomocysteinemia and reduced enzymatic activity (Chen et al., 2001). Despite this, the enzyme that these mice produce is not thermolabile, precipitating the interest of developing a mouse model that is a true representative model of the human polymorphism.

The new mouse model was developed by introducing the 677C>T polymorphism into the mouse genome using CRISPR and backcrossing it into the C57BL/6J background. These mice reproduce biochemical parameters consistent with human clinical data, such as a thermolabile enzyme with reduced enzymatic activity in the liver and elevated plasma homocysteine levels (Reagan et al., accepted in *Journal of Cerebral Blood Flow & Metabolism*, 2022). I was interested in characterizing the sperm DNA

epigenome of these newly developed mouse and the effects of folate deficiency and folic acid supplementation.

Initial analyses were performed using a recently available array, the Illumina Mouse Methylation BeadChip (>285k), followed by WGBS for subsequent samples on a different diet. By the time the project started, no studies had been published regarding the new mouse methylation array. A few articles have now been published (Garcia-Prieto et al., 2022; Vrooman et al., 2022; Khodasevich et al., 2022), but to my knowledge, there have not been studies employing this Array on sperm DNA. Likewise, due to its cost and time-consumption, WGBS has not often been used for similar projects. For all these reasons, this project is of great interest because it combines the use of a newly developed mouse model, a recently available array and includes WGBS, which covers ~20 million CpGs (in contrast to ~285k covered by the Array) and it is the gold standard for global methylation analyses.

4.1 Effect of *Mthfr* 677TT genotype on *Mthfr* 677C>T mice and sperm DNA methylation

A greater tendency for hypermethylation in sperm DNA has been shown before in BALB/c *Mthfr*$^{+/-}$ mice (Aarabi et al., 2018), and the same is true for men with the 677TT polymorphism (Chan et al., 2019). Notably, the knockout mouse study was analyzed using reduced representation bisulfite sequencing (RRBS) which targets primarily classical CpG island and other high-GC content sequences. Due to the way it is performed, this analysis provides a biased sequence selection and does not cover non-CpG areas, genome-wide CpGs, and CpGs in areas without the enzyme restriction site (Meissner et al., 2005).

For the human study, a vast proportion of the observed changes in sperm DNA were found to be in dynamic sperm sites (sites with 20-80% methylation, which are more likely to be affected than sites with 0-20% or 80-100% methylation) analyzed with a customized capture technique for sperm. These sites would not have been detected with other available approaches (Chan et al., 2019). Such dynamic sites were discovered through WGBS, which displayed nearly 1 million CpGs possessing intermediate levels of methylation. These sites along with 2 million commonly assessed CpGs were used to create the customized capture panel for targeted interrogation of the human sperm methylome and test its ability to detect the effects of altered folate metabolism. The vast majority (>80%) of CpGs that were found to be altered in men homozygous to the *MTHFR* 677C>T polymorphism and a high dose of FASD, were located in dynamic sperm CpGs.

Initial results obtained with the Illumina Array determined that the TT genotype was responsible for the largest effect among all groups of genotypes and diets. When comparing the TT vs CC genotypes on a control diet, we observed hyper- and hypomethylation, with a predominance for the latter. However, the difference in methylation was minor, most of them having between 0 and 5% change. This might be explained by the fact that the Array does not cover intergenic regions well. Indeed, studies have consistently shown that, in the sperm genome, intergenic regions tend to be the most affected, suggesting that these genomic areas are the most susceptible to MTHFR deficiency as well as different folic acid status (Aarabi et al., 2018, Karahan et al, 2021; Chan et al., 2022 [submitted to Andrology]). In the Array, we found that introns were most affected at about 35-45% followed by intergenic areas with ~30-32% representation. This array, as previously mentioned, does not cover intergenic areas as well as other

technologies. As stated in a recent study validating the Illumina Mouse Methylation Array (Garcia Prieto et al., 2022), “from a functional standpoint, 70.4% of the CpG sites were located on gene bodies and 16.9% in intergenic regions”.

Subsequent analyses with WGBS further confirmed that the TT genotype caused the largest effect out of all diet and genotype groups, once again displaying hyper- and hypomethylation, with hypomethylation being more prominent, differing to what we hypothesized. Here, intergenic regions were the most affected areas, as we expected. The DMTs exhibited a larger difference in methylation compared to the Illumina Array, however, the total number of DMTs was not extensive. This suggests that these mice and their sperm DNA are not greatly affected by the *Mthfr* 677C>T polymorphism. For this reason, it might be interesting to perform similar studies on a BALB/c background, as it has been shown that these mice are more susceptible to MTHFR deficiency (Chan et al., 2010), which might also help explain population differences in infertility among men with the same polymorphism.

4.2 Effects of folate deficiency on *Mthfr* 677C>T mice and sperm DNA methylation

Previous experiments employing folate deficient diets have not shown affected overall health. These studies have been focused on a lifetime exposure to a 7-fold folate deficient diet (Lambrot et al., 2013; Ly et al., 2017; Chan et al., 2022) and early development or post-weaning exposures to a 0 mg/kg folate deficient diet (Swayne et al., 2012). With that knowledge, in this project we did not expect the mice exposed to the folate deficient diet to present overall adverse health outcomes. This was confirmed by the normal parameters in body weights and reproductive organ weights.

In the case of sperm counts, a lifetime exposure to 7-fold folate deficient diet from a previous study caused a significant decrease in sperm counts of BALB/c mice (Ly et al., 2017). The same was true for mice exposed at the post-weaning period to a 0 mg/kg folate deficient diet (Swayne et al., 2012). In the present study, a 7-fold folate deficient diet was used, however the exposure occurred during the postnatal period, at 4 weeks of age. We did not encounter a reduction in total sperm counts, suggesting that prenatal exposures as well as during early development are more prone to have this effect, as opposed to postnatally. This could be perhaps because MTHFR is highly expressed during gonad development, when DNA methylation patterns are being established.

The effects on the DNA methylome assessed in this study revealed both hypo- and hypermethylation, with a predominance of hypomethylation. Similar findings have been shown on BALB/c mice with reduced representation bisulfite sequencing (RRBS) (Chan et al., 2022). In Lambrot et al., 2013, they found small changes in the DNA methylation of folate-deficient animals assessed by methylated DNA immunoprecipitation. Additionally, they investigated whether folate availability could alter the sperm histone modification pattern by measuring global levels of methylation of histone H3 at K9 and K4 in epididymal sperm extracts. They reported significantly reduced H3K4 and K9 monomethylation and reduced H3K9 trimethylation, indicating that histone methylation levels in sperm are sensitive to folate availability, and that different histone methyltransferases may possess different grades of sensitivity to altered methyl donor pools.

Other studies have further explored the role of histone marks. For example, it was demonstrated that sperm H3K4me2/3 were enriched at promoters of genes that are

expressed in the embryo, and that H3K4me2/3 establishment during spermatid to sperm transition was key for embryonic gene expression (Teperek et al., 2016). Another publication showed that overexpressing a H3K4 lysine to methionine mutant in mouse embryo had as a consequence an abnormal zygotic genome activation in the paternal pronucleus. This correlated with decreased H3K4me1/3 and resulted in developmental arrest of H3K4 mutant overexpressing embryos, further underscoring the importance of paternal histone H3K4me in embryo development (Aoshima et al., 2015).

Furthermore, a prior study generated transgenic mice that overexpressed the histone demethylase KDM1A in developing sperm only. Interestingly, the overexpression of this demethylase during spermatogenesis resulted in differential H3K4me2/3 enrichment at sperm promoters, early postnatal death of offspring and a high number of congenital anomalies in offspring (Lismer et al., 2020; Siklenka et al., 2015). A recently published study demonstrated that a folate deficient diet alters histone H3K4me3 in sperm at developmental genes and putative enhancers. Moreover, using a genetic mouse model in which sires have pre-existing altered H3K4me2/3 in sperm, they demonstrated that a folate deficient diet exacerbates the alterations in sperm H3K4me3 and embryonic gene expression, resulting in an increased severity of developmental defects on C57BL6/J mice (Lismer et al., 2021). All these studies help highlight the importance of histone marks in these crucial processes, emphasizing the need to include similar analyses such as in the project presented here.

Interestingly, the results obtained here showed that the folate deficient diet did not substantially increase the number of DMCs when combined with the TT genotype. TT CD vs CC CD displayed 141 more DMCs than TT FD vs CC CD. When placing the CC mice

on the folate deficient diet, the number of DMCs very closely resembled those caused by the TT genotype alone. While there was a clear effect of folate deficiency in the sperm DNA methylation of these mice, it did not have an added effect on the TT genotype as opposed to what we hypothesized. Moreover, the overall changes observed here were quite small, with most of the beta methylation difference being between ≤5%. This suggests that the postnatal exposure to the folate deficient diet does not have a major effect on the DNA methylome. It is also possible that histones are more susceptible to low folate availability than sperm DNA methylation marks.

A caveat of this experiment, however, is that we employed the Illumina Mouse Methylation Array to assess the folate deficient samples. As mentioned earlier, other investigations used RRBS to determine the impact of different folate diets, which targets a large portion of intergenic regions and includes areas with intermediate or dynamic levels of methylation (20-80% methylation) which have been shown to be more environmentally sensitive (Chan et al., 2019). The Array used here targets mainly genic areas that tend to be highly hyper- or hypomethylated and not responsive to different folate status. Comparably, as demonstrated in Chan et al., 2019 the Infinium EPIC methylation Array (850k) used in human studies covers a minute portion of sites with intermediate methylation. A larger-scale analysis would be necessary to confirm if a postnatal exposure to a folate deficient diet does not have a larger effect on sperm DNA methylation when combined with the TT genotype.

4.3 Effects of folic acid supplementation on sperm DNA methylation

Dietary folate is a significant source of methyl groups which are required for proper DNA methylation processes. In turn, folate is essential for adequate spermatogenesis as

it aids in the synthesis, repair, and methylation of DNA. Randomized studies have shown that high doses of folic acid supplementation (5mg daily intake; >10 times the recommended daily allowance) results in improved sperm concentration in infertile men (Wong et al., 2002; Ebisch et al., 2007). However, because folic acid supplementation in infertile men occurs during adult spermatogenesis, when DNA methylation patterns are being actively maintained and remodeled in male germ cells, a study (Aarabi et al., 2015) sought to determine the possible impact of this supplementation on the sperm epigenome. Using RBBS, they studied 30 men with idiopathic infertility and found a significant global loss of methylation across different regions of the sperm genome, mainly intergenic regions. This effect was greater in men homozygous for the *MTHFR* C677T polymorphism.

A follow-up study using BALB/c *Mthfr*$^{+/-}$ mice showed the same hypomethylating effect of a high dose of folic acid supplementation (Aarabi et al., 2018). These results were further confirmed by the customized capture approach in Chan et al., 2019, where men supplemented with the same dose of folic acid for 6 months presented predominantly hypomethylation, with a large proportion of affected dynamic sites. This suggests that high circulating and testicular folate levels down-regulate MTHFR, in turn decreasing methyl group availability and therefore resulting in a loss of DNA methylation.

A different randomized trial (Jenkins et al., 2021) assessed the impact of zinc and folic acid supplementation on sperm DNA methylation. In this study they supplemented men seeking fertility treatment with 5mg of folic acid and 30mg of elemental zinc, for 6 months or a placebo. Sperm DNA methylation was analyzed using the EPIC methylation array (Illumina). They found no differences between the treatment and placebo. While

these two studies used the same folic acid dose, there are two important distinctions that could explain the dissimilar results. First, in Jenkins et al., 2021, they included a supplementation of zinc. Second and most importantly, in Aarabi et al., 2015, they used RRBS to assess methylation, allowing them to report alterations at intergenic regions, repeat regions and transposable elements, all of which are not covered well on the EPIC array.

Other epigenetic studies have pointed out other detrimental effects of high doses of folic acid supplementation on sperm DNA. A study indicated an increased variance in methylation of imprinted genes after supplementing BALB/c mice with a high dose of folic acid (Ly et al., 2017). At a moderate dose, a recent study showed that a 5-fold folic acid supplemented diet led to hyperactivity-like behavior and memory impairment in 3-week-old mice pups (Luan et al., 2022). As well, they found enhanced mRNA degradation in E17.5 cerebrum and expression changes of genes involved in neurotransmission, neural growth and development, and angiogenesis. In addition, they also reported a downregulation of MTHFR and altered choline/methyl metabolism in the liver. A previous study showed that a 10x FASD resulted in a pseudo-MTHFR deficiency in murine liver (Bahous et al., 2017), so it was interesting to see that a 5x FASD also has the capacity of producing the same effect. While the ratio of phosphorylated to non-phosphorylated isoforms of MTHFR was determined to be reduced in CT and TT in the newly developed mice (Reagan et al., 2022), here we did not assess the ratio due to 5x FASD.

Contrary to previous results observed using a high dose of folic acid supplementation, here the data from WGBS at a 10x coverage and ≥5% methylation difference revealed that a moderate dose of folic acid resulted in a bigger effect on

hypermethylation in both genotypes. Additionally, the supplemented diet was able to reduce the total number of DMTs caused by the TT genotype, with 360 total DMTs in the TT CD vs CC CD group and 216. Comparisons of methylation values for TT CD, CC CD, and TT FASD determined that the of folic acid supplementation helped correct 44 out of 49 tiles that were hypermethylated by the TT genotype and 82 out of 84 tiles that were hypomethylated. This was determined by assessing tiles found to be hypomethylated in TT CD vs CC CD that gained methylation, and the opposite as well, meaning tiles found to be hypermethylated that decreased methylation when supplemented with FASD. This suggests that a moderate dose in individuals with the 677C>T polymorphism might partially help alleviate abnormal DNA methylation in sperm. Importantly, however, the results observed here, obtained by WGBS analysis, did not reveal a major effect caused by the TT genotype or the FASD diet. It may be that these mice are not greatly affected by this polymorphism or a postnatal exposure to a moderate dose of FASD. It might be interesting to determine whether a higher dose of supplementation, as well as a prenatal exposure, cause larger effects on the sperm DNA methylome.

Of note, FASD seemed to have caused a tendency for reduced sperm counts in both genotypes, even though it was not statistically significant. Because normal reproductive organ weights were observed, suggesting normal production of sperm, the reason behind the small reduction of sperm counts might be explained by altered chromatin structure caused by the folic acid supplemented diet, which might have caused some sperm heads to be destroyed during the process of sonication and, in turn, not being accounted for. Similarly, in Ly et al., 2017, 20x FASD showed a significant decrease

in sperm counts, whereas in 10x FASD they appeared to be lower but was not statistically significant.

4.4 Comparing control samples analyzed by both Illumina Array and WGBS

Six control samples were analyzed with both technologies, allowing us to compare the results obtained. After performing Pearson correlations for all samples using both datasets, we found that high and low methylated sites correlated well ($R^2 \geq 0.94$), however, sites with intermediate methylation levels displayed more variability. This is reminiscent to a recently published study (Khodasevich et al., 2022) where the authors compared DNA methylation datasets of EPIC microarray and SeqCap Epi Enrichment System (SeqCap) for *PON1* and 9 additional candidate genes. They reported overall high concordance (r= 0.84) between SeqCap and EPIC DNA methylation, among highly methylated and minimally methylated regions, but considerable discrepancy in moderately methylated regions, where SeqCap measurement exhibited greater within-site variation, similar to our findings with WGBS and the Illumina Array.

4.5 Corresponding affected genes from Illumina Array and WGBS data.

4.5.1 Illumina Array

The group with the largest effect (TT CD vs CC CD), revealed certain crucial closest genes annotated in intergenic regions. For instance, with a 10% difference in methylation, the *Aga* gene, annotated closest to an intergenic area, was found to be hypomethylated with -0.10 beta difference. This gene provides instructions for producing aspartylglucosaminidase, an important enzyme active in lysosomes. Disruptions to this gene are known to affect mental functioning and movement in humans (Arvio et al., 2016). Mice homozygous for targeted mutations that inactivate this gene die prematurely and

share most of the clinical, biochemical, and histopathological characteristics of human aspartylglycosaminuria (Danyukova et al., 2018).

Another affected gene closest to an intergenic area was *Rbmy,* with -0.06 beta difference, this gene is located on the Y chromosome and encodes a male germ cell-specific RNA-binding protein associated with spermatogenesis. A study suggested that this gene is a new candidate oncogene specific for male liver cancer (Tsuei et al., 2004). Of interest, the gene *Foxg1* was found annotated closest to an intergenic area and was hypomethylated as well with -0.08 beta difference. This gene provides instructions for making a protein known as forkhead box G1. Changes in this gene cause *FOXG1* syndrome in humans, a rare disorder characterized by impaired development and structural abnormalities in the brain. This gene is also largely expressed in mice, such as in the central nervous system embryo ectoderm, embryo endoderm, hemolymphoid system, and sensory organ. Mouse models with disrupted *Foxg1* have been used to investigate the implications of this gene in neurological disorders (Pringsheim et al., 2019).

4.5.2 WGBS

Genic regions of DMTs were filtered for each different group to take a closer look at genes that were affected. Here, the top 20 most affected genes were listed, the highest difference in methylation being ~20% change and lowest ~8% for the majority of the groups (Table 5). The most affected group, TT CD vs CC CD, displayed interesting affected genes. In the hypermethylated group, the most affected one was *Fgf12*, fibroblast growth factor 12. It acts upstream of or within several processes which include adult locomotor behavior, neuromuscular processes, and positive regulation of sodium ion

transport (Hanada et al., 2018). It is expressed in several structures, including the central nervous system. In humans this gene has been implicated in developmental end epileptic encephalopathy (Verheyen et al., 2020).

Another affected gene was *Zfp608* which is expressed in the cerebral cortex and cortical plate, and it is involved in the negative regulation of transcription by RNA polymerase II (Ayoub et al., 2011). Predominantly expressed in the brain, *Lmo3* is involved in important neurodevelopmental processes in various brain areas including the amygdala (Reisinger et al., 2020). Also expressed in the brain, *Lsamp,* limbic system-associated membrane protein, has been implicated in locomotion, anxiety, fear reaction*,* learning and social behavior and adaptation. Human data also links *LSAMP* gene to several psychiatric disorders (Innos et al., 2013). In the hypomethylated group, *Pcdhb19* is part of a cadherin-superfamily proteins that play key roles in the morphogenesis and function of the brain, including the formation and maintenance of the neuroepithelium, neurite extension, migration of neuronal cells, synaptogenesis, and synaptic plasticity (Hirano et al., 2012). Also of interest, *Snurf* is a *Snrpn* upstream reading frame protein, these are major gene expression regulatory elements (Barbosa et al., 2013). In humans, this gene is located within the Prader-Willi syndrome critical region on chromosome 15 and transcripts from this gene initiate at an imprinting center that are paternally imprinted. Alterations here are associated with parental imprint switch failure, which may cause Angelman syndrome or Prader-Willi syndrome (Cao et al., 2017).

In the CC FASD vs CC CD hypermethylated group *Pde4b* was found to be altered. This gene is involved in several processes which include cellular response to epinephrine stimulus; negative regulation of adenylate cyclase-activating adrenergic receptor

signaling pathway and others. It is expressed in adult cortex, frontal lobe, and other tissues (Campbell et al., 2017). In humans, altered activity of this gene has been associated with schizophrenia and bipolar affective disorder (Millar et al., 2005; Fatemi et al., 2008). Furthermore, in CC FASD vs CC CD hypomethylated, *Ntm*, neurotrimin, is predicted to be involved in the negative regulation of neuron projection development and is expressed in various tissues including the central nervous system (Fearnley et al., 2021). In humans, a deletion of this gene has been associated with Jacobsen syndrome (Sheth et al., 2014). Relatedly, *Nrxn3* is a gene that codes for neurexin and was found altered in the same group. Neurexins are essential presynaptic cell adhesion molecules that are linked to schizophrenia and autism (Aoto et al., 2013).

A comparison of all affected genes, including those closest to an intergenic area, was done between groups following the same direction (hyper- or hypomethylated). When examining the effect of folic acid supplementation on the TT genotype, that is, when searching for commonalities between the groups TT FASD vs CC CD & TT CD vs CC CD, *Lmo3* remained affected (Table 6). In the hypomethylated group, *Ntm* was also shared, along with *Grip1os2,* a glutamate receptor interacting protein. Lastly, when comparing the effects of folic acid supplementation on both genotypes, that is TT FASD vs CC CD & CC FASD vs CC CD, *Ufm1* and *Slc35b3* were shared in the hypermethylated group as genes located closest to intergenic regions. Other comparisons were performed but no commonalities were found.

CHAPTER 5: CONCLUSIONS AND FUTURE DIRECTIONS

Overall, the general and reproductive health of the newly developed *Mthfr* 677C>T mice was unaffected by either the TT genotype or folic acid diets. Out of all groups studied, the one that consistently showed the largest effect was TT CD vs CC CD, both in the Illumina Array and WGBS datasets. This group showed both hyper- and hypomethylation, with a more prominent effect of hypomethylation, contrary to what was observed in previous studies done with RRBS. Moreover, the folic acid deficient diet did not exacerbate the number of DMCs found in the TT genotype, as TT FD vs TT CD had fewer DMCs than TT CD vs CC CD. Lastly, a moderate dose of folic acid supplementation displayed hyper- and hypomethylating effects, with a predominance for hypermethylation on both genotypes, contrary to what has been observed before with high doses of folic acid supplementation. As well, the moderate dose used here, decreased the number of tiles seen in TT genotype on a control diet, particularly hypomethylated. TT FASD vs CC CD showed the least amount of DMTs, in contrast to TT CD vs CC CD. The supplementation was also able to correct affected tiles that were differentially methylated in the TT CD vs CC CD group.

This project is the first one to characterize the sperm DNA epigenome of the newly developed *Mthfr 677*C>T mice, employing a recently developed Array and WGBS. Moreover, by the time this study started, it was the first project to use employ the Illumina Mouse Methylation BeadChip on mouse sperm DNA. In here we have performed the baseline characterization of these new mice allowing the possibility to investigate clinical issues related to infertility, such as advanced paternal age and other diets.

5.1 Future directions

After performing a comprehensive characterization of the sperm DNA epigenome of the *Mthfr* 677C>T mice, other interesting questions arise. The Illumina Array offered a practical genome-wide coverage of the samples analyzed; however, it would also be advantageous to investigate the effects of folate deficiency using WGBS and determine if any larger-scale effects can be found. This analysis would also help further ascertain if folate deficiency does not, in fact, deteriorate the effects caused by the TT genotype.

In addition, as follow-up studies, it would be interesting to repeat the above experiments in BALB/c mice, as it has been shown that these mice are more susceptible to MTHFR deficiency and different intakes of folic acid (Chan et al., 2010; Ly et al, 2017; Aarabi et al., 2018 Chan et al., 2022 [submitted in Andrology]).

Likewise, additional research should focus on the effect of the *Mthfr* 677C>T genotype and different folic acid diets on the offspring of these mice. That is because genetic and environmental exposures, like the ones studied in this project, can impact the epigenome, such as DNA methylation, and have the potential of leading to heritable alterations in gene expression. These types of abnormalities might impact major epigenomic reprogramming events, such as during embryo development, when DNA methylation and histone methylation are being actively removed, with the exception of imprinted genes, repeat sequences and regions involved in the epigenetic inheritance of disease.

Lastly, as a faithful representation of the human *MTHFR* 677>T polymorphism, this mouse model could be used to investigate other pathologies, such as the numerous neurological disorders listed before, vascular disorders as well as other diseases. Other

epigenetic marks, such as histone modifications would provide additional useful information.

References

Aapola, U., Lyle, R., Krohn, K., Antonarakis, S. E., & Peterson, P. (2001). Isolation and initial characterization of the mouse *Dnmt3l* gene. Cytogenetics and Cell Genetics, 92(1-2), 122–6.

Aarabi, M., Christensen, K. E., Chan, D., Leclerc, D., Landry Mylène, Ly, L., Rozen, R., & Trasler, J. (2018). Testicular MTHFR deficiency may explain sperm DNA hypomethylation associated with high dose folic acid supplementation. Human Molecular Genetics, 27(7), 1123–1135. https://doi.org/10.1093/hmg/ddy021

Aarabi, M., San Gabriel, M.C., Chan, D., Behan, N.A., Caron, M., Pastinen, T., Bourque, G., MacFarlane, A.J., Zini, A. and Trasler, J. (2015) High-dose folic acid supplementation alters the human sperm methylome and is influenced by the *MTHFR* C677T polymorphism. Hum. Mol. Genet., 24, 6301–6313.

Adler, I.-D. (1996). Comparison of the duration of spermatogenesis between male rodents and humans. *Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis*, *352*(1), 169–172. https://doi.org/10.1016/0027-5107(95)00223-5

Agarwal, A., Baskaran, S., Parekh, N., Cho, C.-L., Henkel, R., Vij, S., Arafa, M., Panner Selvam, M. K., & Shah, R. (2021). Male infertility. *Lancet (London, England)*, *397*(10271), 319–333. https://doi.org/10.1016/S0140-6736(20)32667-2

Akalin, A., Kormaksson, M., Li, S., Garrett-Bakelman, F. E., Figueroa, M. E., Melnick, A., & Mason, C. E. (2012). Methylkit: a comprehensive r package for the analysis of genome-wide DNA methylation profiles. Genome Biology, 13(10), 87. https://doi.org/10.1186/gb-2012-13-10-r87

Allen, M. J., Lee, C., Lee, J. D., Pogany, G. C., Balooch, M., Siekhaus, W. J., & Balhorn, R. (1993). Atomic force microscopy of mammalian sperm chromatin. Chromosoma, 102(9), 623–630. https://doi.org/10.1007/BF00352310

Aoshima, K., Inoue, E., Sawa, H., & Okada, Y. (2015). Paternal h3k4 methylation is required for minor zygotic gene activation and early mouse embryonic development. Embo Reports, 16(7), 803–812. https://doi.org/10.15252/embr.201439700

Aoto, J., Martinelli, D. C., Malenka, R. C., Tabuchi, K., & Südhof, T. C. (2013). Presynaptic neurexin-3 alternative splicing trans-synaptically controls postsynaptic ampa receptor trafficking. Cell, 154(1), 75–88. https://doi.org/10.1016/j.cell.2013.05.060

Arvio, M., & Mononen, I. (2016). Aspartylglycosaminuria: a review. *Orphanet Journal of Rare Diseases*, *11*(1), 162–162.

Ayoub, A. E., Oh, S., Xie, Y., Leng, J., Cotney, J., Dominguez, M. H., Noonan, J. P., & Rakic, P. (2011). Transcriptional programs in transient embryonic zones of the cerebral cortex defined by high-resolution mRNA sequencing. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(36), 14950–14955.

Bahous, R. H., Jadavji, N. M., Deng, L., Cosín-Tomás, M., Lu, J., Malysheva, O., Leung, K.-Y., Ho, M.-K., Pallàs, M., Kaliman, P., Greene, N. D. E., Bedell, B. J., Caudill, M. A., & Rozen, R. (2017). High dietary folate in pregnant mice leads to pseudo-MTHFR deficiency and altered methyl metabolism, with embryonic growth delay and short-term memory impairment in offspring. *Human Molecular Genetics*, *26*(5), 888–900. https://doi.org/10.1093/hmg/ddx004

Bannister, A. J., & Kouzarides, T. (2011). Regulation of chromatin by histone modifications. Cell Research, 21(3), 381–95. https://doi.org/10.1038/cr.2011.22

Barbosa, C., Peixeiro, I., & amp; Romao, L. (2013). Gene expression regulation by upstream open reading frames and human disease. Plos Genetics, 9(8). https://doi.org/10.1371/journal.pgen.1003529

Bellvé, A. R., Millette, C. F., Bhatnagar, Y. M., & O'Brien, D. A. (1977). Dissociation of the mouse testis and characterization of isolated spermatogenic cells. The Journal of Histochemistry and Cytochemistry: Official Journal of the Histochemistry Society, 25(7), 480–94.

Bestor, T. H. (2000). The DNA methyltransferases of mammals. *Human Molecular Genetics*, *9*(16), 2395–402.

Brykczynska, U., Hisano, M., Erkek, S., Ramos, L., Oakeley, E. J., Roloff, T. C., Beisel, C., Schübeler, D., Stadler, M. B., & Peters, A. H. F. M. (2010). Repressive and active histone methylation mark distinct promoters in human and mouse

spermatozoa. Nature Structural & Molecular Biology, 17(6), 679–87. https://doi.org/10.1038/nsmb.1821

Cao, Y., AlHumaidi, S. S., Faqeih, E. A., Pitel, B. A., Lundquist, P., & Aypar, U. (2017). A novel deletion of *Snurf/Snrpn* exon 1 in a patient with Prader-Willi-like phenotype. European Journal of Medical Genetics, 60(8), 416–420. https://doi.org/10.1016/j.ejmg.2017.05.003

Campbell, S. L., Groen, T., Kadish, I., Smoot, L. H. M., & Bolger, G. B. (2017). Altered phosphorylation, electrophysiology, and behavior on attenuation of *Pde4b* action in hippocampus. Bmc Neuroscience, 18(1), 1–20. https://doi.org/10.1186/s12868-017-0396-6

Chan, D., Cushnie, D. W., Neaga, O. R., Lawrance, A. K., Rozen, R., & Trasler, J. M. (2010). Strain-specific defects in testicular development and sperm epigenetic patterns in 5,10-methylenetetrahydrofolate reductase-deficient mice. Endocrinology, 151(7), 3363–73. https://doi.org/10.1210/en.2009-1340

Chan, D., Delbès, G., Landry, M., Robaire, B., & Trasler, J. M. (2012). Epigenetic alterations in sperm DNA associated with testicular cancer treatment. *Toxicological Sciences: An Official Journal of the Society of Toxicology*, *125*(2), 532–43. https://doi.org/10.1093/toxsci/kfr307

Chan, D., Ly, L., Martinez Duncker Rebolledo, E., Martel, J., Mylene L., Scott-Boyer M., Droit, A & Trasler, J. (2022). Transgenerational impact of grand-paternal lifetime exposures to both folic acid deficiency and supplementation on genome-wide DNA methylation in male germ cells. *Andrology*, 1–36.

Chan, D., Shao, X., Dumargne, M.-C., Aarabi, M., Simon, M.-M., Kwan, T., Bailey, J. L., Robaire, B., Kimmins, S., San Gabriel, M. C., Zini, A., Librach, C., Moskovtsev, S., Grundberg, E., Bourque, G., Pastinen, T., & Trasler, J. M. (2019). Customized methylc-capture sequencing to evaluate variation in the human sperm dna methylome representative of altered folate metabolism. Environmental Health Perspectives, 127(8), 087002–087002. https://doi.org/10.1289/EHP4812

Chen, Z., Karaplis, A. C., Ackerman, S. L., Pogribny, I. P., Melnyk, S., Lussier-Cacan, S., Chen, M. F., Pai, A., John, S. W., Smith, R. S., Bottiglieri, T., Bagley, P., Selhub, J., Rudnicki, M. A., James, S. J., & Rozen, R. (2001). Mice deficient in

methylenetetrahydrofolate reductase exhibit hyperhomocysteinemia and decreased methylation capacity, with neuropathology and aortic lipid deposition. *Human Molecular Genetics*, *10*(5), 433–43.

Christensen KE, Mikael LG, Leung KY, Lévesque N, Deng L, Wu Q, Malysheva OV, Best A, Caudill MA, Greene ND, Rozen R (2015). High folic acid consumption leads to pseudo-MTHFR deficiency, altered lipid metabolism, and liver injury in mice. Am J Clin Nutr. Mar;101(3):646-58. doi: 10.3945/ajcn.114.086603. Epub 2015 Jan 7. PMID: 25733650; PMCID: PMC4340065.

Clare, C. E., Brassington, A. H., Kwong, W. Y., & Sinclair, K. D. (2019). One-carbon metabolism: linking nutritional biochemistry to epigenetic programming of long-term development. *Annual Review of Animal Biosciences*, *7*(1), 263–287. https://doi.org/10.1146/annurev-animal-020518-115206

Clermont, Y. (1972). Kinetics of spermatogenesis in mammals: seminiferous epithelium cycle and spermatogonial renewal. *Physiological Reviews*, *52*(1), 198–236.

Culty, M. (2009). Gonocytes, the forgotten cells of the germ cell lineage. *Birth Defects Research Part C: Embryo Today: Reviews*, *87*(1), 1–26. https://doi.org/10.1002/bdrc.20142

Danyukova, T., Ariunbat, K., Thelen, M., Brocke-Ahmadinejad, N., Mole, S. E., & Storch, S. (2018). Loss of *Cln7* results in depletion of soluble lysosomal proteins and impaired motor reactivation. *Human Molecular Genetics*, *27*(10), 1711–1722. https://doi.org/10.1093/hmg/ddy076

Davis, T. L., Yang, G. J., McCarrey, J. R., & Bartolomei, M. S. (2000). The h19 methylation imprint is erased and re-established differentially on the parental alleles during male germ cell development. *Human Molecular Genetics*, *9*(19), 2885–94.

Dawlaty, M. M., Breiling, A., Le, T., Raddatz, G., Barrasa, M. I., Cheng, A. W., Gao, Q., Powell, B. E., Li, Z., Xu, M., Faull, K. F., Lyko, F., & Jaenisch, R. (2013). Combined deficiency of tet1 and tet2 causes epigenetic abnormalities but is compatible with postnatal development. Developmental Cell, 24(3), 310–23. https://doi.org/10.1016/j.devcel.2012.12.015

Deaton Aimée M, & Bird, A. (2011). CpG islands and the regulation of transcription. Genes & Development, 25(10), 1010–1022. https://doi.org/10.1101/gad.2037511

de Waal, E., Mak, W., Calhoun, S., Stein, P., Ord, T., Krapp, C., et al. (2014). In vitro culture increases the frequency of stochastic epigenetic errors at imprinted genes in placental tissues from mouse concepti produced through assisted reproductive technologies. *Biology of reproduction*, *90*(2).

Dejeux, E., El abdalaoui, H., Gut, I. G., & Tost, J. (2009). Identification and quantification of differentially methylated loci by the pyrosequencing technology. Methods in Molecular Biology (Clifton, N.j.), 507, 189–205. https://doi.org/10.1007/978-1-59745-522-0_15

Denomme, M. M., McCallie, B. R., Parks, J. C., Schoolcraft, W. B., & Katz-Jaffe, M. G. (2017). Alterations in the sperm histone-retained epigenome are associated with unexplained male factor infertility and poor blastocyst development in donor oocyte IVF cycles. Human Reproduction (Oxford, England), 32(12), 2443–2455. https://doi.org/10.1093/humrep/dex317

Ducker, G. S., & Rabinowitz, J. D. (2017). One-carbon metabolism in health and disease. *Cell Metabolism*, *25*(1), 27–42. https://doi.org/10.1016/j.cmet.2016.08.009

Ebisch, I. M. W., Thomas, C. M. G., Peters, W. H. M., Braat, D. D. M., & Steegers-Theunissen, R. P. M. (2007). The importance of folate, zinc and antioxidants in the pathogenesis and prevention of subfertility. Human Reproduction Update, 13(2), 163–74.

Edwards, J., Yarychkivska, O., Boulard, M., & Bestor, T. (2017). DNA methylation and DNA methyltransferases. *Epigenetics & Chromatin*, *10*(1), 1–10.

Erkek, S., Hisano, M., Liang, C.-Y., Gill, M., Murr, R., Dieker, J., Schübeler, D., van der Vlag, J., Stadler, M. B., & Peters, A. H. F. M. (2013). Molecular determinants of nucleosome retention at CpG-rich sequences in mouse spermatozoa. Nature Structural & Molecular Biology, 20(7), 868–75. https://doi.org/10.1038/nsmb.2599

Fatemi, S. H., King, D. P., Reutiman, T. J., Folsom, T. D., Laurence, J. A., Lee, S., Fan, Y.-T., Paciga, S. A., Conti, M., & Menniti, F. S. (2008). Pde4b polymorphisms and decreased pde4b expression are associated with

schizophrenia. Schizophrenia Research, 101(1-3), 36–49. https://doi.org/10.1016/j.schres.2008.01.029

Fearnley, S., Raja, R., & Cloutier, J.-F. (2021). Spatiotemporal expression of iglon family members in the developing mouse nervous system. Scientific Reports, 11(1). https://doi.org/10.1038/s41598-021-97768-5

Friso, S., Choi, S.-W., Girelli, D., Mason, J. B., Dolnikowski, G. G., Bagley, P. J., Olivieri, O., Jacques, P. F., Rosenberg, I. H., Corrocher, R., & Selhub, J. (2002). A common mutation in the 5, 10-methylenetetra-hydrofolate reductase gene affects genomic DNA methylation through an interaction with folate status. Proceedings of the National Academy of Sciences of the United States of America, 99(8), 5606–5611.

Frosst, P., Blom, H. J., Milos, R., Goyette, P., Sheppard, C. A., Matthews, R. G., Boers, G. J., den, H. M., Kluijtmans, L. A., van, den H. L. P., & et, al. (1995). A candidate genetic risk factor for vascular disease: a common mutation in methylenetetrahydrofolate reductase. Nature Genetics, 10(1), 111–3.

Garcia-Prieto, C. A., Álvarez-Errico, D., Musulen, E., Bueno-Costa, A., N Vazquez, B., Vaquero, A., & Esteller, M. (2022). Validation of a DNA methylation microarray for 285,000 CpG sites in the mouse genome. *Epigenetics*, *1-9*, 1–9. https://doi.org/10.1080/15592294.2022.2053816

Garner, J. L., Niles, K. M., McGraw, S., Yeh, J. R., Cushnie, D. W., Hermo, L., Nagano, M. C., & Trasler, J. M. (2013). Stability of DNA methylation patterns in mouse spermatogonia under conditions of MTHFR deficiency and methionine supplementation. Biology of Reproduction, 89(5), 125–125. https://doi.org/10.1095/biolreprod.113.109066

Gershman, A. (2022). Epigenetic patterns in a complete human genome. Science, 6588, 58–58.

Glaser, J., Iranzo, J., Borensztein, M., Marinucci, M., Gualtieri, A., Jouhanneau, C., Teissandier, A., Gaston-Massuet, C., & Bourc'his, D. (2022). The imprinted zdbf2 gene finely tunes control of feeding and growth in neonates. Elife, 11. https://doi.org/10.7554/eLife.65641

Goll, M. G., & Bestor, T. H. (2005). Eukaryotic cytosine methyltransferases. Annual Review of Biochemistry, 74, 481–514.

Gong M, Dong W, He T, Shi Z, Huang G, Ren R, Huang S, Qiu S, Yuan R. *MTHFR* 677C>T polymorphism increases the male infertility risk: a meta-analysis involving 26 studies. PLoS One. 2015 Mar 20;10(3): e0121147. doi: 10.1371/journal.pone.0121147. PMID: 25793386; PMCID: PMC4368707.

Greenberg, M. V. C., & Bourc'his, D. (2019). The diverse roles of DNA methylation in mammalian development and disease. Nature Reviews. Molecular Cell Biology, 20(10), 590–607. https://doi.org/10.1038/s41580-019-0159-6

Haaf, T., & Ward, D. C. (1995). Higher order nuclear structure in mammalian sperm revealed by in situ hybridization and extended chromatin fibers. Experimental Cell Research, 219(2), 604–11.

Hammoud, S. S., Nix, D. A., Hammoud, A. O., Gibson, M., Cairns, B. R., & Carrell, D. T. (2011). Genome-wide analysis identifies changes in histone retention and epigenetic modifications at developmental and imprinted gene loci in the sperm of infertile men. Human Reproduction (Oxford, England), 26(9), 2558–69. https://doi.org/10.1093/humrep/der192

Hammoud, S., Liu, L., & Carrell, D. T. (2009). Protamine ratio and the level of histone retention in sperm selected from a density gradient preparation. Andrologia, 41(2), 88–94. https://doi.org/10.1111/j.1439-0272.2008.00890.x

Hanada, Y., Nakamura, Y., Ozono, Y. *et al.* Fibroblast growth factor 12 is expressed in spiral and vestibular ganglia and necessary for auditory and equilibrium function. *Sci Rep* 8, 11491 (2018). https://doi.org/10.1038/s41598-018-28618-0

Hayashi, K., Ohta, H., Kurimoto, K., Aramaki, S., & Saitou, M. (2011). Reconstitution of the mouse germ cell specification pathway in culture by pluripotent stem cells. *Cell*, *146*(4), 519–32. https://doi.org/10.1016/j.cell.2011.06.052

Hermann, A., Gowher, H., & Jeltsch, A. (2004). Biochemistry and biology of mammalian DNA methyltransferases. *Cellular and Molecular Life Sciences Cmls*, *61*(19-20), 2571–2587. https://doi.org/10.1007/s00018-004-4201-1

Hess, R.A., de Franca, L.R. (2009). Spermatogenesis and Cycle of the Seminiferous Epithelium. In: Cheng, C.Y. (eds) Molecular Mechanisms in Spermatogenesis. Advances in Experimental Medicine and Biology, vol 636. Springer, New York, NY. https://doi-org.proxy3.library.mcgill.ca/10.1007/978-0-387-09597-4_1

Hirano, K., Kaneko, R., Izawa, T., Kawaguchi, M., Kitsukawa, T., & Yagi, T. (2012). Single-neuron diversity generated by protocadherin-β cluster in mouse central and peripheral nervous systems. Frontiers in Molecular Neuroscience, 5. https://doi.org/10.3389/fnmol.2012.00090

Innos, J., Leidmaa, E., Philips, M.-A., Sütt, S., Alttoa, A., Harro, J., Kõks, S., & Vasar, E. (2013). *lsamp* -/- mice display lower sensitivity to amphetamine and have elevated 5-ht turnover. *Biochemical and Biophysical Research Communications*, *430*(1), 413–418. https://doi.org/10.1016/j.bbrc.2012.11.077

International, H.G.S.C. (2001). Initial sequencing and analysis of the human genome. *Nature*, *409*, 860–921. https://doi.org/10.1038/35057062

Ito, S., Shen, L., Dai, Q., Wu, S. C., Collins, L. B., Swenberg, J. A., He, C., & Zhang, Y. (2011). Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science*, *333*(6047), 1300–1303

Jenkins, T., Aston, K., Carrell, D., DeVilbiss, E., Sjaarda, L., Perkins, N., Mills, J. L., Chen, Z., Sparks, A., Clemons, T., Chaney, K., Peterson, C. M., Emery, B., Hotaling, J., Johnstone, E., Schisterman, E., & Mumford, S. L. (2022). The impact of zinc and folic acid supplementation on sperm DNA methylation: results from the folic acid and zinc supplementation randomized clinical trial (fazst). Fertility and Sterility, 117(1), 75–85. https://doi.org/10.1016/j.fertnstert.2021.09.009

Kang, S. S., Wong, P. W., Susmano, A., Sora, J., Norusis, M., & Ruggie, N. (1991). Thermolabile methylenetetrahydrofolate reductase: an inherited risk factor for coronary artery disease. American Journal of Human Genetics, 48(3), 536–545.

Karahan, G., Chan, D., Shirane, K., McClatchie, T., Janssen, S., Baltz, J. M., Lorincz, M., & Trasler, J. (2021). Paternal MTHFR deficiency leads to hypomethylation of young retrotransposons and reproductive decline across two successive generations. Development (Cambridge, England), 148(13). https://doi.org/10.1242/dev.199492

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

Kebede, A. F., Schneider, R., & Daujat, S. (2015). Novel types and sites of histone modifications emerge as players in the transcriptional regulation contest. Febs Journal, 282(9), 1658–1674. https://doi.org/10.1111/febs.13047

Kelly, T. L. J., Li, E., & Trasler, J. M. (2003). 5-aza-2′-deoxycytidine induces alterations in murine spermatogenesis and pregnancy outcome. *Journal of Andrology*, *24*(6), 822–830. https://doi.org/10.1002/j.1939-4640.2003.tb03133.x

Khodasevich, D., Smith, A. R., Huen, K., Eskenazi, B., Cardenas, A., Holland, N. (2022). Comparison of DNA methylation measurements from epic beadchip and seqcap targeted bisulphite sequencing in pon1 and nine additional candidate genes. Epigenetics, 1-12, 1–12. https://doi.org/10.1080/15592294.2022.2091818

Kobayashi, H., Sato, A., Otsu, E., Hiura, H., Tomatsu, C., Utsunomiya, T., Sasaki, H., Yaegashi, N., & Arima, T. (2007). Aberrant DNA methylation of imprinted loci in sperm from oligospermic patients. *Human Molecular Genetics*, *16*(21), 2542–51.

Kouzarides, T. (2007). Chromatin modifications and their function. Cell, 128(4), 693–705.

Kriaucionis, S., & Heintz, N. (2009). The nuclear DNA base 5-hydroxymethylcytosine is present in purkinje neurons and the brain. *Science*, *324*(5929), 929–930.

Kubo, N., Toh, H., Shirane, K. et al. DNA methylation and gene expression dynamics during spermatogonial stem cell differentiation in the early postnatal mouse testis. BMC Genomics **16,** 624 (2015). https://doi.org/10.1186/s12864-015-1833-5

La Salle, S., & Trasler, J. M. (2006). Dynamic expression of DNMT3a and DNMT3b isoforms during male germ cell development in the mouse. *Developmental Biology*, *296*(1), 71–82. https://doi.org/10.1016/j.ydbio.2006.04.436

Lambrot, R., Xu, C., Saint-Phar, S., Chountalos, G., Cohen, T., Paquet, M., Suderman, M., Hallett, M., & Kimmins, S. (2013). Low paternal dietary folate alters the mouse sperm epigenome and is associated with negative pregnancy outcomes. Nature Communications, 4(1). https://doi.org/10.1038/ncomms3889

Lawrence, M., Daujat, S., & Schneider, R. (2016). Lateral thinking: how histone modifications regulate gene expression. Trends in Genetics, 32(1), 42–56. https://doi.org/10.1016/j.tig.2015.10.007

Lee, H.-C., Jeong, Y.-M., Lee, S. H., Cha, K. Y., Song, S.-H., Kim, N. K., Lee, K. W., & Lee, S. (2006). Association study of four polymorphisms in three folate-related enzyme genes with non-obstructive male infertility. *Human Reproduction (Oxford, England)*, *21*(12), 3162–70.

Lesch, B. J., & Page, D. C. (2014). Poised chromatin in the mammalian germ line. Development (Cambridge, England), 141(19), 3619–26. https://doi.org/10.1242/dev.113027

Lesch, B. J., Silber, S. J., McCarrey, J. R., & Page, D. C. (2016). Parallel evolution of male germline epigenetic poising and somatic development in animals. Nature Genetics, 48(8), 888–94. https://doi.org/10.1038/ng.3591

Liao, H.-F., Tai, K.-Y., Chen, W. S.-C., Cheng, L. C. W., Ho, H.-N., & Lin, S.-P. (2012). Functions of DNA methyltransferase 3-like in germ cells and beyond. *Biology of the Cell*, *104*(10), 571–587. https://doi.org/10.1111/boc.201100109

Liew, S.-C., & Gupta, E. D. (2015). Methylenetetrahydrofolate reductase (*MTHFR*) C677T polymorphism: epidemiology, metabolism, and the associated diseases. European Journal of Medical Genetics, 58(1), 1–10. https://doi.org/10.1016/j.ejmg.2014.10.004

Lismer, A., Dumeaux, V., Lafleur, C., Lambrot, R., Brind'Amour, J., Lorincz, M. C., & Kimmins, S. (2021). Histone h3 lysine 4 trimethylation in sperm is transmitted to the embryo and associated with diet-induced phenotypes in the offspring. Developmental Cell, 56(5), 671–686. https://doi.org/10.1016/j.devcel.2021.01.014

Lismer, A., Siklenka, K., Lafleur, C., Dumeaux, V., & Kimmins, S. (2020). Sperm histone h3 lysine 4 trimethylation is altered in a genetic mouse model of transgenerational epigenetic inheritance. Nucleic Acids Research, 48(20), 11380–11393. https://doi.org/10.1093/nar/gkaa712

Liu, Y., Zhang, Y., Yin, J., Gao, Y., Li, Y., Bai, D., He, W., Li, X., Zhang, P., Li, R., Zhang, L., Jia, Y., Zhang, Y., Lin, J., Zheng, Y., Wang, H., Gao, S., Zeng, W., & Liu, W. (2019). Distinct h3k9me3 and DNA methylation modifications during mouse spermatogenesis. Journal of Biological Chemistry, 294(49), 18714–18725. https://doi.org/10.1074/jbc.RA119.010496

Luan, Y., Cosín-Tomás, M., Leclerc, D., Malysheva, O. V., Caudill, M. A., & Rozen, R. (2022). Moderate folic acid supplementation in pregnant mice results in altered sex-specific gene expression in brain of young mice and embryos. Nutrients, 14(5). https://doi.org/10.3390/nu14051051

Ly, L., Chan, D., Aarabi, M., Landry, M., Behan, N. A., MacFarlane, A. J., & Trasler, J. (2017). Intergenerational impact of paternal lifetime exposures to both folic acid deficiency and supplementation on reproductive outcomes and imprinted gene methylation. Molecular Human Reproduction, 23(7), 461–477. https://doi.org/10.1093/molehr/gax029 Lyon, M., (1961). Gene action in the x-chromosome of the mouse (mus musculus l.). Nature, 190(4773), 372–373. https://doi.org/10.1038/190372a0

Marushige, Y., & Marushige, K. (1975). Transformation of sperm histone during formation and maturation of rat spermatozoa. The Journal of Biological Chemistry, 250(1), 39–45.

Meissner, A., Gnirke, A., Bell, G. W., Ramsahoye, B., Lander, E. S., & Jaenisch, R. (2005). Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Research*, *33*(18), 5868–77.

Meissner, A., Mikkelsen, T. S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B. E., Nusbaum, C., Jaffe, D. B., Gnirke, A., Jaenisch, R., & Lander, E. S. (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature, 454(7205), 766.

Millar, J. K., Douglas, H. R. B., Pickard, B. S., Mackie, S., James, R., Christie, S., Buchanan, S. R., Malloy, M. P., Chubb, J. E., Huston, E., Baillie, G. S., Thomson, P. A., Hill, E. V., Brandon, N. J., Rain, J.-C., Camargo, L. M., Whiting, P. J., Houslay, M. D., Muir, W. J., & Porteous, D. J. (2005). *Disc1* and *Pde4b* are interacting genetic factors in schizophrenia that regulate camp signaling. *Science*, *310*(5751), 1187–1191.

Miousse, I. R., Tobacyk, J., Melnyk, S., James, S. J., Cheema, A. K., Boerma, M., Hauer-Jensen, M., & Koturbash, I. (n.d.). One-carbon metabolism and ionizing radiation: a multifaceted interaction. *Biomolecular Concepts*, *8*(2), 83–92. https://doi.org/10.1515/bmc-2017-0003

Molaro, A., Falciatori, I., Hodges, E., Aravin, A. A., Marran, K., Rafii, S., McCombie, W. R., Smith, A. D., & Hannon, G. J. (2014). Two waves of de novo methylation during mouse germ cell development. *Genes & Development*, *28*(14), 1544–1549. https://doi.org/10.1101/gad.244350.114

Molyneaux, K. A., Stallock, J., Schaible, K., & Wylie, C. (2001). Time-lapse analysis of living mouse germ cell migration. *Developmental Biology*, *240*(2), 488–498. https://doi.org/10.1006/dbio.2001.0436

Morselli, M., Pastor, W. A., Montanini, B., Nee, K., Ferrari, R., Fu, K., Bonora, G., Rubbi, L., Clark, A. T., Ottonello, S., Jacobsen, S. E., & Pellegrini, M. (2015). In vivo targeting of de novo DNA methylation by histone modifications in yeast and mouse. *Elife*, *4*, 06205. https://doi.org/10.7554/eLife.06205

Mruk, D. D., & Cheng, C. Y. (2004). Sertoli-sertoli and sertoli-germ cell interactions and their significance in germ cell movement in the seminiferous epithelium during spermatogenesis. *Endocrine Reviews*, *25*(5), 747–806.

Niles, K. M., Yeh, J. R., Chan, D., Landry, M., Nagano, M. C., & Trasler, J. M. (2013). Haploinsufficiency of the paternal-effect gene *Dnmt3l* results in transient DNA hypomethylation in progenitor cells of the male germline. *Human Reproduction (Oxford, England)*, *28*(2), 519–30. https://doi.org/10.1093/humrep/des395

Ohno, R., Nakayama, M., Naruse, C., Okashita, N., Takano, O., Tachibana, M., Asano, M., Saitou, M., & Seki, Y. (2013). A replication-dependent passive mechanism modulates DNA demethylation in mouse primordial germ cells. Development (Cambridge, England), 140(14), 2892–903. https://doi.org/10.1242/dev.093229

Okano, M., Xie, S., & Li, E. (1998). Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. Nature Genetics, 19(3), 219–20.

Paulsen, M., & Ferguson-Smith, A. C. (2001). DNA methylation in genomic imprinting, development, and disease. *The Journal of Pathology*, *195*(1), 97–110. https://doi.org/10.1002/path.890

Popp, C., Dean, W., Feng, S., Cokus, S. J., Andrews, S., Pellegrini, M., Jacobsen, S. E., & Reik, W. (2010). Genome-wide erasure of DNA methylation in mouse

primordial germ cells is affected by aid deficiency. Nature, 463(7284), 1101–1105. https://doi.org/10.1038/nature08829

Pringsheim, M., Mitter, D., Schröder, S., Warthemann, R., Plümacher, K., Kluger, G., Baethmann, M., Bast, T., Braun, S., Büttel, H.-M., Conover, E., Courage, C., Datta, A. N., Eger, A., Grebe, T. A., Hasse-Wittmer, A., Heruth, M., Höft, K., Kaindl, A. M., … Brockmann, K. (2019). Structural brain anomalies in patients with FOXG1 syndrome and in *Foxg1* +/- mice. *Annals of Clinical and Translational Neurology*, *6*(4), 655–668. https://doi.org/10.1002/acn3.735

Quinlan, A. R., & Hall, I. M. (2010). Bedtools: a flexible suite of utilities for comparing genomic features. Bioinformatics, 26(6), 841–842. https://doi.org/10.1093/bioinformatics/btq033

Reagan, A., Christensen, K., Graham, L., Bedwell, A., Eldridge, K., Speedy, R., Figueiredo, L., Persohn, S., Bottiglieri, T., Sasner, M., Territo, P., Rozen, R., Howell, G. (2022). The 677C>T variant in methylenetetrahydrofolate reductase causes morphological and functional cerebrovascular deficits in mice. Accepted in Journal of Cerebral Blood Flow & Metabolism; doi: https://doi.org/10.1101/2021.12.16.472805

Reik, W., Dean, W., & Walter Jörn. (2001). Epigenetic reprogramming in mammalian development. *Science*, *293*(5532), 1089–1093

Reisinger, S. N., Bilban, M., Stojanovic, T., Derdak, S., Yang, J., Cicvaric, A., Horvath, O., Sideromenos, S., Zambon, A., Monje, F. J., Boehm, S., & Pollak, D. D. (2020). *Lmo3* deficiency in the mouse is associated with alterations in mood-related behaviors and a depression-biased amygdala transcriptome. *Psychoneuroendocrinology*, *111*.

Riggs, A. D. (1975). X inactivation, differentiation, and DNA methylation. *Cytogenetics and Cell Genetics*, *14*(1), 9–25.

Saffroy, R., Benyamina, A., Pham, P., Marill, C., Karila, L., Reffas, M., Debuire, B., Reynaud, M., & Lemoine, A. (2008). Protective effect against alcohol dependence of the thermolabile variant of *MTHFR*. Drug and Alcohol Dependence, 96(1-2), 30–6. https://doi.org/10.1016/j.drugalcdep.2008.01.016

Saitou, M., Barton, S. C., & Surani, M. A. (2002). A molecular programme for the specification of germ cell fate in mice. *Nature*, *418*(6895), 293–300.

Santos-Rebouças C B, & Pimentel, M. M. G. (2007). Implication of abnormal epigenetic patterns for human diseases. *European Journal of Human Genetics*, *15*(1), 10–17.

Sasaki, H., & Matsui, Y. (2008). Epigenetic events in mammalian germ-cell development: reprogramming and beyond. Nature Reviews. Genetics, 9(2), 129–40. https://doi.org/10.1038/nrg2295

Seisenberger, S., Andrews, S., Krueger, F., Arand, J., Walter, J., Santos, F., Popp, C., Thienpont, B., Dean, W., & Reik, W. (2012). The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells. Molecular Cell, 48(6), 849–62. https://doi.org/10.1016/j.molcel.2012.11.001

Sheth, F., Nayak, D., Ghambhir, P., Sheth, J., Datar, C., Pandit, A., Desai, M., & Patel, B. (2014). Partial deletion of distal long arm encompassing Jacobsen syndrome. Molecular Cytogenetics, 7, 1–1. https://doi.org/10.1186/1755-8166-7-S1-P47

Shirane, K. (2022). The dynamic chromatin landscape and mechanisms of DNA methylation during mouse germ cell development. Genes & Genetic Systems, 97(1), 3–14. https://doi.org/10.1266/ggs.21-00069

Siklenka, K., Erkek, S., Godmann, M., Lambrot, R., McGraw, S., Lafleur, C., Cohen, T., Xia, J., Suderman, M., Hallett, M., Trasler, J., Peters, A. H. F. M., & Kimmins, S. (2015). Disruption of histone methylation in developing sperm impairs offspring health transgenerationally. Science, 350(6261), 651–651.

Strogantsev, R., Yamazawa, K., Shi, H., Gould, P., Goldman-Roberts, M., McEwen, K., Ferguson-Smith, A. C., Krueger, F., Sun, B., & Pedersen, R. (2015). Allele-specific binding of *Zfp57* in the epigenetic regulation of imprinted and non-imprinted monoallelic expression. Genome Biology, 16(1).

Susiarjo, M., Sasson, I., Mesaros, C., & Bartolomei, M. S. (2013). Bisphenol a exposure disrupts genomic imprinting in the mouse. PLoS genetics, 9(4), e1003401.

Swayne, B. G., Kawata, A., Behan, N. A., Williams, A., Wade, M. G., MacFarlane, A. J., & Yauk, C. L. (2012). Investigating the effects of dietary folic acid on sperm count,

DNA damage and mutation in Balb/c mice. Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis, 737(1-2), 1–7.

Teperek, M., Simeone, A., Gaggioli, V., Miyamoto, K., Allen, G. E., Erkek, S., Kwon, T., Marcotte, E. M., Zegerman, P., Bradshaw, C. R., Peters, A. H. F. M., Gurdon, J. B., & Jullien, J. (2016). Sperm is epigenetically programmed to regulate gene transcription in embryos. Genome Research, 26(8), 1034–1046.

Trasler, J. M. (2006). Gamete imprinting: setting epigenetic patterns for the next generation. *Reproduction, Fertility and Development*, *18*(2), 63–63.

Tsuei, D.-J., Hsu, H.-C., Lee, P.-H., Jeng, Y.-M., Pu, Y.-S., Chen, C.-N., Lee, Y.-C., Chou, W.-C., Chang, C.-J., Ni, Y.-H., & Chang, M.-H. (2004). RBMY, a male germ cell-specific RNA-binding protein, activated in human liver cancers and transforms rodent fibroblasts. *Oncogene*, *23*(34), 5815–22.

Van Guelpen, B., Hultdin, J., Johansson, I., Hallmans, G., Stenling, R., Riboli, E., Winkvist, A., & Palmqvist, R. (2006). Low folate levels may protect against colorectal cancer. Gut, 55(10), 1461–1466. https://doi.org/10.1136/gut.2005.085480

Verheyen, S., Speicher, M. R., Ramler, B., & Plecko, B. (2020). Childhood-onset epileptic encephalopathy due to *Fgf12* exon 1–4 tandem duplication. *Neurology Genetics*, *6*(5), 494. https://doi.org/10.1212/NXG.0000000000000494

von Meyenn, F., Berrens, R. V., Andrews, S., Santos, F., Collier, A. J., Krueger, F., Osorno, R., Dean, W., Rugg-Gunn, P. J., & Reik, W. (2016). Comparative principles of DNA methylation reprogramming during human and mouse in vitro primordial germ cell specification. *Developmental Cell*, *39*(1), 104–115. https://doi.org/10.1016/j.devcel.2016.09.015

Vrooman, L. A., Rhon-Calderon, E. A., Suri, K. V., Dahiya, A. K., Lan, Y., Schultz, R. M., & Bartolomei, M. S. (2022). Placental abnormalities are associated with specific windows of embryo culture in a mouse model. Frontiers in Cell and Developmental Biology, 10. https://doi.org/10.3389/fcell.2022.884088

Walsh, C. P., & Bestor, T. H. (1999). Cytosine methylation and mammalian development. *Genes & Development*, *13*(1), 26–34.

Waterston, R. H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J. F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P., Antonarakis, Mouse Genome Sequencing Consortium. (2002). Initial sequencing and comparative analysis of the mouse genome. Nature, 420(6915), 520–62.

Wei, J.-W., Huang, K., Yang, C., & Kang, C.-S. (2017). Non-coding RNAs as regulators in epigenetics (review). Oncology Reports, 37(1), 3–9. https://doi.org/10.3892/or.2016.5236

Wong, W. Y., Merkus, H. M. W. M., Thomas, C. M. G., Menkveld, R., (2002). Effects of folic acid and zinc sulfate on male factor subfertility: a double-blind, randomized, placebo-controlled trial. Fertility and Sterility, 77(3), 491–498. https://doi.org/10.1016/S0015-0282(01)03229-0

Yadav, U., Kumar, P., Gupta, S., & Rai, V. (2017). Distribution of *MTHFR* C677T gene polymorphism in healthy north Indian population and an updated meta-analysis. Indian Journal of Clinical Biochemistry: Ijcb, 32(4), 399–410.

Yamada, K., Chen, Z., Rozen, R., & Matthews, R. G. (2001). Effects of common polymorphisms on the properties of recombinant human methylenetetrahydrofolate reductase. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(26), 14853–14858.

Yamaguchi, S., Hong, K., Liu, R., Shen, L., Inoue, A., Diep, D., Zhang, K., & Zhang, Y. (2012). *Tet1* controls meiosis by regulating meiotic gene expression. *Nature*, *492*(7429), 443–447. https://doi.org/10.1038/nature11709

Zhao, J., Lu, P., Wan, C., Huang, Y., Cui, M., Yang, X., Hu, Y., Zheng, Y., Dong, J., Wang, M., Zhang, S., Liu, Z., Bian, S., (2021). Cell-fate transition and determination analysis of mouse male germ cells throughout development. *Nature Communications*, *12*(1). https://doi.org/10.1038/s41467-021-27172-0