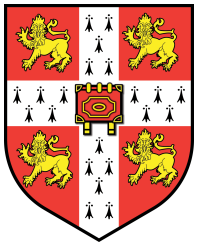
# Investigating the Epigenetic Mechanism Behind Transgenerational Inheritance in Mice with Abnormal Folate Metabolism

Georgina Emma Tallulah Blake



A dissertation submitted for the degree of
Doctor of Philosophy
at the University of Cambridge

St John's College
December 2018

**Investigating the Epigenetic Mechanism Behind Transgenerational Inheritance in Mice with Abnormal Folate Metabolism**

**Georgina Blake**

Exposure to environmental stressors can impact our health and that of future generations even when they are not similarly exposed. How disease risk is inherited is unclear. My thesis focuses on a mouse model of transgenerational inheritance in which folate metabolism is disrupted by a mutation in Methionine synthase reductase (*Mtrr*$^{gt}$). Remarkably, *Mtrr*$^{+/gt}$ heterozygosity leads to an increased likelihood of a wide spectrum of congenital malformations in their wildtype offspring for at least four generations. Folate metabolism is required for DNA synthesis and cellular methylation. Folate and its metabolism have also been linked to spermatogenesis and male fertility. My thesis aims to explore three possible mechanisms for the transgenerational inheritance of congenital malformations in the *Mtrr*$^{gt}$ model: germ cell morphology and function abnormalities, genetic instability and altered germ cell epigenetic patterns. We first considered if the *Mtrr*$^{gt}$ mutation affected testes morphology, spermatogenesis or sperm parameters. These were largely normal in *Mtrr*$^{gt}$ males. Next, we performed whole genome DNA sequencing of embryos to determine whether abnormal folate metabolism affects genetic stability. Importantly, the frequency of structural variants and single nucleotide polymorphisms (SNPs) were similar in C57Bl/6 control and *Mtrr*$^{gt/gt}$ embryos indicating that the *Mtrr*$^{gt}$ mutant genome is relatively stability. We did however identify an increase in SNP and SV frequency at the *Mtrr* locus, linked to the generation of the *Mtrr*$^{gt}$ mice in a 129/P2 genetic background prior to backcrossing into the C57Bl/6 background. Subsequently, we identified a large number of differentially methylated regions (DMRs) in sperm DNA of *Mtrr*$^{+/+}$, *Mtrr*$^{+/gt}$, and *Mtrr*$^{gt/gt}$ males compared to C57Bl/6 control sperm. Few DMRs were associated with underlying SNPs or SVs. However, no sperm DMRs in *Mtrr*$^{+/gt}$ males persisted in embryonic or adult tissues of the wildtype F1 or F2 generations. Despite this, we identified two genes, *Hira* and *Rn45s*, that were adjacent to DMRs and were misexpressed in the F2 and F3 generation embryos. This suggested that abnormal DNA methylation in sperm may influence gene expression two generations later despite reprogramming events. Additionally, we identified a number of differentially expressed small non-coding RNAs in *Mtrr*$^{+/gt}$ and *Mtrr*$^{gt/gt}$ sperm compared to C57Bl/6 control sperm. Overall, a complete understanding of the mechanisms behind transgenerational inheritance of phenotypes in the *Mtrr*$^{gt}$ model remains elusive. Unravelling the mechanisms of transgenerational epigenetic inheritance could have important implications for the future prediction and prevention of human diseases.

# Declaration

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text.

It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text.

It does not exceed the prescribed word limit for the relevant Degree Committee.

Georgina Blake
March 28, 2019

# Acknowledgements

I am eternally grateful to Erica Watson for the endless opportunities provided to me, the constant support, kind advice and faith in me. It is a pleasure to be a member of the Watson lab. Thanks to Anne Ferguson-Smith for her great enthusiasm for this project and for the challenging and insightful discussions of my data. Thank you to all the members of Watson lab, past and present. I am indebted to Russell Hamilton and Xiaohui Zhao for all the bioinformatic support they have provided during this project. I am grateful for our collaboration with Eric Miska and Katharina Gapp on the sncRNA-seq project. I extend thanks to the following people for their generous advice and support: Tessa Bertozzi, Claire Senner, Wayo Matsushima, Lizzie Radford and all those working in G5 and the CTR. I am extremely grateful to my funders, Wellcome Trust, St John's College and Cambridge Philosophical Society, for supporting me financially. I thank the Directors and members of the Wellcome Trust Developmental Mechanisms PhD programme, for the education, friendship and discussion forum the programme has provided. I also need to thank all the staff of the Combined Animal Facility for their help with our mouse colony. Thanks also to my friends and family for their support, care packages and coffee dates over the last few years. Finally, I want to thank my husband, Matt, without whom none of this would have been possible.

# Contents

# List of Figures

# List of Tables

# Abbreviations

| | |
|---|---|
| 5caC | 5-carboxylcytosine |
| 5fC | 5-formylcytosine |
| 5hmC | 5-hydroxymethylcytosine |
| 5mC | 5-methylcytosine |
| 5-methyl-THF | 5-methyltetrahydrofolate |
| 6mA | $N^6$-methyldeoxyadenosine |
| ADAMTS5 | A disintegrin-like and metallopeptidase (reprolysin type) with thrombospondin type 1 motif, 5 (aggrecanase-2) |
| $A^{vy}$ | *Agouti viable yellow* |
| ANOVA | Analysis of variance |
| ATP2B2 | ATPase, $Ca^{2+}$ transporting, plasma membrane 2 |
| bp | Base pair |
| BHMT | Betaine-homocysteine methyltransferase |
| BND | Break point |
| BTAF1 | B-TFIID TATA-box binding protein associated factor 1 |
| CBS | Cystathionine-$\beta$-synthase |
| CBLN-2 | Cerebellin 2 precursor protein |
| CD99L2 | CD99 antigen-like 2 |
| CGI | CpG island |
| Chr | Chromosome |
| CMTM1 | CKLF-like MARVEL transmembrane domain containing 1 |
| CNV | Copy number variant |
| CTS8 | Cathepsin 8 |
| DAB | 3,3-diaminobenzidine |
| DEL | Deletion |
| DHF | Dihydrofolate |
| DMR | Differentially methylated region |
| DNMT | DNA methyltransferase |
| DOCK10 | Dedicator of cytokinesis 10 |
| dTMP | Deoxythymidine monophosphate |
| dUMP | Deoxyuridine monophosphate |
| DUP | Duplication |
| E | Embryonic day |
| EFCAB6 | EF-hand calcium binding domain 6 |
| ELISA | Enzyme-linked immunosorbent assay |
| ESC | Embryonic stem cell |
| F | Filial generation |

| | |
|---|---|
| FGFBP3 | Fibroblast growth factor binding protein 3 |
| FOLR1 | Folate receptor 1 |
| FRM1OS | Fragile X mental retardation 1, opposite strand |
| GAS1 | Growth arrest specific 1 |
| GI | Glucose intolerance |
| GPR98 | Adhesion G protein-coupled receptor V1 (ADGRV1) |
| gt | Gene-trap |
| H3K27me3 | Histone 3 lysine 27 trimethylation |
| H3K4me3 | Histone 3 lysine 4 trimethylation |
| H3K9me2 | Histone 3 lysine 9 dimethylation |
| HCCS | Holocytochrome c synthetase |
| H & E | Haematoxylin and Eosin |
| HFD | High-fat diet |
| Hom/Het | Homozygous/Heterozygous ratio |
| HSD17B3 | Hydroxysteroid (17-beta) dehydrogenase 3 |
| HS6ST3 | Heparan sulfate 6-O-sulfotransferase 3 |
| IAP | Intracisternal A particle |
| IgG | Immunoglobulin G |
| IDE | Insulin degrading enzyme |
| IL9R | Interleukin 9 receptor |
| IGV | Integrative Genomics Viewer |
| INV | Inversion |
| INS | Insertion |
| IQSEC3 | IQ motif and Sec7 domain 3 |
| IR | Insulin resistance |
| IVF | *In vitro* fertilisation |
| KIF11 | Kinesin family member 11 |
| KLHL3 | Kelch-like 3 |
| KO | Knock-out |
| KRTAP20-2 | Keratin associated protein 20-2 |
| LINE | Long-interspersed element |
| LTR | Long-terminal repeat |
| lnc-RNA | Long non-coding RNA |
| MARCH5 | Membrane-associated ring finger (C3HC4) 5 |
| MAT | Methionine adenosyltransferase |
| Mb | Megabase (million base pairs) |
| MeDIP-Seq | Methylated DNA Immunoprecipitation followed by sequencing |
| miRNA | microRNA |
| MSUS | Maternal separation and unpredictable maternal stress |
| mt-tRNA | Mitochondrial tRNA |
| MTHFR | Methylene tetrahydrofolate reductase |
| MTR | Methionine synthase (also known as MS) |
| MTRR | Methionine synthase reductase |

| | |
|---|---|
| MTERF3 | Mitochondrial transcription termination factor 3 |
| ncRNAs | non-coding RNAs |
| NMD | Nonsense-mediated decay |
| NSG2 | Neuron specific gene family member 2 |
| NSUN2 | NOL1/NOP2/Sun domain family member 2 |
| NTC | No template control |
| OGFRL1 | Opioid growth factor receptor-like 1 |
| PAS | Periodic acid-Schiff |
| PBS | Phosphate buffered saline |
| PCA | Principle component analysis |
| PGC | Primordial germ cell |
| piRNA | Piwi-interacting RNA |
| PTCH1 | Patched 1 |
| Ref. | Reference allele |
| rp(k)m | Reads per (kilobase) million |
| rRNA | Ribosomal RNA |
| rsRNA | rRNA fragment |
| RT | Room temperature |
| RT-qPCR | Reverse transcription and quantitative polymerase chain reaction |
| SAM | S-adenosylmethionine |
| SD | Standard deviation |
| SINE | Short-interspersed element |
| SLC19A1 | Solute carrier 19a1 (also known as RFC1) |
| sncRNA | Small non-coding RNA |
| sncRNA-seq | Small-non-coding RNA sequencing |
| SNP | Single nucleotide polymorphism |
| SRD5A1 | Steroid 5 alpha-reductase 1 |
| SV | Structural variant |
| TEI | Transgenerational epigenetic inheritance |
| TET | Ten-eleven translocase |
| THF | Tetrahydrofolate |
| tPA | Tissue plasminogen activator |
| TPBPA | Trophoblast specific protein alpha |
| tRNA | Transfer RNA |
| tsRNA | tRNA fragment |
| Ts/Tv | Transition/ Transversion ratio |
| TSS | Transcription start site |
| UTR | Untranslated region |
| UQCRB | Ubiquinol-cytochrome c reductase binding protein |
| WGS | Whole genome sequencing |
| ZFP | Zinc finger protein |

# Chapter 1

# Introduction

This chapter contains material published in the following review and textbook chapter: Blake et al. (2018); Blake and Watson (2016). Permission was granted by the publisher for their inclusion in this thesis.

## 1.1 Transgenerational epigenetic inheritance

In recent years, it has become apparent that the environment experienced during an individual's lifetime may impact not only their health but also that of their descendants. Environmental stressors (e.g. poor diet, toxins, or psychological stress) also influence the epigenome (Beck et al., 2017; Radford et al., 2014; Gapp et al., 2014; Sharma et al., 2016). Abnormal epigenetic patterns could be inherited over multiple generations and contribute to disease. This is the phenomenon we call epigenetic inheritance. How commonplace epigenetic inheritance is and the underlying mechanisms remain uncertain, though studies over recent years have improved our understanding of the molecular pathways responsible for it.

### 1.1.1 Concepts of transgenerational epigenetic inheritance

We define epigenetic inheritance as the transmission of non-DNA base sequence information between generations via the germline (Daxinger and Whitelaw, 2012; Heard and Martienssen, 2014). Epigenetic changes in the parental generation (F0) occurring after exposure to an environmental insult increase the risk for specific phenotypes in subsequent generations (F1, F2, F3, etc.) even when they are not exposed to the insult themselves. To be considered transgenerational epigenetic inheritance (TEI), the phenotype must persist in the F2 generation or later when inherited via the paternal lineage and the F3 generation or later via the maternal lineage (Daxinger and Whitelaw, 2012; Heard and Martienssen, 2014) (Figure 1.1). In the paternal lineage, an F0 male and his germ cells that are destined to become the F1 generation are both directly exposed to the environmental insult. Therefore, to be considered TEI a phenotype and/or altered epigenetic patterns must be observed in the F2 generation or beyond, because the F2 generation is the first generation that was not directly exposed to the insult. In the maternal lineage, if environmental exposure occurs while a female is pregnant, the mother, the foetus (F1 generation) and its primordial germ cells (F2 generation) are all directly exposed. Thus, the persistence of phenotypes/epigenetic changes in the F3 generation or beyond is considered TEI. I use the term intergenerational inheritance to describe transmission between

generations (e.g. parent to offspring) that does not meet the criteria for true transgenerational inheritance.

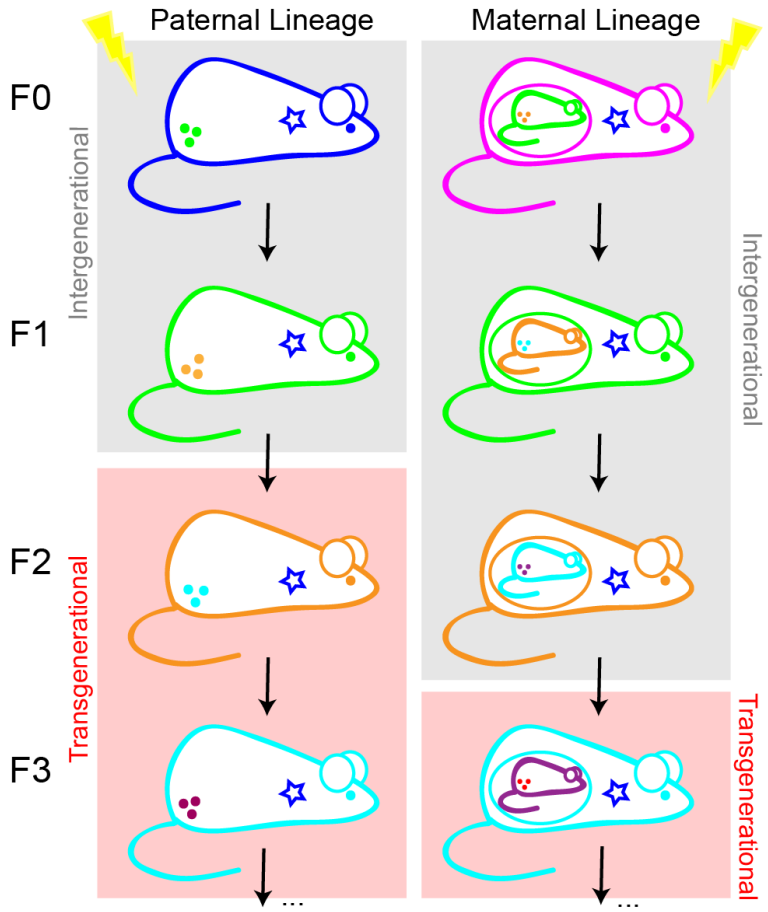



Figure 1.1: **Transgenerational epigenetic inheritance (TEI) within the paternal and maternal lineages.** A schematic example in mice of how epigenetic changes and phenotypes (blue star) induced by an environmental insult environmental insult (lightening bolt) in the first generation (F0) may be inherited via the germline over several generations (F1, F2, F3, etc.). Figure from Blake and Watson, 2016.

Our understanding of the mechanisms of TEI is in its infancy. It is clear that exposure to environmental stressors may alter the epigenome and that information is transmitted from the exposed individual to their offspring. There are two key paradigms of epigenetic inheritance: replication and reconstruction (Miska and Ferguson-Smith, 2016) (Figure 1.2). The initial dogma in the field of TEI was that germline epigenetic patterns were altered in the exposed generation and that these were inherited directly by the offspring, much like DNA.

This so called replicative inheritance requires epigenetic marks, like DNA itself, to be directly copied between cell divisions and generations. Importantly it also requires epigenetic marks to escape the reprogramming events that 'wipe the epigenetic slate clean' on transmission to the next generation (Figure 1.3). Two waves of reprogramming, the first in the post-fertilisation embryo and the second in developing germ cells, involve widespread erasure of DNA methylation patterns and chromatin remodelling (Reik and Surani, 2015).

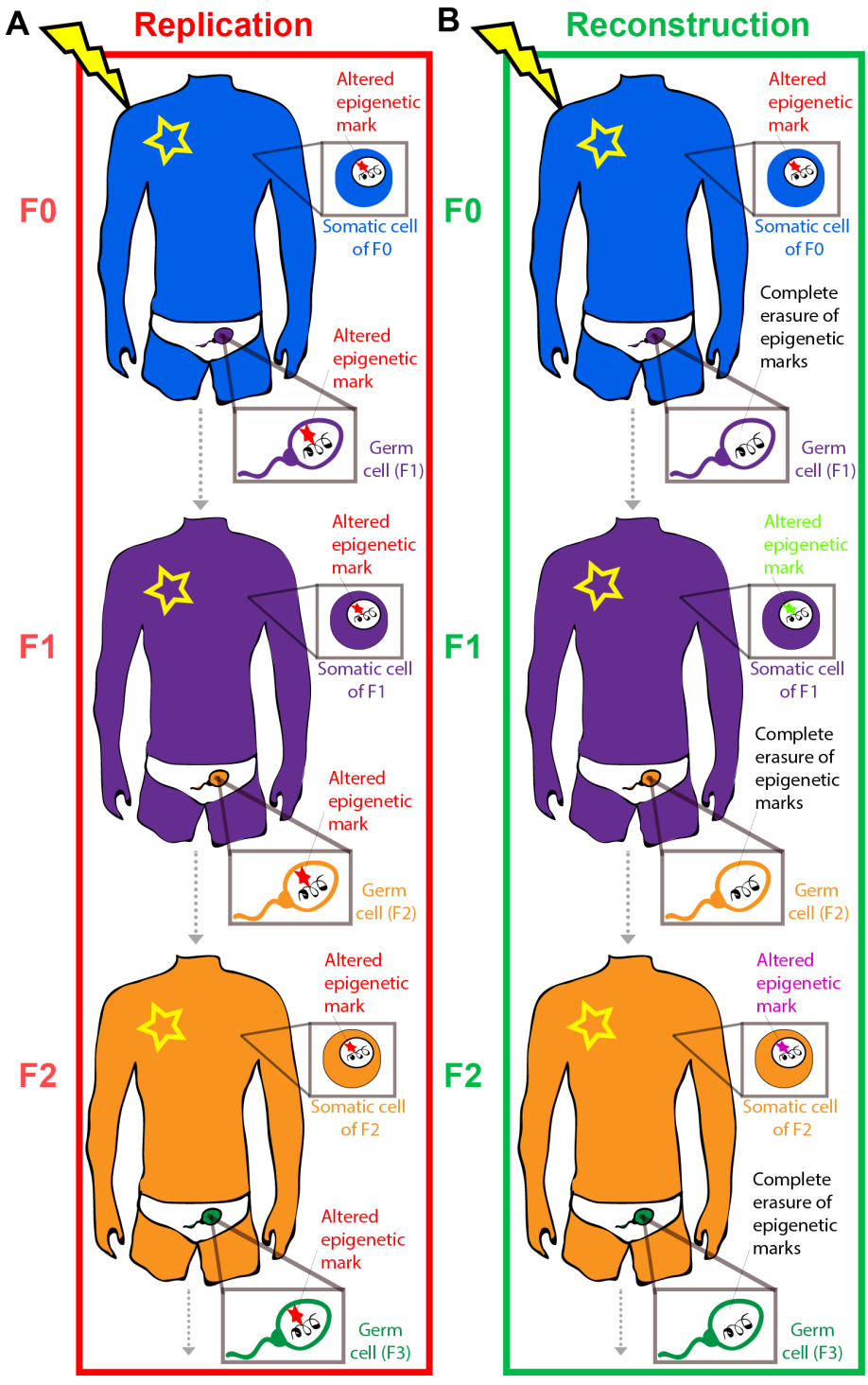

A
Replication
B
Reconstruction
F0
F1
F2
Altered epigenetic mark
Somatic cell of F0
Germ cell (F1)
Somatic cell of F1
Germ cell (F2)
Somatic cell of F2
Germ cell (F3)
Complete erasure of epigenetic marks

Figure 1.2: **Replication versus reconstruction models of epigenetic inheritance.** Two potential mechanisms of epigenetic inheritance caused by an environmental stressor (lightning bolt) include replication and reconstruction. **(A)** Replicative inheritance requires that abnormal epigenetic marks (denoted by red stars) be directly inherited via the germline between generations (dotted arrow). In order to do so, these marks must escape two waves of epigenetic reprogramming that occur in the developing germ cells and the post-fertilisation zygote. **(B)** Reconstructive inheritance suggests that epigenetic marks (denoted by red, green and pink stars) caused by an environmental stressor (lightning bolt) and responsible for phenotypes undergo normal reprogramming in the germline and zygote but are then reconstructed or recreated in each successive generation. Figure from Blake et al., 2018

Epigenetic reprogramming begins shortly after fertilisation (before DNA replication) and continues until about E3.5, after which epigenetic patterns begin to be re-established (Smith et al., 2012; Smallwood et al., 2011; Santos et al., 2013) (Figure 1.3). Initially there is rapid global active DNA demethylation of the paternal pronucleus thought to require base-excision repair (BER) (Oswald et al., 2000; Gu et al., 2011; Santos et al., 2013; Hajkova et al., 2010). The DNA demethylase Ten eleven translocase 3 (TET3) may be involved in protecting newly demethylated DNA from reacquisition of DNA methylation rather than in demethylation itself (Amouroux et al., 2016). The maternal genome primarily undergoes passive replication dependent demethylation achieved by DNMT1 exclusion (Santos et al., 2002). Some active demethylation of the maternal pronucleus also occurs (Guo et al., 2014). Imprinted regions and some repetitive loci are protected from this global demethylation (Smallwood et al., 2011; Lane et al., 2003). Additionally, extensive reprogramming of histone modifications occurs shortly after fertilisation, following the replacement of protamines in the paternal pronucleus by histones (van der Heijden et al., 2006; Marcho et al., 2015).

Reprogramming in primordial germ cells (PGCs) presents a second obstacle to inheritance beyond a single generation (Figure 1.3). As PGCs migrate to the genital ridge of the developing embryo, DNA methylation is passively removed (Seisenberger et al., 2012; Kobayashi et al., 2013). A second wave of DNA demethylation at around E11.5 involves active removal of residual DNA methylation including at imprinted loci and CpG islands (CGIs) on the

X-chromosome (Hackett et al., 2013). DNA demethylation on PGCs requires BER but does not directly require TET1 mediated methyl-cytosine oxidation (Hajkova et al., 2010; Hill et al., 2018). Once DNA demethylation is complete the germ cells enter mitotic (male) or meiotic (female) arrest (Reik et al., 2001). In male germ cells, DNA methylation patterns are re-established prior to birth (Davis et al., 2000). Sperm DNA methylation is then maintained during spermatogenesis, although there is evidence that some genes acquire their methylation during sperm development (Oakes et al., 2007). In oocytes, DNA methylation is re-established after birth as the oocytes mature (Smallwood et al., 2011; Kota and Feil, 2010). Histone marks are also extensively remodelled in PGCs (Seki et al., 2005). In sperm, histones are replaced by protamines to increase DNA compaction within the nucleus (Krawetz, 2005). Some histones are retained in sperm at two distinct sites: promoters of developmentally regulated genes (Erkek et al., 2013) and gene-poor repeat regions (Carone et al., 2014). Nucleosomes containing histones are retained genome wide in the mouse oocyte (Erkek et al., 2013). Together, these two extensive waves of reprogramming limit the possibilities of epigenetic inheritance. However, loci that are resistant to reprogramming in the post-fertilisation zygote and in PGCs have been identified (Hackett et al., 2013; Smallwood et al., 2011; Kobayashi et al., 2012) and these provide scope for replicative inheritance.

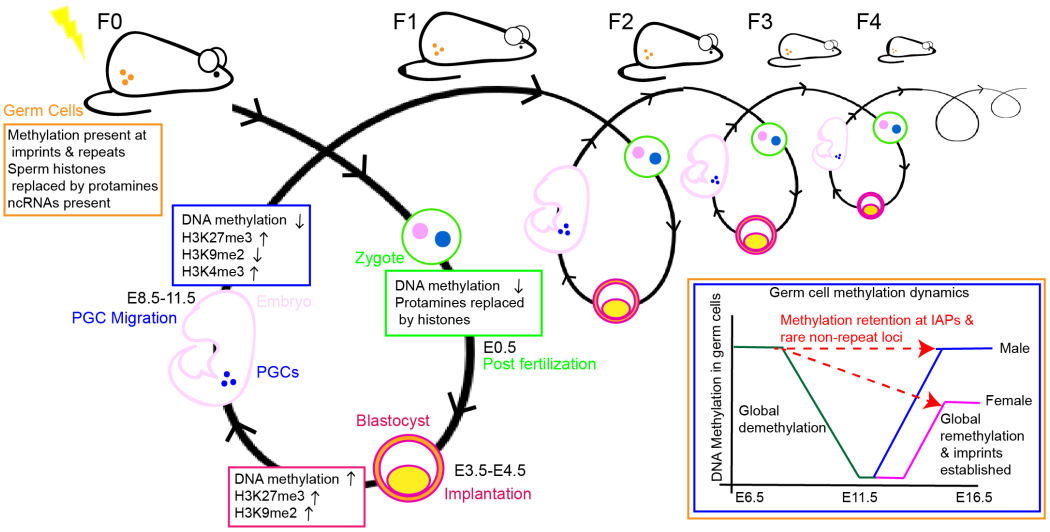



Figure 1.3: **Epigenetic reprogramming in the early embryo and primordial germ cells.** During the first wave of epigenetic reprogramming in the post-fertilisation embryo the maternal and paternal genomes are demethylated, and sperm protamines are replaced by histones. Later, *de novo* DNA methylation and dynamic changes to histone modifications occur during blastocyst formation and implantation. In the second wave of epigenetic reprogramming, as primordial germ cells (PGCs) migrate to the genital ridges in the embryo, genome-wide DNA demethylation occurs and histone marks are altered. Thereafter, re-establishment of epigenetic marks occurs in the maturing gametes in a sex-specific manner. The majority of histones are removed and replaced by protamines during spermatogenesis. Germ cell DNA methylation dynamics are indicated in the graph (bottom right, adapted from Heard and Martienssen, 2014). An environmental insult (lightening bolt) can influence epigenetic patterns in the germline. H3K27me3, histone 3 lysine 27 trimethylation; H3K4me3, histone 3 lysine 4 trimethylation; H3K9me2, histone 3 lysine 9 dimethylation; ncRNAs, non-coding RNAs; IAP, intracisternal A particle; E, embryonic day. Figure from Blake and Watson, 2016.

Reconstructive inheritance was postulated by Eva Jablonka as an alternative theoretical concept to replicative inheritance. The reconstruction model avoids the challenges presented by epigenetic reprogramming by suggesting that epigenetic marks responsible for a phenotype undergo normal reprogramming in the germline and zygote but are then recreated in each successive generation (Jablonka, 2013). Traces of the parental epigenetic landscape may be sufficient to reconstitute the parental epigenetic state in the offspring and thus bring about a phenotype (Jablonka, 2013). For example, the reconstruction of abnormal epigenetic marks and phenotypes may be driven via altered cellular

signalling, metabolism or ncRNAs (Miska and Ferguson-Smith, 2016). This mechanism may lead to serial reconstruction of an induced phenotype over multiple generations, such that an abnormal phenotype in the F0 generation programs the same defect in the F1 generation, and so on (Aiken and Ozanne, 2014). Despite their seemingly contradictory messages, the replication and reconstruction models need not be juxtaposed. In reality, there may be an element of direct inheritance of some epigenetic marks (replication) and reconstruction of additional marks linked to an ancestral phenotype.

### 1.1.2 Mechanisms of TEI

Epigenetic marks, namely DNA methylation, histone modifications and non-coding RNAs, have been postulated as the information carrying vectors between generations. These three candidates have been studied in many TEI paradigms. It is hypothesised that stable alterations in epigenetic marks induced by an environmental insult (epimutations) may affect gene expression or chromatin stability leading to increased risk of a phenotype or disease in the offspring (Miska and Ferguson-Smith, 2016). Within individual cells, DNA methylation, histone modifications and ncRNAs are interconnected and interdependent. In the context of epigenetic inheritance they likely act in a coordinated and collaborative manner to cause disease phenotypes.

#### 1.1.2.1 DNA methylation

In humans and other mammals, methylation of cytosine residues (5mC) is the predominant form of DNA methylation. DNA methylation in mammals is generally present in the context of CpG dinculeotides (Bird, 2002). DNA methylation is dispersed across the genome, with a large proportion of CpGs methylated in somatic cells (Bird, 2002). However, DNA methylation patterns are dynamic. DNA methylation patterns differ in time and space, changing during development and differing between cell types (Messerschmidt et al., 2014; Luo et al., 2018). A family of DNA methyltransferase (DNMT) enzymes establish and maintain DNA methylation. DNMT1 ensures methylation is recapitulated on to the daughter strand of DNA after replication has occurred (Almouzni and Cedar, 2016). DNMT3a and DNMT3b establish de novo DNA methylation patterns as occurs in germ cells (Almouzni and Cedar, 2016). Alternatively, the Ten eleven translocases (TETs) remove methylation via a 5-hydroxymethylcytosine (5hmC) intermediate and then further oxidise 5hmC into 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) (Rasmussen and Helin, 2016). It is now becoming clear that 5hmC may have functional roles itself (Ficz et al., 2011). DNA methylation in gene control regions is generally associated with gene repression, although its exact role in gene regulation is likely locus dependent (Deaton and Bird, 2011). Beyond regulation of gene expression, a key role of DNA methylation is to silence repetitive DNA (Slotkin

and Martienssen, 2007). DNA methylation has many properties that make it an attractive mechanistic candidate for epigenetic inheritance including that: 1) it can be environmentally modulated; 2) machinery exists to replicate methylation patterns onto newly synthesised DNA and thus it is mitotically heritable; and 3) methylated loci resistant to epigenetic reprogramming in the zygote and germline have been identified. Disruption of the DNA methylation machinery may also play an important role in initiating TEI.

#### 1.1.2.2 Histone modifications

DNA in somatic cells is packaged around an octomeric core of histone proteins (H2A, H2B, H3 and H4) forming a nucleosome (Marmorstein and Trievel, 2009). Histone proteins are subject to chemical modifications (e.g. methylation, acetylation etc.) on their N-terminal tails (Marmorstein and Trievel, 2009). These modifications can regulate, in part, chromatin packaging and nucleosome positioning on DNA. Histone modifications, like DNA methylation, are dynamic in space and time. Certain histone modifications are generally recognised as being associated with gene repression e.g. trimethylation of lysine 27 on histone 3 (H3K27me3), whereas acetylation of the same residue (H3K27ac) is generally associated with gene activation (Lawrence et al., 2016). Histone modifications serve to regulate gene expression though the recruitment of protein complexes and by controlling the accessibility of the DNA to the transcription machinery (Lawrence et al., 2016).

Histone modifications are a plausible mechanistic candidate in TEI as they can be environmentally modulated. Some histone modifications in the germline and zygote are retained during epigenetic reprogramming. Histones are propagated onto newly assembled chromatin by complex machinery (Marmorstein and Trievel, 2009), however how histone codes are retained during mitosis and meiosis is still largely unresolved (Erkek et al., 2013). Additionally, most histones are removed during spermatogenesis and replaced with protamines (Casas and Vavouri, 2014). Protamine modifications that are similar to histone modifications are present in sperm (Brunner et al., 2014), though their functional importance remains uncertain, particularly during epigenetic inher-

itance. Disruption of the machinery that regulates histone modification may also be an important initiator of TEI (Siklenka et al., 2015; Greer et al., 2016).

#### 1.1.2.3 non-coding RNA

Germ cells have extensive regulatory non-coding RNA (ncRNA) profiles. ncRNAs can be classified broadly into two classes: 1) long-ncRNA, greater than 200 nucleotides in length (lncRNAs) (Lee et al., 2012) and 2) small-ncRNA, less than 200 nucelotides (sncRNA) (Bouckenheimer et al., 2018). sncRNAs include micro-RNAs (miRNAs), small-interfering-RNAs (siRNAs), piwi-interacting-RNAs (piRNAs) and transfer-RNAs (tRNAs) amongst others. ncRNAs have diverse roles from direct regulation of gene expression to acting to localise other epigenetic pathways. For example, piRNAs mediate transposon silencing in the germline (Holoch and Moazed, 2015).

ncRNAs are in some respects uniquely suited to transfer information between generations. Both sperm and oocytes contain a range of ncRNA species (Veselovska et al., 2015; Rando, 2016). These can be environmentally modulated (Beck et al., 2017; Sharma et al., 2016). There is evidence sperm can take up ncRNAs from the surrounding somatic tissues (Vojtech et al., 2014; Cossetti et al., 2014). For example, maturing sperm might receive sncRNAs (e.g., tRNA fragments) from small vesicles called exosomes released from the epididymis (Vojtech et al., 2014). Thus, the theoretically impassable wall between somatic cells and the germline (the so-called Weismann barrier) may be circumvented and penetrated by RNAs. This has important implications for our understanding of epigenetic inheritance resulting from environmental insults experienced in adulthood. Additionally, RNAs can carry chemical modifications, such as methylation (Zhang et al., 2016). This provides an additional layer of epigenetic information that might be transmitted to the next generation. However, how epigenetic messages transmitted by RNA are perpetuated over multiple cell divisions let alone multiple generations currently remains unclear.

### 1.1.3 Models of epigenetic inheritance

There is a myriad of studies in the field of intergenerational and transgenerational epigenetic inheritance. Epigenetic inheritance can seemingly be initiated or modulated by mutations in the epigenetic machinery, either DNA/RNA methylation enzymes or histone modification complexes (Table 1.1). This suggests a conserved role for the epigenetic machinery in regulating epigenetic inheritance across a range of species . Alternatively, epigenetic inheritance may be instigated by environmental perturbations, with various insults leading to offspring phenotypes in a range of model systems (Table 1.2). These models provide interesting insight into the mechanisms of epigenetic inheritance. However, studies often investigate only parent to offspring transmission, and are therefore confounded by exposure of the germ cells that become the F1 generation to the environmental insult. The use of locus-specific approaches, rather than unbiased genome-wide techniques, and a failure to assess the germline itself are further limitations affecting many epigenetic inheritance studies.

Table 1.1: **Epigenetic inheritance paradigms involving mutations in the epigenetic machinery**.

| Model | Wildtype offspring Phenotype | Generations assessed | Molecular Mechanism | Reference |
|---|---|---|---|---|
| DNA METHYLTRANSFERASE 1 (MET1) mutation in *Arabidopsis Thaliana* | Late flowering phenotype | Several | Hypomethylation at tandem repeats upstream of FLOWERING WAGENINGEN (FWA) gene causing ectopic FWA expression. Methylation changes at other loci involved. | Kankel et al. (2003); Soppe et al. (2000) |
| Deletion of *Spr-5*, an H3K4me2 demethylase, in *C. elegans* | Progressive decline in fertility but increased longevity | 20 | Global accumulation of H3K4me2 and an increase in 6mA levels. A network of chromatin regulators also play a role. | Greer et al. (2016, 2011) |
| Overexpression of *Kdm1a*, an H3K4 demethylase, in mice | Birth defects (skin, skeletal and limb abnormalities), increased neonatal mortality, altered gene expression | Up to F3 | F0 sperm: reduced H3K4me2 in spermatozoa (not sperm), DNA methylation unaltered, altered RNA profile | Siklenka et al. (2015) |
| *Dnmt2*$^{-/-}$ mice fed HFD. DNMT2 is the RNA methyltransferase | Rescue of glucose intolerance and insulin resistance phenotype normally observed in offspring of fathers fed a HFD | F1 | *Dnmt2*$^{-/-}$prevents modification of fragmented tRNAs induced by HFD | Zhang et al. (2018) |
| Haploinsufficiency for *Setdb1*, a H3K9 methyltransferase, in male mice | When crossed to $A^{vy/a}$ females, offspring had increased tendency to be yellow than controls | F1 | *in trans* paternal effect as $A^{vy}$ inherited from dams. Hypomethylation at ERV elements in *Setdb1*$^{+/-}$ sperm | Daxinger et al. (2016) |

A non-exhaustive table of selected studies in which the phenotypes of the wildtype offspring are impacted by a parent with a mutation in the epigenetic machinery. HFD: High fat diet, ERV: endogenous retrovirus.

Table 1.2: **Epigenetic inheritance paradigms instigated by environmental perturbations**.

| Model/Exposure | Offspring Phenotype | Generations assessed | Molecular Mechanism | Reference |
|---|---|---|---|---|
| *In utero* undernutrition of F1 mice | F2 adult: reduced muscle mass, increased adiposity, glucose intolerance. F2 embryo at E16.5: increased lipid abundance | F2 | F1 sperm: differentially methylated regions (DMRs) identified. F2 tissues: DMR DNA methylation normal, genes misexpressed | Radford et al. (2014) |
| Pre-diabetic mouse model (HFD and streptozotocin exposure) | Prediabetes including glucose intolerance,↓ insulin sensitivity. | F1 and F2 | F0 sperm: differentially methylated regions (DMRs) identified. F1: Some DMRs present in pancreatic islets. F2: DNA methylation at DMRs normalised | Wei et al. (2014) |
| *In utero* vinclozolin exposure in rats | Reduced fertility, prostate and kidney disease, immune system abnormalities, tumour development | Up to F4 | F1 and F3 generation sperm have DNA methylation changes, although at different loci. F3 sperm: altered histone modifications (not seen in F1 and F2 sperm), ncRNA (primarily tRNAs) misexpressed. CNVs increased in F3 sperm. | Anway et al. (2005); Beck et al. (2017); Schuster et al. (2016); Ben Maamar et al. (2018) |
| Western-diet (high-fat, high-sugar) in mice | Obesity, glucose intolerance and insulin resistance | F1 | F0 sperm miRNAs misexpressed. miR19b when injected into 1 cell embryos could induced metabolic alterations similar to diet induced phenotype. | Grandjean et al. (2015) |

| Model and Exposure | Offspring Phenotype | Genera-tions assessed | Molecular Mechanism | Reference |
|---|---|---|---|---|
| High-fat diet in mice | F1: Obesity, glucose intolerance and insulin resistance, transcriptional changes in islets | F1 | Sperm miRNAs and tsRNAs misexpressed. tsRNA modifications required, DNA methylation changes in islets but do not correspond to misexpressed genes | Chen et al. (2016a) |
| High-fat diet in mice | Obesity, insulin resistance and sub-fertility | F2 | Sperm miRNAs misexpressed, decreased global germ-cell methylation | Fullston et al. (2013) |
| Low-protein diet in mice | Altered hepatic cholesterol biosynthesis | F1 | tsRNAs, miRNAs, piRNAs and let-7 misexpressed in sperm. tsRNAs may induce phenotype by repression of genes associated with the endogenous retroelement MERVL. | Sharma et al. (2016) |
| Environmental enrichment (physical and mental exercise) in mice | Enhanced synaptic plasticity, cognition and memory in F1 but not F2 offspring | F2 | Sperm miRNAs misexpressed (particularly miR212/132) | Benito et al. (2018) |
| Exercise in mice | Reduced anxiety and fear memory in males | F1 | Sperm miRNAs and tsRNAs misexpressed | Short et al. (2017) |
| Air pollutant exposure in pregnant mice | Increased likelihood of allergic asthma | F3 | DNA methylation changes in dendritic cells | Gregory et al. (2017) |

| Model and Exposure | Offspring Phenotype | Genera-tions assessed | Molecular Mechanism | Reference |
|---|---|---|---|---|
| High temperature induced expression of a multicopy fluorescent reporter array under the control of a heat shock promoter in *C. elegans* | Elevated reporter gene expression. Persistence of elevated gene expression depends upon copy number of reporter and duration of high temperature exposure. | 14 | H3K9me3 methyltransferase, *set-25*, activity reduced by high temperatures leading to altered histone methylation at transgene | Klosin et al. (2017) |
| *In utero* Bisphenol-A (BPA), bis(2-ethylhexyl)phthalate (DEHP) and dibutyl phthalate (DBP) exposure in rats | Increased body weight, testis and ovary diseases, pubertal abnormalities | F3 | Differential DNA methylation in F0 sperm | Manikkam et al. (2013) |
| *In utero* exposure to jet fuel (JP-8) in mice | F3: primordial follicle loss and polycystic ovarian disease, obesity. Additional prostate and kidney abnormalities in F1 | F3 | Differential DNA methylation in F0 sperm | Tracey et al. (2013) |
| *In utero* valproic acid exposure in mice | Autism-like behaviours | F3 | Misexpression of excitatory and inhibitory synaptic markers in frontal cortices of F1 and F3 | Choi et al. (2016) |

| Model and Exposure | Offspring Phenotype | Genera-tions assessed | Molecular Mechanism | Reference |
|---|---|---|---|---|
| Maternal separation and unpredictable maternal stress (MSUS) in mice | Behavioural phenotypes: ↑ risk taking, fear, despair. Metabolic: ↓ body weight, insulin hypersensitivity, glucose intolerance | F3 | DNA methylation changes at glucorticoid receptor promoter. Sperm miRNA, lncRNA and sncRNA misexpressed. | Gapp et al. (2014, 2018, 2016) |
| Starvation-induced developmental arrest in *C. elegans* | Increased lifespan | F3 | Misexpression of small RNAs primarily endo-siRNAs | Rechavi et al. (2014) |
| Olfactory fear conditioning in F0 mice | F1 and F2: Increased behavioural sensitivity to odour used in conditioning but not other odours | F2 | DNA hypomethylation of *Olfr151* odourant receptor in sperm of the F0 and F1 generations | Dias and Ressler (2014) |
| Female mice fed a high-fat/high-sugar diet | F1: Glucose intolerance, hyperinsulinemia. F1-F3 mitochondrial dysfunction and abnormal morphology in muscle | F3 | F1 and F2: oocyte mitochondrial dysfunction | Saben et al. (2016) |

A non-exhaustive table of selected studies in which the phenomenon of epigenetic inheritance has been explored. HFD: high fat diet ↓: decreased, ↑: increased. E: embryonic day

#### 1.1.3.1 The *Agouti*$^{vy}$ model

The *Agouti viable yellow* ($A^{vy}$) model is the classic mammalian model of TEI from which much of our mechanistic understanding has been derived (Morgan et al., 1999) (Figure 1.4). The *Agouti* gene (a) controls mouse coat colour. The $A^{vy}$ allele contains an intracisternal A particle (IAP) element containing an ectopic promoter upstream of the *Agouti* transcription start site (TSS). This regulates *Agouti* gene expression, therefore coat colour, in a DNA methylation-dependent manner in $A^{vy}$ individuals (Morgan et al., 1999; Miltenberger et al., 1997) (Figure 1.4 A). Methylation of the IAP suppresses ectopic *Agouti* mRNA expression to generate brown (pseudo-agouti) mice (Morgan et al., 1999). Hypomethylation at the IAP element drives ectopic *Agouti* expression leading to mice with a yellow coat colour (Morgan et al., 1999). $A^{vy}$ mice also have obesity and a diabetic-like phenotype, owing to disruption of their satiety signalling pathways (Cropley et al., 2016; Miltenberger et al., 1997).

Remarkably, the methylation status of the $A^{vy}$ allele, and therefore, the coat colour and metabolic phenotypes, can be inherited over multiple generations via the maternal line (Morgan et al., 1999). For example, a yellow $A^{vy/a}$ mouse has an increased propensity to have yellow offspring when mated to an *a/a* male (Morgan et al., 1999) (Figure 1.4 B). The methylation status of the agouti gene can also be influenced by environmental factors, such as nutrition and endocrine disruptor exposure (Cropley et al., 2010; Dolinoy et al., 2007). Feeding $A^{vy}$ females a methyl-rich diet reduces the number of yellow offspring in their litters (Cropley et al., 2010; Dolinoy et al., 2007) (Figure 1.4 C). Whether DNA methylation patterns at the IAP element normalise in response to increased methyl groups remains unclear (Cropley et al., 2010; Dolinoy et al., 2007). However, the $A^{vy}$ locus undergoes epigenetic reprogramming during early embryogenesis including the removal of IAP methylation following maternal transmission of the allele (Blewitt et al., 2006). This suggests a mechanism of inheritance that might be independent of DNA methylation or reconstruction-based. Indeed, sperm ncRNA profiles, including tRNA fragments, are abnormal in the sperm of F1 males derived from mating an $A^{vy}$ obese male to wildtype lean female (Cropley et al., 2016). This suggests RNA-based mech-

anisms may contribute to inheritance and reconstruction of abnormal DNA methylation profiles at the IAP element.

The $A^{vy}$ model highlights several key points regarding TEI. It involves a repetitive element, the epigenetic marks are responsive to environmental modulation in the exposed generation and several epigenetic factors are dysregulated together. I discuss some of these in section 1.1.4, and also address the limitations of applying findings from the $A^{vy}$ model to TEI more generally.

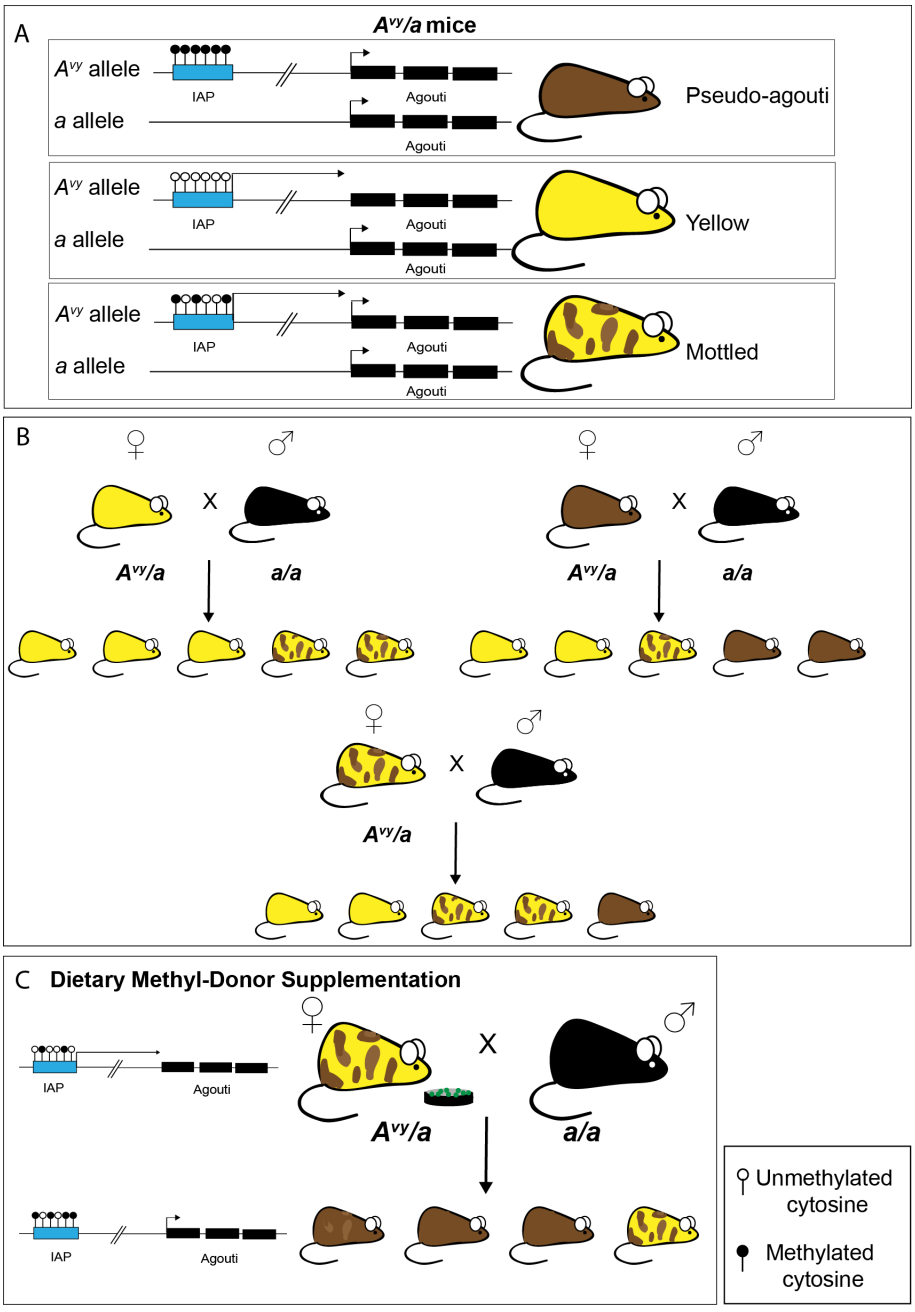


Figure 1.4: ***Agouti viable yellow ($A^{vy}$)*: a classic example of epigenetic inheritance**. **(A)** An intracisternal A particle (IAP) upstream of the *Agouti* coat colour gene regulates *Agouti* gene expression in a DNA methylation-dependent manner in $A^{vy}$ mice. When the IAP element is fully methylated, *Agouti* expression is driven from the nascent promoter resulting in mice with a brown (pseudoagouti) coat colour. When the IAP element is unmethylated, a promoter in the IAP element drives ectopic *Agouti* expression leading to a yellow coat colour and a metabolic phenotype. Partial methylation at the IAP element leads to a mottled coat colour. **(B)** The methylation status of the *Agouti* allele, and therefore the coat colour and metabolic phenotypes, are heritable over multiple generations via the maternal line. A yellow $A^{vy/a}$ female will have an increased propensity to have yellow offspring compared to a brown $A^{vy/a}$ female when mated to an *a/a* male due to epigenetic inheritance at the $A^{vy}$ allele. In the schematic, only $A^{vy/a}$ offspring are depicted, *a/a* progeny have been excluded for simplification. **(C)** The methylation status of the *Agouti* gene can be influenced by environmental factors, such as dietary methyl donor supplementation. A methyl-rich diet fed to a mottled coat coloured $A^{vy/a}$ female reduces the number of yellow offspring in her litters with respect to unsupplemented controls (see panel B). Figure based upon images in Morgan et al., 1999. Figure from Blake et al., 2018

## 1.1.4 Common themes and conflicts in the study of TEI

### 1.1.4.1 The repetitive genome

A large part of most genomes are made up of repetitive elements, such as transposons and retrovirus-derived sequences. These need to be silenced to prevent their transposition into other regions of the genome resulting in mutation (Crichton et al., 2014). Epigenetic mechanisms are vital to their suppression. Epigenetic marks at repetitive elements may be more likely to be resistant to epigenetic reprogramming than unique loci, thus providing scope for heritability (Hackett et al., 2013; Smallwood et al., 2011; Kobayashi et al., 2012). Numerous models of TEI have implicated repetitive loci in their mechanism of inheritance, for example the heat-shock induced expression of a multi-copy fluorescent reporter array in *C. elegans* (Klosin et al., 2017) and in $A^{vy/a}$ mice. The $A^{vy}$ allele is classified as a metastable epiallele, defined as an IAP element that has variable methylation between isogenic individuals (Kazachenka et al., 2018). Recently, using genome-wide screens, many more metastable epialleles have been identified (Kazachenka et al., 2018). Like the $A^{vy}$ allele, the methylation at these IAP elements is reprogrammed after fertilization. However, unlike

$A^{vy}$, metastable epiallele methylation is then reestablished in a variable manner in the offspring rather than reflecting the methylation status of the parent. The inheritance of methylation status only occurs at exceptional loci (Kazachenka et al., 2018). Additionally, most did not affect transcription (Kazachenka et al., 2018). This suggests that the $A^{vy}$ locus is very much the exception rather than the rule and cautions against making generalisations about epigenetic inheritance based on our findings regarding the $A^{vy}$ allele. However, this does not preclude that the repetitive genome may still play an important role in TEI.

#### 1.1.4.2 Interactions between mechanisms

There is a significant level of interaction between DNA methylation, histone modifications and their related machineries. Histone modification patterns can influence DNA methylation via interactions with the DNA methyltransferase enzymes. For example, DNMT3 enzymes each contain an ADD domain (ATRX-DNMT3-DNMT3L) that recognises unmodified H3 and is inhibited by H3K4 methylation (Otani et al., 2009). Interactions between histones and the DNA methylation machinery can also be facilitated by adaptor proteins. The multi-domain protein UHRF1 can act as an adaptor between histone methylation and DNMT1 and may play a role in mediating H3K9 methylation associated DNA methylation (Zhao et al., 2016; Du et al., 2015). An excellent example of the histone modification-DNA methylation interaction occurs in oocytes. The unique DNA methylation landscape of oocytes, with DNA methylation in transcribed gene bodies, is associated with the acquisition of H3K6me3 and the KDM1A mediated loss of H3K4me3 at CGIs (Gahurova et al., 2017; Stewart et al., 2015). Furthermore, interplay between DNA methylation and histone modifications is important in the protection of imprinted loci from reprogramming in the early embryo. The binding of a range of factors and cofactors (e.g. STELLA) to H3K9me2 and H3K9me3 may play a role in protecting parental DNA methylation at imprinted domains (Nakamura et al., 2007; Quenneville et al., 2011). Furthermore, UHRF1 and histone binding proteins act as part of a complex to recruit DNMT1 and the H3K9 methyltransferase SETDB1 and exclude TET enzymes at imprinted genes (Messerschmidt, 2012).

An increasing number of TEI studies have combined analyses of DNA methylation, histone modifications and ncRNAs in a single model system. For example DNA methylation, histone modifications and ncRNAs have been assessed in the sperm of F3 generation rats ancestrally exposed to vinclozolin (Ben Maamar et al., 2018). However, studying the F3 generation reveals the endpoint epigenetic status not those epimutations responsible for instigating inheritance. Additionally, how epigenetic modifications interact is often not explored. For example, DNA methylation or histone modification patterns might regulate small ncRNA expression. This, in turn, may direct DNA/histone methylation patterns to ultimately establish an interactive feedback system (Miska and Ferguson-Smith, 2016). Further collaborative, multifaceted, approaches to assess all epigenetic pathways in a single model are required to fully explore epigenetic mechanisms of inheritance.

#### 1.1.4.3 The bandwidth of inheritance

It is debated in the field of epigenetic inheritance whether specific information regarding an adverse environment is transmitted to the next generation or whether a more general signal of suboptimal conditions is inherited. This is described in terms of the extent or bandwidth of information that can be communicated (Rando, 2016). Evidence supporting the transfer of specific information comes from a multigenerational study assessing olfactory fear conditioning in mice (Dias and Ressler, 2014). In this study, F0 males were conditioned to associate a specific odour with a foot shock (Dias and Ressler, 2014). The F1 and F2 generations derived from these conditioned males have increased behavioural sensitivity to the conditioned odour and not other odours even though they had not been conditioned themselves (Dias and Ressler, 2014). Importantly, social transmission of odour sensitivity was excluded using *in vitro* fertilization and cross-fostering (Dias and Ressler, 2014). Alternatively, recent evidence for inheritance of a general message comes from exposing mice to nicotine (Vallaster et al., 2017). This exposure primes their male F1 offspring for increased survival when exposed to toxic levels of nicotine or cocaine (Vallaster et al., 2017). This suggests a general enhanced xenobiotic

resistance is inherited (Vallaster et al., 2017). It may be that the bandwidth of inheritance is epigenetic inheritance paradigm specific.

#### 1.1.4.4 Correlating epigenetic changes to phenotypes

It is rare that researchers are able to correlate the epimutations identified directly to phenotypic outcomes. In many models of epigenetic inheritance, offspring generations have various phenotypes involving numerous tissues and biological pathways. However, the epigenetic changes reported in such models can often be limited to a restricted number of loci. It is possible that epimutations impacting expression of a key transcriptional regulator or epigenetic modifier could lead to a cascade of transcriptional events that result in a broad spectrum of phenotypes.

Additionally when considering DNA methylation it is important to recognise that the DNA methylation status at any given cytosine residue is binary. Each cytosine can be methylated or unmethylated. However, moderate changes in CpG methylation (<10%) are often reported in intergenerational inheritance models in which DNA methylation is assessed (for example in Radford et al. (2014)). These are often associated with complex disease phenotypes. In general, the degree of DNA methylation change that can lead to dramatic alternations in gene expression and severe phenotypes remains poorly explored. Understanding the biological significance of small DNA methylation changes that result from environmental stressors is a challenge that should not be ignored.

In the case of ncRNAs, causality between germline misexpression of ncRNA species and offspring phenotypes has been demonstrated to some extent using microinjection studies. Microinjection of ncRNA species misregulated in exposed male sperm into naive fertilised oocytes has been shown to be able to at least partially recapitulate offspring phenotypes (Chen et al., 2015; Gapp et al., 2014; Sharma et al., 2016). While not definitive, these studies go some way to demonstrating that specific germline epigenetic changes can impact gene expression and offspring phenotype.

#### 1.1.4.5 Allowing for phenotypic and epigenetic variability

Under normal circumstances, considerable epigenetic variation exists between individuals (Shea et al., 2015). Therefore, this must be considered when assessing epigenetic changes in models of epigenetic inheritance. This is rarely the case, with many studies pooling samples together for analysis. Remarkably, one study showed epigenetic variation as a result of diet manipulation is less than natural epigenetic variation between individuals (Shea et al., 2015). It is currently unclear whether environmental insults affect the epigenome stochastically in each individual and/or within the germ cell population, or whether there are hotspots within the epigenome such that a population is similarly affected. Single cell methylome and transcriptome technologies may permit us to better understand this heterogeneity and epigenome sensitivity.

#### 1.1.4.6 Considering genetic effects

Epigenetic inheritance is often defined as inheritance that is entirely independent of the DNA-base sequence. However, the genome and epigenome are intrinsically linked. Very few researchers actually confirm that genetic factors do not play a role in the inheritance of phenotypes in their TEI paradigm. The reasons this must be considered are twofold. Firstly, epigenetic instability may promote genetic instability. For example, DNA methylation is a known mutagen since spontaneous deamination of methyl-cytosine forms thymine and thus generates a C$\rightarrow$T transition mutation (Robertson and A.Jones, 2000). Secondly, the environmental stressors that affect the epigenome and lead to inheritance of phenotypes may be mutagenic themselves e.g. vinclozolin (Skinner et al., 2015). Genetic background effects also influence whether phenotypes and/or epigenetic changes are associated with specific environmental stressors (Guerrero-Bosagna et al., 2012). This may be due to differences in interactions between the epigenome and underlying genome. Only a few epigenetic inheritance studies have taken take into account potential genetic effects (Oey et al., 2015; Skinner et al., 2015). Altogether, how the epigenome interacts with the genome needs to be explored further in models of epigenetic inheritance.

#### 1.1.4.7 Confounding influences

Investigations into the mechanisms of TEI need to be well designed to avoid confounding factors. Studies of inheritance through the maternal line can be confounded by the intrauterine environment and maternal care behaviours. For example, stressful male-female mating interactions could impact the stress hormones embryos are exposed to *in utero* and the care-giving behaviours experienced from their mothers (Champagne and Meaney, 2006; Bohacek and Mansuy, 2017). Female mate choice can also impact offspring outcomes, with females investing less in pups sired by poor quality non-preferred males (Drickamer et al., 2003). Cross fostering, embryo transfer and assisted reproductive technologies can be used to rule out these effects (Bohacek and Mansuy, 2017). It is for these reasons that most studies of TEI focus on paternal inheritance, as males were thought to pass solely genetic and epigenetic information to their offspring. Yet paternal inheritance itself is not without confounders. For example mate choice effects, seminal fluid composition or perhaps even the microbiome could influence offspring outcomes (Watkins et al., 2018; Heard and Martienssen, 2014). Additionally, litter and cage effects need to be considered as littermates tend to be more similar than non-littermates (Carone et al., 2010). Many reports to date are unable to exclude these confounding factors, although the field is beginning to consider them more thoroughly.

### 1.1.5 The relevance of TEI to human populations

There is much circumstantial evidence for intergenerational inheritance in human populations. However, very few epigenetic studies have been completed. TEI is difficult to study in human populations due to long generation times, genetic diversity and variable environmental conditions. Retrospective analyses form most of the body of evidence for intergenerational inheritance, with most studies only observing parent-child transmission. The seminal works in the field of human intergenerational inheritance explore the role of the Dutch Hunger Winter, a famine in Amsterdam from October 1944 to May 1945, on offspring health (Painter et al., 2008). *In utero* exposure to famine, particularly during the first trimester, resulted in reduced infant birth weights and metabolic disorders in the F1 offspring (Painter et al., 2008; Ravelli et al., 1999).

Remarkably, F2 adult offspring of men exposed *in utero* had increased BMIs compared to offspring of unexposed fathers (Veenendaal et al., 2013). Women exposed to famine *in utero* had first-born babies with increased birth weights (Lumey et al., 1995) and their granddaughters had higher mean ponderal indices (relationship between weight and length) at birth compared to grandsons (Painter et al., 2008). The effect of diet on offspring health was studied further using accurate records of food consumption for the population of Overkalix, Sweden (Pembrey, 2010). Fluctuations in food supply during childhood (prior to puberty) can impact offspring health in a sex-specific manner (Bygren et al., 2014). For instance, abundant food during a paternal grandfathers childhood resulted in diabetes and reduced survival of his grandchildren (Kaati et al., 2002), whereas paternal grandmothers had granddaughters with an increased risk of cardiovascular disease (Bygren et al., 2014). Recent studies have also suggested that air pollutants (Robledo et al., 2015), post-traumatic stress disorder (Nadler et al., 1985; Yehuda et al., 2001) and smoking (Miller et al., 2014) can also impact the health of an exposed individual's children and grandchildren.

However, our mechanistic understanding of human intergenerational inheritance is limited. Some studies have shown locus specific DNA methylation changes in offspring of exposed individuals. For example, *in utero* famine exposure correlated with decreased DNA methylation in the regulatory control region of the growth gene *IGF2* in blood samples (Heijmans et al., 2008). Similarly, genome-wide DNA methylation analysis of individuals exposed to famine *in utero* found a small increase (0.7-2.7%) in methylation at specific CpG sites associated with genes involved in growth, development, and metabolism in whole blood samples (Tobi et al., 2015). Furthermore, cigarette smoking has been shown to alter sperm DNA methylation patterns (Jenkins et al., 2017).

Altogether, these studies suggest that intergenerational inheritance of increased disease risk does occur in human populations and that epigenetic factors may play a role. However, much further work will be required to achieve a full understanding whether true TEI occurs in human populations and the mechanisms behind it. Animal models are likely to play a vital part in this research. Developing a full understanding of the phenomenon of TEI could

have important implications for predicting and preventing human disease in the future.

## 1.2 Folate, health and disease

Folate is a vitamin that has important roles in health and disease (Silva et al., 2017). Folate deficiency in humans is classically associated with megaloblastic anaemia (Herbert, 1962). It has been known for many decades that folate is important for foetal development, most famously in the prevention of neural tube defects (Group, 1991). In the United States, since 1998 there has been mandatory fortification of grain products with folic acid. This has coincided with a decrease in the incidence of birth defects and is often heralded as a preventative medicine and public health success story (Crider et al., 2011). However despite this, very little is known about the molecular mechanisms underpinning the influence of folate on health and disease.

### 1.2.1 Folate metabolism

Folate cannot be synthesised *de novo* in mammals but rather it is obtained from the diet (Matherly, 2001). Folic acid is the synthetic form of folate used for oral supplementation and food fortification (Silva et al., 2017). Folate and methionine metabolism are integrated as part of one-carbon metabolism (Figure 1.5). Folate is absorbed by the intestine and transported into cells via folate receptors (e.g. FOLR1) or folate transporters (e.g. SLC19A1) (Ratnam et al., 1989). Once internalised folate is reduced to dihydrofolate (DHF). DHF is converted to tetrahydrofolate (THF) by Dihydrofolate reductase and subsequently converted to 5-10-methylene-tetrahydrofolate by Serine hydroxymethyltransferase (Xu and Sinclair, 2015). 5-10-methylene-tetrahydrofolate is then reduced to 5-methyltetrahydrofolate (5-methyl-THF) by Methylenetetrahydrofolate reductase (MTHFR) (Zhu et al., 2014). 5-methyl-THF is the main circulating form of folate (Zhu et al., 2014) (Figure 1.5). A methyl group from 5-methyl-THF is then transferred to homocysteine by Methionine synthase (MTR, also known as MS). This forms methionine and THF (Shane and Stokstad, 1985) (Figure 1.5). MTR is activated by Methionine synthase reductase (MTRR) through the

reductive methylation of its cofactor, vitamin B12 (Yamada et al., 2006). Methionine is converted to S-adenosylmethionine (S-AdoMet or SAM) by Methionine adenosyltransferase (MAT) enzymes (Xu and Sinclair, 2015). SAM is the sole methyl donor for cellular substrates including proteins, RNA and DNA (Xu and Sinclair, 2015; Singh and Jaiswal, 2013). Beyond cellular methylation, one-carbon metabolism is also required for DNA synthesis. A methyl group from 5-10-methylene-THF is also used in the conversion of deoxyuridine monophosphate (dUMP) to deoxythimidine monophosphate (dTMP) by Thymidylate synthetase (Xu and Sinclair, 2015). dTMP is the DNA base thymidine. Due to its key role in the generation of both SAM and thymidine, folate metabolism is required for maintaining the epigenetic landscape and genetic stability of a cell.

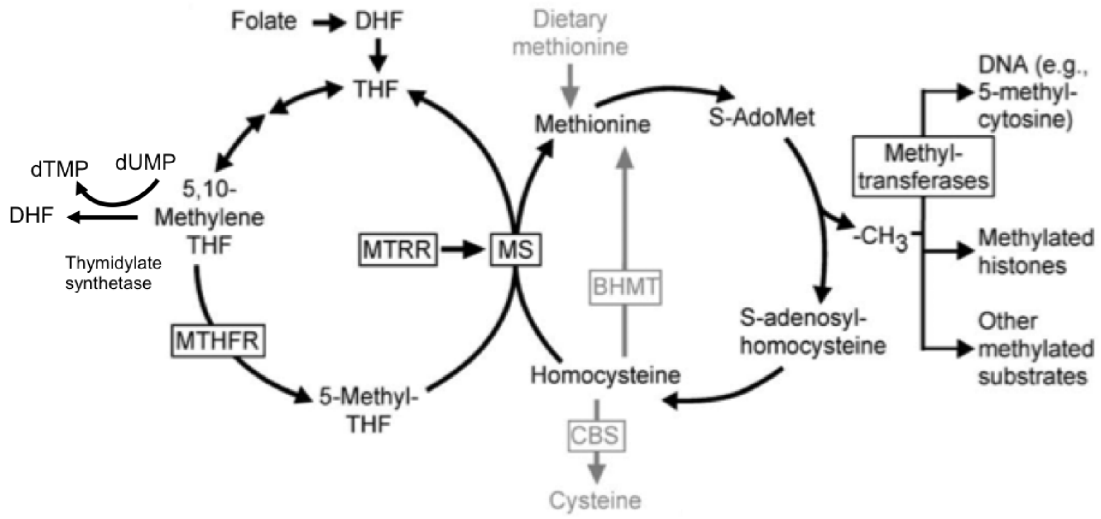



Figure 1.5: **The folate and methionine cycles.** MTHFR converts 5,10-methylene-THF to 5-Methyl-THF. MTR (MS), activated by MTRR, transfers a methyl group from 5-methyl-THF to homocysteine to form methionine and THF. Methionine is the precursor of SAM, which is a methyl donor for cellular substrates. Thymidylate synthetase converts dUMP to dTMP, for DNA synthesis, using 5,10-methylene-THF as a methyl donor. Figure adapted from Padmanabhan et al., 2013.

### 1.2.2 Folate, disease and offspring health

Folate has been linked to a wide spectrum of diseases including allergy, cancer, metabolic disorders, neurological and behavioural problems and developmen-

tal abnormalities (reviewed in Silva et al. (2017)). Importantly, abnormal folate status or disrupted folate metabolism have not only been linked to disease in the exposed generation but also to disease in their offspring. The roles of both dietary folate deficiency and supplementation have been explored in human populations and animal models. Rodent models in which folate metabolism is disrupted by genetic mutations have been developed and human populations with polymorphisms in key folate metabolism enzymes have been studied. The literature regarding folate, disease and offspring health is extensive but often lacking consensus or even conflicting.

Here I highlight some key studies examining the role of folate on offspring health, primarily considering foetal growth and development. These are the primary phenotypes observed in a model of abnormal folate metabolism resulting from a mutation in the gene *Mtrr* (*Mtrr*$^{gt}$) that is the focus of this thesis. I also highlight studies examining the role of folate in adult disease, focusing on the effects on male fertility and testes function. The evidence for the molecular mechanisms by which folate may influence offspring health is also considered. These are themes that are explored further in the *Mtrr*$^{gt}$model in this thesis.

#### 1.2.2.1 Folate and developmental abnormalities

*In utero* exposure to maternal dietary folate deficiency leads to embryonic and pre-weaning phenotypes. Folate is primarily known for its role in neural tube development. Folate supplementation in the periconceptual period has been shown to drastically reduce the incidence of neural tube defects in human populations (Czeizel and Duds, 1992; Group, 1991). During human pregnancy, low concentrations of folate and hyperhomocysteinemia (increased serum homocysteine concentration) are also associated with increased risks of pre-term delivery, low birth weight, preeclampsia, foetal growth retardation and placental abruption or infarction (Ray and Laskin (1999), reviewed in Scholl and Johnson (2000)). A range of similar embryonic phenotypes are seen in mice fed a low folate diet or with disrupted folate metabolism. These include developmental delay, growth restriction, congenital malformations (including neural tube defects), placental defects and increased reproductive failure (e.g. miscarriage or post-partum litter death) (Tee and Reinberg, 2014; Momb et al.,

2013; Deng et al., 2008; McKay and Mathers, 2015). However, the prevalence of these defects varies across models. Knock-out mutations in folate metabolism genes, e.g. *Mthfd1* (MacFarlane et al., 2009) and *Mtr* (Swanson et al., 2001), are homozygous lethal. In mice, dietary folate deficiency alone is not sufficient to cause neural tube defects (Heid et al., 1992). Folate is also vital for pre-implantation embryo development. Evidence from *in vitro* culture, demonstrates that depletion of folate, by the addition of the folate antagonist methotrexate, inhibits the development of mouse embryos beyond the first cell division and bovine and ovine embryos beyond day eight (Kwong et al., 2010; O'Neill, 1998). Intriguingly, very high doses of folic acid in mice have also been reported to negatively impact embryo development. Supplementation with high doses of folic acid during pregnancy was associated with embryonic loss, developmental delay and heart defects (septal defects and decreased ventricular wall thickness) at embryonic day (E) 14.5 (Mikael et al., 2012). Together, these data suggest that folate is required at an optimal level throughout pregnancy to support normal embryonic development.

Folate deficiency in males has also been shown to have effects on embryonic development. Paternal folate deficiency, from prior to breeding until weaning, led to an increase in birth defects, including craniofacial defects, placental phenotypes and musculoskeletal abnormalities, in the offspring (Lambrot et al., 2013). Abnormal folate metabolism resulting from the $Mtrr^{gt}$mutation in grandfathers has also been shown to increase the incidence of congenital malformations (including heart and neural tube defects) in their wildtype grandprogeny (Padmanabhan et al., 2013). However, in comparison to the wealth of studies considering the role of folate during pregnancy, there is a sparsity of studies investigating paternal folate status and offspring health. Further research into the role of folate deficiency or abnormal folate metabolism in males is required. Altogether, these data highlight an important role for both maternal and paternal folate status in ensuring normal embryonic development.

#### 1.2.2.2 Folate and adult disease

Dietary folate deficiency and disrupted folate metabolism are also associated with a range of adult diseases including hepatic, haematological, cardiac,

metabolic and neurological phenotypes (Chanson et al., 2005; Pogribny et al., 2013; Eckart et al., 2013; Christensen et al., 2013; Ash et al., 2013; Padmanabhan et al., 2018). Folate deficiency and abnormal folate metabolism also impact male fertility and testes function in mice and humans (Lambrot et al., 2013; Murphy et al., 2011; Wong et al., 2002; Ebisch et al., 2003; Boxmeer et al., 2009; Singh and Jaiswal, 2013). For example, the C677T mutation in the gene *MTHFR* is associated with reduced fertility in men (Bezold et al., 2001; A et al., 2007). Knocking out the gene *Mthfr* in mice leads to a testes phenotype, though the severity of the defect depends upon the genetic background of the mouse. For instance, the *Mthfr*$^{-/-}$ loss of function mutation in mice with a BALB/c genetic background results in fewer proliferating germ cells in the early postnatal period and infertility in adulthood (Chan et al., 2010). In contrast, a milder testes phenotype is observed in C57Bl/6 mice that are homozygous for the *Mthfr*$^{-/-}$ mutation (Chan et al., 2010). Sperm count is reduced and abnormal testes morphology is observed despite normal fertility (Chan et al., 2010). Similarly, in some human populations, the *Mtrr* A66G mutation is associated with reduced fertility (Lee et al., 2006). However, in different populations those with the *MTRR* A66G mutation have normal fertility (Ni et al., 2015; Montjean et al., 2011; Mfady et al., 2014). This suggests that genetic background may influence the impact of the *MTRR* A66G mutation. Additionally, in men, infertility correlated with reduced folate concentrations in serum compared to fertile individuals (Murphy et al., 2011). Indeed, dietary supplementation with folic acid improves fertility in subfertile men (Wong et al., 2002; Ebisch et al., 2003), though the molecular mechanism of this recovery is unclear. Furthermore, dietary folate deficiency in mice is associated with delayed onset of meiosis during spermatogenesis (Lambrot et al., 2013). Abnormalities in spermatogenesis and fertility may be associated with genetic and epigenetic aberrations in sperm which have the potential to impact offspring development (Schulte et al., 2010; Kobayashi et al., 2007).

Importantly, *in utero* exposure to folate deficiency also results in adult phenotypes in the offspring. Adult offspring from folate-depleted dams have a spectrum of metabolic dysfunction (McKay et al., 2016; Barnett et al., 2015). For example, in sheep, adult offspring of folate-deficient mothers (fed a methyl-

deficient diet lacking vitamin B12, folate and methionine) had increased body weight and adiposity, insulin resistance and altered immune responses to antigenic challenge (Sinclair et al., 2007). These adult phenotypes in the offspring of folate-deficient mothers have the potential to impact the development of their offspring. In such a manner, the negative effect of folate deficiency experienced in one generation may be perpetuated.

Studies investigating the effects of folate deficiency and/or supplementation and disrupted folate metabolism have not revealed a clear picture regarding how folate impacts the health of both exposed individuals and their offspring. The phenotypes observed vary in penetrance, frequency and occurrence between models. Discrepancies arise between dietary and genetic studies, genetic studies with mutations in different genes and even between studies using seemingly similar dietary paradigms. For example, SAM levels were decreased in some models of abnormal folate metabolism e.g. *Mtrr*$^{gt/gt}$mice (Elmore et al., 2007) but not others e.g. *Mthfd1*$^{gt/+}$ mice (MacFarlane et al., 2009). Disparities between dietary and genetic studies, and between genetic studies with mutations in different folate metabolism enzymes, may be explained by the differing biochemical outcomes resulting from disruption of the folate and methionine cycles at different points. Meanwhile, a lack of consistency with respect to the severity of folate depletion or supplementation or the duration and timing of the dietary manipulation may account for conflicting findings in dietary manipulation studies. Furthermore, studies often combine folate deficiency with other challenges e.g. high fat diet, concomitant selenium deficiency, genetic variants in folate metabolism enzymes. While vital for studying gene-nutrition interaction and perhaps more accurately mimicking human populations, these models make it difficult to tease apart the molecular mechanisms by which folate acts to influence offspring health. Rigorous folate-specific studies in animal models and human populations will be required to further our understanding. However, the variability observed between models and conflicting nature of the folate literature may reflect stochasticity in the underlying mechanism by which folate deficiency or abnormal folate metabolism causes disease.

### 1.2.3 Mechanisms of folate-related disease

Due to its key role in the generation of both thymidine and SAM, folate metabolism is required for both DNA synthesis and cellular methylation reactions (Figure 1.5). Disruption of these key processes is thought to account for the array of negative outcomes associated with folate deficiency or disruptions to folate metabolism (Xu and Sinclair, 2015). However, the exact mechanisms by which folate deficiency, supplementation or abnormal metabolism act to cause phenotypes is still not well understood.

Many studies have shown both global and locus-specific DNA methylation alterations resulting from folate deficiency and supplementation *in utero* (McKay et al., 2011; Richmond et al., 2018; Sie et al., 2013). It has been suggested that DNA methylation changes resulting from abnormal folate metabolism may be stochastic, thus accounting for the spectrum of phenotypes observed (Padmanabhan et al., 2013). However, many studies were unable to identify DNA methylation changes following folate deficiency. For example, female rats fed a folate-deficient diet from two weeks prior to breeding had no changes in global DNA methylation in maternal or foetal liver tissue (Maloney et al., 2007). Unfortunately most studies examining DNA methylation in the context of offspring phenotypes only assess DNA methylation in offspring somatic tissues. They do not investigate DNA methylation in the parental germline, thus are unable to determine if epigenetic patterns are being directly inherited or if an abnormal *in utero* environment leads to dysregulation of DNA methylation in the offspring. A couple of exceptional studies have however identified abnormalities in sperm DNA methylation patterns in folate-deficient male mice (Lambrot et al., 2013) or males deficient for the folate metabolism enzyme MTHFR (Chan et al., 2010). Overall, these findings suggest that folate can influence the epigenome. However, all studies to date have failed to provide substantive evidence that any DNA methylation changes observed directly lead to the phenotypes reported.

Other studies have focused on the role of the folate cycle in the production of thymidine and the effects disrupting this may have on DNA synthesis and cell proliferation. It has been speculated that the rapid cell division in embryogenesis makes it vulnerable to low folate conditions (Imbard et al.,

2013). Indeed, murine/bovine/ovine embryo development *in vitro* can be halted using methotrexate (a folate antagonist). This can be partially rescued by supplementation of the culture medium with thymidine (Kwong et al., 2010; O'Neill, 1998). This suggests that development was partly halted as a result of defective DNA synthesis (Kwong et al., 2010). Additionally, folate deficiency has been associated with genetic instability in mice and cultured human lymphocytes. Folate-deficient BALB/c mice showed increased DNA fragmentation index and increased mutation frequency at expanded simple tandem repeats (Swayne et al., 2012). Cultured human lymphocytes exposed to folate-deficient culture conditions have DNA strand breakage and uracil misincorporation (Duthie and Hawdon, 1998). Altogether, this suggests that folate deficiency can negatively impact genetic stability and DNA synthesis. This could clearly contribute to diverse disease phenotypes. Phenotypic variability may result from differences in the degree of genetic instability or the location of genetic mutations arising from it. However, the approaches used in many studies have generally been low resolution, often identifying the presence of DNA damage but not where in the genome this occurs. Identifying mutations in genes associated with the phenotypes observed following folate deficiency will be required to specifically demonstrate a genetic cause.

It has also been postulated that the effects of folate deficiency may be mediated by changes in the individual's metabolic signalling or immune system. It has been suggested that dysregulation of the inflammatory response may be involved in mediating negative pregnancy outcomes associated with hyperhomocysteinemia (Mikael et al., 2013). In *Mthfr*$^{+/-}$ mice fed folate deficient diets the expression of *ApoAI*, a modulator of immune function, was increased in liver and offspring placenta (Mikael et al., 2013). Additionally, hyperhomocysteinemia may lead to elevated inflammatory cytokine levels (Forges et al., 2007). Furthermore, nitric oxide (NO) signalling, which has been implicated in neurodegenerative diseases and plays an important role in many aspects of sperm functionality, may be disrupted as a result of hyperhomocysteinemia (Singh and Jaiswal, 2013; Knott and Bossy-Wetzel, 2009). Reactive oxygen species (ROS) levels are known to be elevated by abnormal folate metabolism or availability, and are associated with oxidative stress (Ho et al., 2003). Oxida-

tive stress in sperm can cause DNA damage, affect the epigenetic reprogramming that occurs in the early embryo and thus consequently impair embryo development (Wyck et al., 2018). As such these factors may act downstream of or in concert with epigenetic and genetic factors to contribute to the phenotypes associated with folate deficiency. It will be necessary to investigate multiple mechanistic candidates in models of abnormal folate metabolism or folate deficiency to get a full molecular understanding of how folate is linked to disease.

## 1.3 The *Mtrr*$^{gt}$ model of TEI

I study the phenomenon of TEI in a mouse model of abnormal folate metabolism. In this model, the progression of the folate cycle is disrupted by a mutation in the gene *Methionine synthase reductase (Mtrr)* (Elmore et al., 2007; Padmanabhan et al., 2013; Deng et al., 2008). The MTRR enzyme is required for the progression of the folate and methionine cycles (Figure 1.5). MTRR is a dual flavoprotein and contains an N-terminal flavin mononucleotide (FMN)-binding domain and a C-terminal flavin adenine dinucleotide (FAD)-binding domain. MTRR catalyses NAPDH dependent reduction of MTR, transferring a electron through covalently bound FMN and FAD cofactors (Elmore et al., 2007).

The *Mtrr* mutation was generated by insertion of a gene-trap (gt) vector into intron 9 of the *Mtrr* locus (*Mtrr*$^{gt}$) (Padmanabhan et al., 2013; Elmore et al., 2007). Gene-trap insertion was performed in 129/P2 embryonic stem cells. These were injected into C57Bl/6 blastocysts. Upon germline transmission the *Mtrr*$^{gt}$ allele was maintained by backcrossing, at least 8 generations, into the C57Bl/6 strain (Elmore et al., 2007). The gene-trap insertion into intron 9 generates a fusion protein of MTRR and $\beta$-galactosidase/neomycin phosphotransferase and removes the C-terminal FAD/NADPH domains of the MTRR protein (Elmore et al., 2007). The *Mtrr*$^{gt}$ mutation is hypomorphic owing to splicing out of the gene-trap, although this seems to occur to differing degrees in a tissue-specific manner (Elmore et al., 2007). *Mtrr* mRNA expression was 19-37% of *Mtrr*$^{+/+}$ levels in *Mtrr*$^{gt/gt}$ mice dependent on the tissue assessed (Padmanabhan et al., 2013). A similar decrease in MTRR protein levels was also observed (Elmore et al., 2007). *Mtrr*$^{gt/gt}$ mice are characterised by a adult

metabolic phenotype including hyperhomocysteinemia (Padmanabhan et al., 2013), increased liver 5-methyl-THF, reduced plasma methionine and reduced heart AdoMet/AdoHcy ratio (Elmore et al., 2007). *Mtrr*$^{gt}$mice also have late onset haematological abnormalities: *Mtrr*$^{gt/gt}$female mice display macrocytic anaemia and splenomegaly whereas *Mtrr*$^{gt/gt}$male mice were not anaemic but had erythrocytic macrocytosis and lymphopenia (Padmanabhan et al., 2018). Overall, the *Mtrr*$^{gt}$mutation leads to adult metabolic and haematological dysfunction and may affect other systems not yet fully characterised.

### 1.3.1 The *Mtrr*$^{gt}$model and developmental phenotypes

The *Mtrr*$^{gt}$mutation is also associated with developmental defects. A range of phenotypes were observed in litters of *Mtrr*$^{+/gt}$intercrosses at both at E10.5 and E14.5 (Padmanabhan et al., 2013; Deng et al., 2008). At E14.5, maternal *Mtrr* deficiency was associated with developmental delay, growth defects, reduced placental weight and an elevated frequency of congenital heart defects including ventral septal defects (Deng et al., 2008). At E10.5, in addition to the phenotypes observed at E14.5, neural tube defects (failed closure in the cranial or spinal cord regions), reversed heart looping, haemorrhage (placental and embryonic) and placental abnormalities (e.g. off-centred chorio-allantoic attachment) were observed (Padmanabhan et al., 2013). Importantly, phenotypes were seen in all embryonic genotypes including *Mtrr*$^{+/+}$embryos (Padmanabhan et al., 2013). This result suggested that parental exposure to the *Mtrr*$^{gt}$mutation was sufficient to cause developmental phenotypes. However, *Mtrr*$^{+/+}$intercrosses revealed females that had abnormal litters always had a *Mtrr*$^{+/gt}$grandparent (Padmanabhan et al., 2013). This suggested embryonic phenotypes were associated with grandparental not parental *Mtrr*$^{gt}$mutation (Padmanabhan et al., 2013).

Highly controlled genetic pedigrees were used to investigate this effect further (Figure 1.6). In the *Mtrr*$^{+/gt}$maternal grandparent pedigrees, a *Mtrr*$^{+/gt}$male or female (F0), respectively was crossed to a C57Bl/6 control mouse (Figure 1.6). The wildtype F1 daughter of this cross was mated to a C57Bl/6 male (Figure 1.6). Phenotypes were then assessed in the wildtype grandprogeny (F2) generation at E10.5. C57Bl/6 mice were used as controls as they have the same

genetic background as the *Mtrr*$^{gt}$mice but have been bred separately from the *Mtrr*$^{gt}$line. These pedigrees demonstrated that when either maternal grandparent was heterozygous for the *Mtrr*$^{gt}$allele (*Mtrr*$^{+/gt}$) the wildtype grandprogeny (F2) had an increased frequency of growth defects (e.g. intrauterine growth restriction, growth enhancement and developmental delay) and congenital malformations (e.g. abnormal placenta development, and heart and neural tube defects) that were not present in control litters (Padmanabhan et al., 2013). Embryo transfer experiments were performed to explore the effects of the F1 maternal uterine environment on embryo development. F2 generation blastocysts derived from a *Mtrr*$^{+/gt}$maternal grandparent were transferred to host mothers prior to implantation. This demonstrated that the maternal grandparental *Mtrr*$^{gt}$allele affected the grandprogeny phenotype in two distinct ways (Padmanabhan et al., 2013). Firstly, the grandparental *Mtrr*$^{gt}$allele resulted in an abnormal uterine environment in the wildtype F1 daughters which was responsible for growth phenotypes observed in the *Mtrr*$^{+/+}$F2 grandprogeny (Padmanabhan et al., 2013). These phenotypes were therefore rescued by embryo transfer. Secondly, the grandparental *Mtrr*$^{gt}$allele caused congenital malformations in the *Mtrr*$^{+/+}$grandprogeny independent of the maternal environment (Padmanabhan et al., 2013), thus these phenotypes were not rescued following embryo transfer. It was hypothesised that these phenotypes resulted from gametic epigenetic inheritance. This hypothesis was further supported by evidence that congenital malformations, but not growth defects, were observed in the F3 and F4 generations derived from a *Mtrr*$^{+/gt}$maternal grandparent (Padmanabhan et al., 2013).

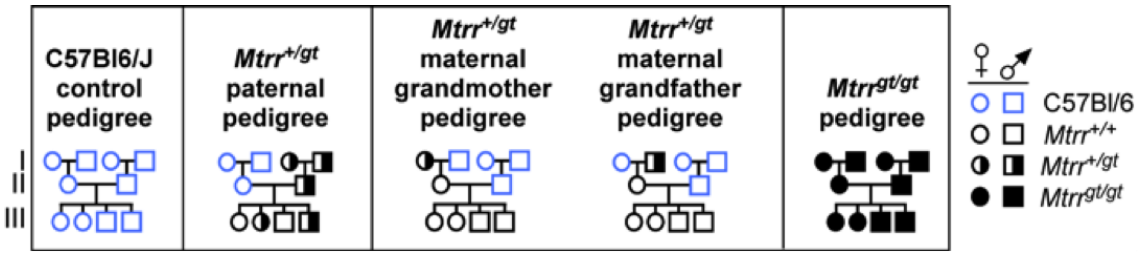



Figure 1.6: **Highly controlled genetic pedigrees used to study inheritance in the *Mtrr*$^{gt}$model.** The following pedigrees are shown: Control C57Bl/6 pedigree, *Mtrr*$^{+/gt}$paternal pedigree, *Mtrr*$^{+/gt}$maternal grandmother and *Mtrr*$^{+/gt}$maternal grandfather pedigrees and *Mtrr*$^{gt/gt}$pedigree. Circle: female, square: male. Blue outline: C57Bl/6 (control). Black outline: *Mtrr*$^{gt}$line; no fill: *Mtrr*$^{+/+}$, half fill: *Mtrr*$^{+/gt}$, complete fill: *Mtrr*$^{gt/gt}$. I, II and III indicate the generations (F0, F1 and F2 respectively). Figure provided by Erica Watson.

A similar pattern of inheritance of developmental abnormalities was observed in the wildtype grandprogeny derived from a *Mtrr*$^{+/gt}$maternal grandmother or a *Mtrr*$^{+/gt}$maternal grandfather. In both pedigrees, approximately 40-47% of wildtype grandprogeny conceptuses were phenotypically normal, 22% growth restricted, 8% developmentally delayed and 12-17% had congenital malformations (Padmanabhan et al., 2013). There was a significant increase in the levels of growth enhancement only in grandprogeny of the *Mtrr*$^{+/gt}$maternal grandfather pedigree (Padmanabhan et al., 2013). In both pedigrees, some litters were more sensitive to grandparental *Mtrr* deficiency than others. In some litters all embryos were normal, in some litters all conceptuses had developmental abnormalities and other litters had varying proportions of both phenotypically normal conceptuses and conceptuses with developmental abnormalities (Padmanabhan et al., 2013). Intriguingly, no phenotypes were observed in the F1 generation derived form a *Mtrr*$^{+/gt}$grandfather, however, phenotypes were observed in the F1 generation derived from a *Mtrr*$^{+/gt}$grandmother (Padmanabhan et al., 2013). The *Mtrr*$^{+/gt}$paternal pedigree was used to demonstrate that a paternal *Mtrr*$^{gt}$allele did not contribute to the inheritance of phenotypes (Figure 1.6). The offspring of *Mtrr*$^{+/gt}$males (derived from *Mtrr*$^{+/gt}$parents) crossed to a C57Bl/6 female had no congenital malformations or growth defects at E10.5 (Padmanabhan et al., 2013). This was used as evidence to suggest that neither the paternal nor paternal grandparental *Mtrr*$^{gt}$allele influenced

developmental phenotypes. However, this cannot be totally excluded as phenotypes may not be expected in these embryos, as they are similar to the F1 generation of the maternal grandfather pedigree (Padmanabhan et al., 2013). The presence of phenotypes should be assessed in the F3 generation offspring derived from the F2 generation of the *Mtrr*$^{+/gt}$paternal pedigree to assess the paternal lineage.

Importantly there was no sexual dimorphism in the developmental abnormalities seen to result from either intrinsic or ancestral *Mtrr*$^{gt}$allele, although assessment of more individuals may make this become apparent (Padmanabhan et al., 2017). However, both female and male phenotypically normal conceptuses derived from *Mtrr*$^{+/gt}$maternal grandparents had increased placental efficiency (characterised by increased embryo weight and reduced placental weight) in each subsequent generation up to the F4 generation. This effect led to excessive growth in female but not male embryos, such that female *Mtrr*$^{+/+}$F4 embryos weighed more than controls (Padmanabhan et al., 2017). This suggested a subtle sexually dimorphic response to ancestral *Mtrr*$^{gt}$deficiency in phenotypically normal embryos (Padmanabhan et al., 2017). Further characterisation of ancestral exposure to the *Mtrr*$^{gt}$allele revealed an adult haematological defect in *Mtrr*$^{+/+}$daughters derived from *Mtrr*$^{+/gt}$parents, in a parent specific manner (Padmanabhan et al., 2018). The wildtype daughters of *Mtrr*$^{+/gt}$females displayed normocytic anaemia while *Mtrr*$^{+/+}$daughters of *Mtrr*$^{+/gt}$males exhibited erythrocytic microcytosis not associated with anaemia (Padmanabhan et al., 2018). Together these data suggest that the *Mtrr*$^{gt}$allele can affect offspring in a sex and parent-of-origin specific manner in certain circumstances. This makes understanding the mechanism by which the *Mtrr*$^{gt}$mutation influences offspring phenotype particularly challenging.

### 1.3.2 The *Mtrr*$^{gt}$model and dysregulation of DNA methylation

Folate metabolism is required for cellular methylation reactions, therefore disrupting folate metabolism would be expected to lead to DNA hypomethylation (Jacob et al., 1998; Ghandour et al., 2002). Stochastic changes in DNA methylation were hypothesised to be responsible for the spectrum of phenotypes observed in the *Mtrr*$^{gt}$model. Widespread tissue-specific dysregulation

of DNA methylation was observed in *Mtrr*$^{gt}$mice (Padmanabhan et al., 2013). Global DNA hypomethylation, measured using a DNA methylation quantification kit, was reported in *Mtrr*$^{+/+}$, *Mtrr*$^{+/gt}$and *Mtrr*$^{gt/gt}$non-pregnant uteri and livers (∼40% to 75% of C57Bl/6 levels) (Padmanabhan et al., 2013). Similarly, the *Mtrr*$^{+/+}$F2 generation placentas at E10.5 of the maternal grandfather and maternal grandmother pedigrees were also hypomethylated regardless of phenotype (∼50% to 60% of C57Bl/6 levels) (Padmanabhan et al., 2013). Global hypomethylation was also reported in the placentas at E10.5 of the *Mtrr*$^{+/+}$F1 generation females of the *Mtrr*$^{+/gt}$maternal grandfather pedigree (derived from a *Mtrr*$^{+/gt}$father and C57Bl/6 mother) and in the placentas of the *Mtrr*$^{+/+}$grandprogeny of the *Mtrr*$^{+/gt}$paternal pedigree. This was despite no developmental phenotypes being observed in these conceptuses (Padmanabhan et al., 2013). This suggested that dysregulation of DNA methylation precedes the appearance of phenotypes in the *Mtrr*$^{gt}$model. Intriguingly, global DNA methylation levels in all embryos at E10.5, including *Mtrr*$^{gt/gt}$embryos and *Mtrr*$^{+/+}$F2 generation embryos from the maternal grandmother or maternal grandfather pedigrees, were comparable to C57Bl/6 controls regardless of genotype or phenotype (Padmanabhan et al., 2013). However, global DNA methylation levels were determined using a DNA methylation quantification kit based method, which may not have provided the most accurate measure of DNA methylation levels. More sensitive techniques for assessing global DNA methylation levels (e.g. mass spectrometry) should be used to study global DNA methylation in the *Mtrr*$^{gt}$model.

Furthermore, as a model for the epigenetic stability of the genome as whole, locus specific methylation at imprinted genes was assessed in *Mtrr*$^{+/+}$grandprogeny placentas at E10.5 using bisulfite pyrosequencing. Dysregulation of DNA methylation, including hypomethylation and hypermethylation, was observed at imprinting control regions (Padmanabhan et al., 2013). The degree of dysregulation of DNA methylation at CpG sites in imprinted loci correlated with the severity of the phenotype (Padmanabhan et al., 2013). Overall the DNA methylation changes at individual CpG sites were relatively small (generally less than ∼10%) (Padmanabhan et al., 2013). However, imprinting control regions may not be representative of the genome as a whole, therefore genome-wide ap-

proaches will be required to confirm whether the dysregulation of DNA methylation occurs more generally. Overall, this suggests that the *Mtrr*$^{gt}$mutation, or ancestral *Mtrr*$^{gt}$mutation, can disrupt DNA methylation patterns in a tissue-specific manner.

I speculate that stochastic dysregulation of DNA methylation across the genome may be responsible for congenital malformations observed in the *Mtrr*$^{gt}$model and may be inherited via the germline. This hypothesis is supported by evidence that congenital malformations persist after embryo transfer, thus are independent of the maternal environment, and are observed in the F3 and F4 generations derived from a *Mtrr*$^{+/gt}$maternal grandparent (Padmanabhan et al., 2013). Additionally, a role for DNA methylation is supported by evidence that knocking out DNA methyltransferases results in DNA demethylation defects and phenotypes similar to those observed in the *Mtrr*$^{gt}$model, such as embryonic lethality by E10.5, neural tube and placenta defects (Hata et al., 2002).

## 1.4 Aims and rationale

The objective of the research presented in this thesis was to explore the potential mechanisms responsible for transgenerational inheritance of congenital malformations in the *Mtrr*$^{gt}$model. Transgenerational inheritance of developmental phenotypes was demonstrated when either maternal grandparent was a carrier of the *Mtrr*$^{gt}$mutation (Padmanabhan et al., 2013). I focus on inheritance in the *Mtrr*$^{+/gt}$maternal grandfather pedigree to avoid the confounding influence of the maternal environment in the F0 generation. As part of one-carbon metabolism the folate cycle is vital for DNA synthesis, thus genetic stability, and cellular methylation reactions, thus epigenetic patterns (Xu and Sinclair, 2015). Folate metabolism has also been shown to play important roles in spermatogenesis and male fertility (Chan et al., 2010). Therefore, I proposed that abnormalities in three key factors in *Mtrr* deficient males could influence offspring phenotype; 1) gross sperm morphology, 2) genetic stability and/or 3) epigenetic modifications.

Firstly, I assessed gross testes morphology, spermatogenesis and mature sperm parameters to determine if any abnormalities resulted from the *Mtrr*$^{gt}$mutation

(Chapter 3). I determined if the expression of the DNA and RNA methylation machinery was disrupted in testes and epididymides as this could contribute to defects in the establishment and/or maintenance of DNA methylation profiles in sperm.

Secondly, I investigated genetic stability in the *Mtrr*$^{gt}$model using whole genome sequencing (Chapter 4). This analysis included consideration of the genetic background of the region flanking the *Mtrr* gene itself.

Thirdly, I examined the epigenome, identifying alterations in DNA methylation patterns (Chapter 5) and sncRNAs (Chapter 7) in sperm. I explored whether DNA methylation changes that I identified in sperm from *Mtrr*$^{+/gt}$males could be inherited by the F1 and F2 generations and if they impact gene expression in these offspring (Chapter 6).

These analyses are the first steps to fully elucidating the mechanism of TEI in the *Mtrr*$^{gt}$model. Understanding this phenomenon may have important implications and relevance to disease inheritance in human populations in the future.

# Chapter 2

# Methods

## 2.1 Animals

Mice were housed at University of Cambridge. This research was regulated under the Animals (Scientific Procedures) Act 1986 Amendment Regulations 2012 following ethical review by the University of Cambridge Animal Welfare and Ethical Review Body. *Mtrr*$^{gt}$ mice were generated as previously described (Elmore et al., 2007). Briefly, insertion of a gene-trap vector, containing a $\beta$-galactosidase/neomycin phosphotransferase cassette, into intron 9 of the *Mtrr* gene was performed in 129/P2 embryonic stem cells (ESCs) (Elmore et al., 2007). ESCs containing the *Mtrr*$^{gt}$allele were injected into C57Bl/6 blastocysts (Elmore et al., 2007). On germline transmission, the *Mtrr*$^{gt}$allele was maintained by backcrossing it into a C57Bl/6 genetic background for at least 8 generations (Elmore et al., 2007; Padmanabhan et al., 2013). *Mtrr*$^{gt/gt}$ mice were derived from *Mtrr*$^{gt/gt}$ intercrosses. *Mtrr*$^{+/+}$ and *Mtrr*$^{+/gt}$ mice were derived from *Mtrr*$^{+/gt}$ intercrosses. *Mtrr*$^{+/+}$, *Mtrr*$^{+/gt}$and *Mtrr*$^{gt/gt}$ mice are distinguishable from each other by a three-primer polymerase chain reaction (PCR) using primers flanking the gene-trap insertion combined with a primer for the gene-trap vector (Padmanabhan et al., 2013). C57Bl/6 mice were used as controls throughout since the *Mtrr*$^{gt}$ mutation was backcrossed into the C57Bl/6 background (Padmanabhan et al., 2013). Control C57Bl/6 mice were housed separately to the *Mtrr* lineage and had never been exposed to the *Mtrr*$^{gt}$mutation. Mice were fed a normal chow diet (Rodent No. 3 breeding chow, Special Diet Services, Essex, UK) *ad libitum* from weaning.

## 2.2 Embryo and placenta collections

All conceptuses were collected at E10.5. Noon on the day the vaginal plug was detected was considered E0.5. Embryos were scored for phenotypes as previously published (Padmanabhan et al., 2013), photographed, measured for crown-rump length and snap frozen in liquid nitrogen. Placentas were snap frozen whole or trophoblast cells were separated from the maternal decidual layer by manual dissection. 129/P2 embryos at E10.5 were collected by E. Watson and N. Padmanabhan.

## 2.3 Testes collection and processing

Testes were collected from 16-20 week old male mice proven fertile by mating to a female mouse. Each testis was weighed at the time of dissection. One testis per individual was snap frozen for molecular analysis and the other was fixed in 4% paraformaldehyde in 1x phosphate buffered saline (PBS) overnight at 4°C. For paraffin embedding, fixed testes were washed in 1x PBS, dehydrated in ethanol, cleared in Histo-Clear II (National Diagnostics) and embedded in paraffin wax. Paraffin blocks were sectioned to 7$\mu$m in a transverse orientation and stained with haematoxylin and eosin (H & E) or Periodic Acid-Schiff (PAS) (Sigma-Aldrich) using standard practices. Histological sections were imaged with a Nanozoomer 2.0RS digital slide scanner (Hamamatsu Photonics UK Ltd) and processed with NDP.view2 viewing software (Hamamatsu Photonics). Measurement of testes morphological parameters was performed using ImageJ (64-bit) software (NIH, USA).

## 2.4 Immunohistochemistry

Testes sections were de-waxed, rehydrated in ethanol and washed in 1x PBS. To quench endogenous peroxidase activity, sections were incubated with 3% hydrogen peroxide in 1x PBS for 30 minutes at room temperature (RT). Antigen retrieval was performed using trypsin tablets (Sigma-Aldrich) for 10 minutes according to the manufacturers instructions. Tissue sections were incubated with blocking serum (5% donkey serum [Sigma-Aldrich], 1% bovine serum albumin [Sigma-Aldrich] in 1x PBS) for 1 hour at RT. Tissue sections were then incubated with primary antibody in blocking serum overnight at 4°C at the following dilutions: 1:100, anti-MTRR (cat no: 26994-1-AP, ProteinTech, UK); 1:50, anti-MTHFR (sc-17079, Santa Cruz Biotechnology, USA ); 1:50, anti-SLC19A1 (sc-47358, Santa Cruz Biotechnology, USA); and 1:50, anti-FOLR1 (sc-28997, Santa Cruz Biotechnology, USA). Tissue sections were washed in 1x PBS, and then incubated with horseradish peroxidase-conjugated donkey anti-rabbit IgG (ab6802, Abcam, UK) or donkey anti-goat IgG (ab6885, Abcam, UK) diluted to 1:300 in blocking serum for one hour at RT. DAB (3,3-diaminobenzidine) chromagen substrate (Abcam) was used to perform the

colorimetric reaction, according to the manufacturers instructions. Histological sections of testes incubated in blocking serum without the primary antibody or the secondary antibody were included as controls. Sections were counterstained with haematoxylin, dehydrated in ethanol, cleared and coverslip-mounted using DPX Mountant (Sigma-Aldrich).

## 2.5 Assessment of spermatogenesis through the staging of seminiferous tubules

Assessment of spermatogenesis was performed by staging seminiferous tubules on PAS-stained histological sections of testes. At least eight sections from three males per genotype were analysed. At least 40 randomly selected seminiferous tubule cross-sections were assessed per section. Stages were assigned as previously described in detail (Carrell and Aston, 2013). Briefly, stages I-VIII were characterised by the presence of round and elongated spermatids and the development and migration of the acrosome over the nucleus. Stage VII and VIII were identified by the migration of the elongated spermatids to the seminiferous tubule lumen. Stages IX-XI were defined by the condensation of chromatin and nuclear shape change in spermatids. Stage XII was identified by the presence of characteristic meiotic figures in spermatocytes.

## 2.6 Testosterone concentration

Peripheral blood was collected using a 26-gauge needle through direct cardiac puncture after cervical dislocation. Blood was allowed to clot for 30 minutes at RT before centrifugation to separate serum from the blood cells (2000xg, 10 minutes, 4°C). Serum was stored at -80°C. Serum testosterone levels were assessed using a testosterone enzyme-linked immunosorbent assay (ELISA) (EIA-1559, DRG International) according to the manufacturer's instructions. According to the manufacturer, the ELISA has a sensitivity of 0.083ng/ml of testosterone, an intra-assay variation of 3.3%, and inter-assay variation of 6.7%.

## 2.7 Sperm counts, viability and morphology

Spermatozoa were collected from the cauda epididymis of at least 3 males per genotype, at least one of which was between 16-20 weeks old. Both cauda epididymides were weighed at the time of dissection. One cauda epididymis per male was minced in pre-warmed 1x PBS and incubated for 15 minutes at 37°C to allow a sperm suspension to form. The sperm suspension was diluted (1:4) in 1x PBS (at RT) and then incubated at 60°C for one minute to inactivate sperm as previously described (Wang, 2003). Spermatozoa were counted using a haemocytometer and counts were normalised to average cauda epididymis weight as previously published (Wang, 2003). Viability was assessed using supravital staining as previously described (Golshan Iranpour and Rezazadeh Valojerdi, 2013). At least 100 sperm per male were assessed. Briefly, a drop of sperm was mixed with 1% eosin (Sigma-Aldrich). After 15 seconds, a drop of 10% aqueous nigrosin (Sigma-Aldrich) was added, mixed and a smear was made. Sperm morphology was assessed in at least 100 sperm per male. Sperm were characterised as normal, headless, hookless or amorphous using published criteria (Wyrobek et al., 1983).

## 2.8 Fertility analysis

Fertility analysis was performed retrospectively. We examined matings of male mice (N=52 C57Bl/6 mice, N=42 $Mtrr^{gt/gt}$mice) with female mice. The time taken for a copulatory plug to form was calculated. To determine whether coitus resulted in pregnancy, uteri were dissected and the presence of conceptuses assessed or the female was allowed to litter.

## 2.9 DNA and RNA extractions

RNA was extracted from tissues using TRI-reagent (Sigma-Aldrich) as per the manufacturer's instructions with an additional RNA precipitation step in 4M LiCl at -20°C overnight. DNA was extracted from tissues using the DNeasy Blood & Tissue Kit (Qiagen), according to manufacturer's instructions. When both DNA and RNA were required form the same tissue, nucleic acid ex-

tractions were performed using the AllPrep DNA/RNA Mini Kit (Qiagen), according to the manufacturer's instructions.

## 2.10 Reverse transcription

cDNA was synthesised using random hexamer primers (Thermo Scientific) and RevertAid H Minus reverse transcriptase (Thermo Scientific) using 1-2$\mu$g of RNA in a 20 $\mu$l reaction, according to manufacturer's instructions.

## 2.11 quantitative PCR (qPCR)

qPCR amplification was conducted using MESA Green qPCR MasterMix Plus for SYBR Assay (Eurogentec Ltd., UK) on an MJ Research DNA Engine Opticon2 thermocycler (BioRad, USA). The following cycling conditions were used unless otherwise stated: 95°C for 10 minutes, 40 cycles: 95°C for 30 seconds, 60°C for 1 minute, followed by melt curve analysis. Transcript levels were normalised to housekeeping genes (GAPDH and/or HPRT). Relative cDNA expression levels were analysed as previously described (Livak and Schmittgen, 2001). Experiments were conducted in at least duplicate with four to eight biological replicates. Details of all qPCR primer sequences and primer concentrations are shown in Table 2.1.

## 2.12 Whole genome sequencing

Six *Mtrr*$^{gt/gt}$ embryos at E10.5 with congenital malformations and two C57Bl/6 phenotypically normal embryos at E10.5 were selected for whole genome analysis. DNA was extracted from whole embryos (Section 2.9). After quality verification using gel electrophoresis, non-degraded DNA was sent to the BGI (Hong Kong) for library preparation and sequencing. Sequencing was performed with 150bp paired-end reads on an Illumina HiSeq machine.

Bioinformatics analysis was performed in collaboration with Dr Russell Hamilton and Dr Xiaohui Zhao, Centre for Trophoblast Research, Cambridge, UK. Quality control processing was performed using Trim Galore (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/), to remove

adapters and low quality bases, and FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Summary metrics were created across all samples using the MultiQC package (http://multiqc.info). Sequencing reads were aligned to the C57Bl/6 reference genome (GRCm38, mm10) using BowTie2 with default parameters (http://bowtie-bio.sourceforge.net/bowtie2/index.shtml). Duplicates were marked using Picard (http://broadinstitute.github.io/picard). Structural variant analysis was performed using Manta (Chen et al., 2016b). Structural variants (SVs) were filtered using vcftools (version 0.1.15) (Danecek et al., 2011).

In order to identify single nucleotide polymorphisms (SNPs), the data was remapped to the mm10 reference genome using BWA (version 0.7.15-r1144-dirty) (Li and Durbin, 2009). Reads were then locally realigned and single nucleotide polymorphisms and short indels identified using GenomeAnalysisTK (GATK, version 3.7) (McKenna et al., 2010). Homozygous variants were called when more than 90% of reads at the locus supported the variant call, whereas variants with at least 30% of reads supporting the variant call were classified as heterozygous. Filtering of variants was performed firstly to remove low quality and biased variant calls and secondly as described in Oey et al. (2015) using vcftools (version 0.1.15). Briefly, this second filtering removed variants identified at 1) simple repeats with a periodicity $<$9 bp, 2) homopolymer repeats $>$ 8 bp, 3) dinucleotide repeats $>$ 14 bp, 4) regions with low mapping quality ($<$40), 5) overlapping annotated repeats or segmental duplications, and 6) where $>$ 3 heterozygous variants fell within a 10 kb region (Oey et al., 2015). The 129P2/OlaHsd mouse genome variation data was downloaded from the Mouse Genomes Project (Keane et al., 2011). Telomere length analysis was performed using TelSeq (Ding et al., 2014).

## 2.13 Telomere length assessment by quantitative PCR

Relative telomere lengths of embryos at E10.5 were assessed using qPCR as previously described (Cawthon, 2002). DNA was denatured by incubation at 95°C for 5 minutes prior to qPCR. qPCR was performed using the telomere (T) and 36B4 (S) primers (Table 2.1) (Callicott and Womack, 2006). The following cycling conditions were used: Telomere primers: 95°C for 10 minutes, 30 cycles:

95°C for 15 seconds, 56°C for 1 minute; 36b4 primer: 95°C for 10 minutes, 35 cycles: 95°C for 15 seconds, 52°C for 20 seconds then 72°C for 30 seconds. The qPCRs were performed in triplicate for each sample. Four biological replicates were used per group. Relative telomere length, presented as a T/S ratio, was calculated as previously described (Cawthon, 2002).

## 2.14 Sanger sequencing

In order to calculate the proportion of false-positive variant calls in the whole genome sequencing data, a number of variants were selected for validation by PCR amplification followed by Sanger sequencing. Sanger sequencing was performed on embryo or adult liver DNA. PCR amplification was performed using HotStar Taq DNA polymerase (Qiagen), using the following conditions: denaturation 95°C for 5min, 35 cycles of amplification: 95°C for 30s, 55°C for 40s, and 72°C for 1 min, and an extension of 72°C for 5 min. Gel electrophoresis was used to ensure a single product. PCR product was purified using MinElute PCR purification kit (Qiagen). Purified PCR samples were sent for Sanger sequencing (Department of Biochemistry, University of Cambridge, UK). Both forward and reverse strands were sequenced. Sanger sequencing data was analysed using Chromas (Technelysium) or MySequence (Version 1.1.6, www.telethon.jp). Primers used for PCR amplification and Sanger sequencing are shown in Table 2.2.

## 2.15 Sperm collection for DNA methylation analysis

Sperm from the cauda epididymis and vas deferens were collected from 16-20 week-old fertile male mice using the technique previously described (Hisano et al., 2013) with the following amendments. Sperm were released for 20 minutes at 37°C in Donners Medium (25 mM $NaHCO_3$, 20 mg/ml BSA, 1 mM sodium pyruvate and 0.53 % (vol/vol) sodium dl-lactate in Donners stock: 135 mM NaCl, 5 mM KCl, 1 mM $MgSO_4$, 2 mM $CaCl_2$ and 30 mM HEPES). Samples were centrifuged at 500 xg (21°C) for 10 minutes. The supernatant was then transferred and centrifuged at 3000 rpm (4°C) for 15 minutes. The majority of

supernatant was discarded and samples centrifuged at 3000 rpm (4°C) for 5 minutes. Further supernatant was discarded and sperm centrifuged at 12000 xg for 1 minute. Sperm were stored at -80°C.

## 2.16 Sperm DNA extraction

Solution A (75 mM NaCl pH 8; 25 mM EDTA) and Solution B (10 mM Tris-HCl pH 8; 10 mM EDTA; 1% SDS; 80 mM DTT) were added to the sperm samples followed by RNAse A incubation (37°C, 60 minutes) and Proteinase K incubation (overnight, 55°C) (Radford et al., 2014). DNA was extracted using phenol/chloroform/isoamyl alcohol mix (25:24:1) (Sigma-Aldrich) as per the manufacturers instructions. DNA was precipitated using 10M ammonium acetate, glycogen (0.1 mg/ml), and 100% ethanol, followed by incubation at -80°C for at least 30 minutes. DNA was collected by centrifugation (13000 rpm, 30 minutes). The pellet was washed twice in 70% ethanol and air-dried prior to resuspension in TE buffer. DNA quality and quantity was confirmed using gel electrophoresis and QuantiFluor dsDNA Sample kit (Promega) as per the manufacturer's instructions.

## 2.17 Bisulfite mutagenesis

Between 250ng-2$\mu$g of DNA per sample underwent bisulfite conversion using the Imprint DNA Modification Kit (Sigma) according to the manufacturer's instructions. When less than 500ng of DNA was undergoing bisulfite conversion the two-step conversion protocol was followed, otherwise the one-step conversion protocol was used. No template samples were run to ensure no contamination occurred during bisulfite conversion.

## 2.18 Pyrosequencing

Pyrosequencing was carried out as previously described to quantify methylation at individual CpG sites (Tost and Gut, 2007). Pyrosequencing primers were designed using PyroMark Assay Design Software 2.0 (Qiagen). Pyrosequencing primer sequences are given in Table 2.3. PCRs were performed using

the HotStarTaq Plus DNA Polymerase (Qiagen) in triplicate using the following PCR conditions: 95°C for 5 minutes, 40 cycles of 94°C for 30 seconds, 56°C for 30 seconds, 72°C for 55 seconds, then 72°C for 5 minutes. 5$\mu$l of PCR product was run on a gel to ensure specificity of amplification. The PCR product was then purified using strepdavidin Sepharose High Performance beads (GE healthcare). DNA-beads were then cleaned with 70% ethanol, 0.4M NaOH and 10mM Tris-acetate (pH 7.6) and then mixed with the pyrosequencing primer and PyroMark annealing buffer (Qiagen) according to the manufacturer's instructions. Pyrosequencing was conducted using PyroMark Gold reagents (Qiagen) on a PyroMark MD pyrosequencer (Biotage). Analysis of methylation status was performed using Pyro Q-CpG software (Biotage).

## 2.19 Mass spectrometry

Sperm DNA was digested into individual nucleoside components using the DNA Degradase Plus kit (Zymo research) according to the manufacturer's instructions. The heat inactivation step was omitted. The DNA was analysed for 5mC and 5hmC by liquid chromatography/mass spectrometry at the Babraham Institute, Cambridge, UK.

## 2.20 Methylated DNA immunoprecipitation and sequencing (MeDIP-Seq)

MeDIP-seq libraries were generated as previously described (Radford et al., 2014; Senner et al., 2012). A total of 3$\mu$g of sperm DNA was sonicated using a Diagenode Bioruptor UCD-200 to 200-700bp fragments. DNA was then end-repaired, dA-tailed and pre-annealed paired-end adapters ligated using the NEBNext DNA Library Prep Master Mix for Illumina kit (NEB). After each step the DNA was cleaned using Agencourt AMPure XP SPRI beads (Beckman Coulter).

MeDIP was performed in triplicate for each sample, with 500ng of DNA per IP. An input sample was reserved for each sample. MeDIP was performed using 1.25$\mu$l of anti-5mC antibody (Eurogentec BI-MECY-0100), the specificity of which has been previously confirmed (Radford et al., 2014). The

immunoprecipitated fraction was isolated using Dynabeads M-280 sheep anti-mouse IgG bead (Invitrogen 112-01D). The unbound fraction was retained. The triplicate IPs for each sample were then pooled and purified using MinElute PCR Purification columns (Qiagen). Libraries were amplified by PCR using Phusion High-Fidelity PCR Master Mix (12 cycles) and barcoded using iPCR tag primers (Table 2.4). Libraries were purified using Agencourt AMPure XP SPRI beads. The efficiency of the IP was verified using qPCR comparing pre-amplification input and IP fractions for regions of known methylation status (methylated: Nanog, H19 Imprinting control region (ICR), unmethylated: H1t, TsH2B ) (Lambrot et al., 2013). Library concentrations were estimated using the Kapa Library Quantification kit (Kapa Biosystems). Fragment size and library concentration were further verified by running on an Agilent High Sensitivity DNA chip on an Agilent 2100 BioAnalyzer. MeDIP libraries were sequenced with 100bp paired-end reads on an Illumina HiSeq at the Babraham Institute, Cambridge, UK.

## 2.21 Analysis of MeDIP-seq data

Quality assessment of reads was performed using FastQCC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Adaptor trimming was performed using Trim Galore (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). Reads were mapped to the GRCm38 (mm10) reference genome using Bowtie2 (http://bowtie-bio.sourceforge.net/bowtie2/index.shtml). All programmes were run with default settings unless otherwise stated. Sample clustering was assessed using principle component analysis (PCA). Briefly, each sample genome was divided up into 5kb windows and read coverage (as a proxy for methylation) for each window was calculated. PCA was performed using the top 500 most variable windows across all samples.

Differential methylation analysis was performed using the MEDIPS package in R (Lienhard et al., 2014). Data quality checks including saturation analysis and genome coverage estimations were performed as part of this package. The following key parameters were defined: BSgenome = BSgenome.Mmusculus.UCSC.mm10, uniq = 1e-3, extend = 300, ws = 500, shift

= 0. The script used is included in the appendix (Chapter 9). Differentially methylated regions (DMRs) were defined as windows (500bp) in which there was a least a 1.5 fold difference in methylation (reads per kilobase million (RPKM)) between C57Bl/6 and *Mtrr* sperm methylation level with a p-value $<$0.01. Adjacent windows were merged using BEDTools (version 2.27.0) (Quinlan and Hall, 2010).

The genomic localisations of DMRs, including association with coding/non-coding regions and CpG islands, was determined using annotation downloaded from UCSC (Meyer et al., 2013). The percentage of DMRs associated with repetitive regions of the genome was calculated using RepeatMasker software (http://www.repeatmasker.org). Gene ontology analysis was performed using DAVID (version 6.8) (Huang da et al., 2009b,a).

## 2.22 Western blotting

Western blotting was performed as previously described (Yung et al., 2007) by Billy Yung, Centre for Trophoblast Research, Cambridge, UK. Briefly, embryo lysates were prepared by homogenisation in lysis buffer (20 mM Tris (pH 7.5), 150 mM NaCl, 1 mM EDTA, 1 mM EGTA, 1% Triton X-100, 2.5 mM sodium pyrophosphate, 1 mM $\beta$-glycerolphosphate, 1 mM $Na_3VO_4$ and complete mini EDTA-free proteases inhibitor cocktail (Roche Diagnostics)). Bicinchoninic acid (Sigma-Aldrich) was used to determine protein concentration of the tissue lysates. Equivalent amounts of protein were resolved by SDS-PAGE and blotted onto nitrocellulose (0.2 $\mu$m). The membrane was stained with Ponceau S solution and scanned. This image was used as a loading control. The membrane was washed and blotted before incubation with the primary antibody: anti-HIRA (cat no. 13307, Cell signalling technology) or anti-$\beta$-actin (Sigma), overnight at 4 °C. The membrane was incubated with HRP-conjugated secondary antibody (anti-rabbit or anti-mouse, 1:10000, GE Healthcare). The signal of resolved protein was analysed by enhanced chemiluminescence (ECL) (Amersham Biosciences) using Kodak X-OMAT androgen receptor (AR) film (Sigma-Aldrich). A flat-bed scanner (Cannon 8000F) was used to scan films. Band intensities were determined with background subtraction using ImageJ (64-bit) software (NIH, USA).

## 2.23 Sperm collection and RNA extraction for small non-coding (snc) RNA sequencing

Sperm were collected from the cauda epididymis from 16-20 week-old male mice. The cauda epididymis was punctured and incubated in pre-warmed M2 medium (Sigma-Aldrich) for 1 hour at 37°C. The supernatant was centrifuged (2000 xg, 2 minutes) and the pellet washed in 1x PBS. Somatic cell lysis was performed by incubation in somatic cell lysis buffer (0.1% SDS, 0.5% Triton-X) for 10 minutes on ice, followed by a further wash in 1x PBS.

RNA was extracted using Tri-Reagent (Sigma-Aldrich) according to the manufacturer's instructions with some modifications. Sperm, in Tri-Reagent, were passed through a 26-gauge needle at least four times prior to addition of chloroform. Glycogen was added as a carrier to the aqueous phase. RNA precipitation was performed overnight. RNA purity was verified using an Agilent Technologies 2100 BioAnalyzer (Agilent Technologies) and quantified using a Qubit fluorometer (Life Technologies).

## 2.24 Small RNA library preparation and sequencing

Small RNA library preparation and sequencing were performed in collaboration with Katharina Gapp (Sanger Institute, Hinxton, UK) and Eric Miska (Department of Genetics and Gurdon Institute, Cambridge UK). The TruSeq Small RNA Library Prep kit (Illumina) was used to generate libraries according to the manufacturer's instructions. 20ng of RNA was used for library preparation. Briefly, total RNA from sperm was adapter ligated (3' and 5' adapters) and reverse transcribed to generate cDNA constructs. PCR amplification was performed, at which stage index sequences were introduced. Libraries were gel purified, with an approximately 145-160bp band size excised, corresponding to adapter ligated miRNAs, piRNAs and other small regulatory RNA molecules. Quality control and library quantification was performed using TapeStation analysis (Agilent Technologies). Sequencing was performed on a HiSeq 2500 (Illumina) machine, with 50bp single-end reads at the Gurdon Institute, Cambridge, UK.

## 2.25 sncRNA-seq analysis

sncRNA-seq analysis was performed in collaboration with Dr Russell Hamilton, Centre for Trophoblast Research, Cambridge, UK. Adapter sequences and low quality bases were removed from RNA-seq reads using cutadapt (Martin, 2011). Quality control processing was performed using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/), and summarised across all samples using the MultiQC package (http://multiqc.info). Reads were aligned to reference genome (GRCm38) using STAR (Dobin et al., 2013). Analysis was performed as described in Gapp et al. (2018). The CCA-3' was trimmed off. Reads were quantified using Salmon (Patro et al., 2017) using an index of tRNA sequences from GtRNAdb (Chan and Lowe, 2016, 2009) (parameters K=15, –perfecthash). For analysis of piRNAs, ncRNAs and miRNAs, reads were aligned and quantified using Salmon and the following indicies of RNAs: piRBAse (Zhang et al., 2014), ftp://ftp.ensembl.org/pub/release-94/fasta/mus_musculus/ncrna/ and miRBase (Kozomara and Griffiths-Jones, 2014), respectively. In parallel read quantification analysis was performed using SPORTS1.0 (Shi et al., 2018), with alignment using Bowtie2 (http://bowtie-bio.sourceforge.net/bowtie2/index.shtml). Differential expression analysis was performed using DESeq2 (Love et al., 2014). Overall sncRNA mapping distribution, rsRNA mapping distribution, and tRNA mapping distributions were determined using SPORTS1.0.

## 2.26 Statistical analyses

Statistical analysis was performed using GraphPad Prism software (version 7). Independent unpaired t-tests were used to analyse RT-qPCR data when comparing two groups only and data of time taken for a copulatory plug to form. Unpaired t-tests, with Welch's correction, were used to analyse structural variant and single nucleotide polymorphism totals, Ts/Tv ratio, Hom/Het ratio and TelSeq determined telomere lengths. Ordinary one-way ANOVA, with Dunnett's or Sidak's multiple comparison testing, was used for RT-qPCR (when more than two groups were assessed), relative telomere length determination by qPCR, testes/male weights, testosterone concentrations, sperm

parameters, histological data, and DMR CpG density. Two-way ANOVAs with Dunnett's, Sidak's or Tukey's multiple comparisons tests were used to analyse data for spermatogenesis staging, bisulfite pyrosequencing, SV frequency by chromosome, DMR distribution at repetitive elements, sperm ncRNA profiles and origin of tsRNA fragments. A binomial test (Wilson/Brown) was used to compare the observed frequency of nucleosome occupancy at DMRs to expected values. The proportion of coitus that resulted in pregnancy was analysed using Fisher's exact test. HIRA protein levels were analysed using a non-parametric Kruskal-Wallis test with Dunn's multiple comparisons test. $p<0.05$ was considered significant.

Table 2.1: **Details of qPCR Primers**.

| Gene | Forward Primer | Reverse Primer | Conc. | Reference |
|---|---|---|---|---|
| HPRT | CAGGCCA-GACTTTGTTGGAT | TTGCGCTCATCTTAG-GCTTT | Vari-able | Rameix-Welti et al. (2014) |
| GAPDH | CATGGCCTTCCGT-GTTCCT | GCGGCACGTCA-GATCCA | Vari-able | Gillich et al. (2012) |
| MTRR | GGGAAATTTGGAGC-TATGTGG | CAGATGAGTCAAGAC-CCCAGT | 200nm | Padmanab-han et al. (2013) |
| MTHFR | AGCTTGAAGCCAC-CTGGACTGTAT | AGACTAGCGTTGCT-GGGTTTCAGA | 200nm | Uthus and Brown-Borg (2006) |
| SLC19A1 | GGGTGTGCTACGT-GACCTTT | ACGGAACTGATCACG-GACTT | 200nm | Kooistra et al. (2013) |
| FOLR1 | GGCCCTGAGGA-CAATTTACA | TCGGGGAACACT-CATAGAGG | 200nm | Kooistra et al. (2013) |
| DNMT1 | CCTAGTTCCGTGGC-TACGAGGAGAA | TCTCTCTCCTCT-GCAGCCGACTCA | 200nm | Kowluru et al. (2016) |
| DNMT3a | CCAGACGGGCAGC-TATTTAC | AGACTCTCCAGAGGC-CTGGT | 300nm | La Salle et al. (2004) |
| DNMT3b | TTCAGTGACCAGTC-CTCAGACACGAA | TCAGAAGGCTGGA-GACCTCCCTCTT | 200nm | La Salle et al. (2004) |
| DNMT3c | CCTGTCAGGAAAGGC-CTGTT | CCTCTTGTCACCGAC-CTC | 200nm | Barau et al. (2016) |
| DNMT2 | AGCCTGTG-GCTTTCAGTATCA | TTGGCT-GACTTTCTTCAAC-TACTGC | 300nm | Tuorto et al. (2012) |
| NSUN2 | GGTAAACCATGACG-CATCC | CCTTCCACTGTGA-GACTCC | 100nm | Tuorto et al. (2012) |
| TET1 | ACACAGTGGTGC-TAATGCAG | AGCATGAACGGGA-GAATCGG | 200nm | Rakoczy et al. (2017) |
| TET2 | ACCTGGCTACTGT-CATTGCTCC | TGCAGTGACTCCTGA-GAATGGC | 50nm | - |
| TET3 | AGGCAGCTAAGCAC-CTCAG | GGCCCCGTAAGATGA-CACAG | 200nm | Hwang et al. (2016) |
| PTCH1 | CCTCTGCTC-CTTGATTGGCA | TCCCCAGTCCTGTCCT-CAAA | 200nm | - |
| CTS8 | TCCTGTGAA-GAATCAGGGCAC | GTGCT-CAGTGGGACCAGTTT | 200nm | - |
| SRD5A1 | CTTGAGCCAGTTTGCG-GTGTA | GCCTCCC-CTGGGTATTTTGTATC | 200nm | - |
| TPBPA | ACTGGAGTGC-CCAGCACAGC | GCAGTTCAGCATC-CAACTGCG | 200nm | - |
| IAP-GAG | AACCAATGC-TAATTTCACCTTGGT | GC-CAATCAGCAGGCGT-TAGT | 200nm | Kim et al. (2014) |
| IAP-3′LTR | GCACATGCGCAGAT-TATTTGTT | CCACATTCGCCGTTA-CAAGAT | 100nm | Kim et al. (2014) |
| LINE1-5′UTR | GGCGAAAG-GCAAACGTAAGA | GGAGTGCTGCGTTCT-GATGA | 100nm | Kim et al. (2014) |
| LINE1-ORF2 | GGAGGGA-CATTTCATTCTCATCA | GCTGCTCTTGTATTTG-GAGCATAGA | 200nm | Kim et al. (2014) |

| Gene | Forward Primer | Reverse Primer | Conc. | Reference |
|---|---|---|---|---|
| SINEB1 | TGAGTTCGAG-GCCAGCCTGGTCTA | ACAGGGTTTCTCTGT-GTAGCCCTG | 100nm | Kim et al. (2014) |
| Telomere | CGGTTTGTT-TGGGTTTGGGTTT - GGGTTTGGGTTTGGGTT | GGCTTGCCTTACCCT-TACCCTTAC CCTTACCCTTACCCT | 300nm | Callicott and Womack (2006) |
| 36B4 | ACTGGTCTAGGACCC-GAGAAG | TCAATGGTGCCTCTG-GAGATT | F:300nm R:500nm | Callicott and Womack (2006) |
| BTAF1 | ACGATGAAGATTTG-GATTATACCC | CAATTCAGCTGCTTG-GAGAGT | 200nm | - |
| FGFBP3 | GCCCTTGCTAGT-GAAGTCCA | GTCTCAGTGAGCTCG-GCATT | 200nm | - |
| KIF11 | GTCCCAGCC-GAGTTCTTTGA | GCATTAGCTTTC-CGCTCTGC | 200nm | - |
| IDE | AATCCGGCCATCCA-GAGAATA | GGGTCTGACAGT-GAACCTATGT | 200nm | - |
| MARCH5 | TTCACCAGGCTTGTCT-CCA | GCATCACTGTCACT-GCTCCA | 150nm | - |
| HIRA | CTC-CATCTTGTCAGGAAGT-GAT | GTTCCTG-GCACTCAGTAAAGAG | 200nm | - |
| RN45S | GCGTGTCA-GACGTTTTTCCC | AGAAAAGAGCGGAG-GTTCGG | 100nm | - |
| DYNLT1A | GAAGACTTCCAGGC-CTCCG | GGTTGACTTTGCTGT-GCTGG | 100nm | - |
| EXOC4 | CACAGCC-TACAGGGGCATTG | TTGGCAGC-GATTTCAAGAGTC | 200nm | - |
| TSHZ3 | GCGCGCAGCAGCC-TATGTTTC | TCAGCCATCCGGT-CACTCGTC | 300nm | - |
| CWC27 | TGATAATG-GCAGCCAGTTTTTCT | CTGTCAGGCGTAG-CATGTTGT | 200nm | - |
| PPM1H | CGACCCTCACAGATA-CACACT | GCCCAGTCGGT-CATTCGAT | 100nm | - |
| NANOG | GGACTGATCG-GCAAACTTTGA | TGGTCCCAAACTCCT-GATCTTC | 150nm | Lambrot et al. (2013) |
| H19 | GCCTCAGTGGTC-GATATGGTTT | AAAGGGACCCC-CTCCAGAA | 200nm | Lambrot et al. (2013) |
| H1t | ACGTAGGTGC-CATGGGTAAGA | CCCGCCTGAATCT-CAAGAGA | 200nm | Lambrot et al. (2013) |
| TsH2B | CCCCGCTTCTCAAC-CTCAA | CAACGTCT-CAAAACAGTTC-CAACT | 200nm | Lambrot et al. (2013) |
| UQCRB | TCTCAGGTCAAAATG-GCGGG | ATCATCTCGCATTAAC-CCCAGT | 200nm | - |
| GAS1 | CCTCTGCACCACGT-GTCTTA | TGGCAGTACC-GAGCTTTAGG | 200nm | - |
| HSD17B3 | CTGAGCACTTCCGGT-GAGAG | ATAAGGGGTCAGCAC-CTGAAT | 200nm | - |

Conc.: Concentration (nm). Where no reference is provided the PCR primers were designed by G Blake.

Table 2.2: **Details of Sanger sequencing primers.**

| SNP | Forward Primer | Reverse Primer |
|---|---|---|
| 1 | CCCACGCCTTCATTGTAA | GTTTCTCCCAAATCGAACC |
| 2 | CTGCTTTCCCTTAACACAG | GCTCTTCTTTCTCATTACCC |
| 3-6 | CTTGGGCTATCATCTTGTTC | TCTATACGTCTCGACCTTTC |
| 7 | TGATGTGGGGTTGGTCAG | AAATCTTTGGGCCGGAG-TAG |
| 8 | GAACCTACACCGAAGTAAC | GTCTGGTCAGGAAGATAA |
| 9-10 | ACTACTAACAAGACTGGGG | CTCAGAAGATCAGTAACGG |
| 11-14 | GACTCCTACAAGTGTTCAC | TGTGAGTACCTGATATGGG |
| 15 | CACTGAGCATCGCGCTTTC | CTACGGTAGGTGAAGA-GAACCT |
| 16 | ATCTTGTTCATATGGGCC-CTGG | CTCCTCTAAGCCTTTGCAC-TAA |
| 17 | AGGCTTGAACACTGC-CAACT | GGTCTGAACATCGTTG-GTCTC |
| 18 | TTGTTGGTCCGTGCTCACTT | GAGATGTGCCTACCTCG-CAA |

Primers were designed using Primer3 or NCBI Primer Blast by G Blake, D Chia or P Laouris.

Table 2.3: **Details of pyrosequencing primers.**

| DMR | PCR Forward Primer | PCR Reverse Primer | Sequencing Primer |
|---|---|---|---|
| H19 | GGGGGGTAG-GATATATGTATTTTT | *ACCTCATAAAACC-CATAACTATAAAAT-CAT | GTGTGTAAA-GATTAGGG |
| Peg3 | TTGGATTGGTTAGA-GAGGAAGT | *ACCTCATAAAACC-CATAACTATAAAAT-CAT | GGAGAGATGTT-TATTTTG |
| E28 | TAGGGA-GATTGATTGTGTTT-TATTAGAT | *ACTACCCTCTTC-CTCTCTTACTCC-TAAAT | AGATTGATTGTGTTT-TATTAGATT |
| E49 | GTATTTTGGGTAG-GTTTGGAATTGAATAT | *ACAACCTCAAA-CATATACTATAAACAC | TTTTAGTTTTTT-TATTGTTGTGAT |
| E50 | AGGTAGATTAG-GTTGAATATTTATA-GAGA | *AAAAACAAAATC-TATAAAACCC-TAAATCC | GGGTTGAAATTAAA-GATAGG |
| E52 | AGTTTTGGAGTTTAG-GTTAGGT | *ATAACCATAT-CAATAAAACCTTCT-TAACA | AGTTTAGGTTAGGT-TAAAG |
| E66 | GGA-GAGGGTTGTTTTTTTT-GGGATAAGT | *ATTTACCCTTCTCCC-CACTACCA | GTGTTTTTGTTTT-TAGTTAGTTATT |
| E74 | GTAAGTTGGAGGTA-GATATAGTTAGAAGT | *ACTCCATTTA-CAACTCTTCACCAT-ACT | ATAAAAGTTTT-TAAATTTATTTGGG |
| E76 | GGAATTTAAGTG-GTTTGTTTATGGGTATA | *TAACCCATTTCCAT-TATTTAAAACAAACT | ATTTTATGAATTTG-GTATATATGTT |
| E79 | TTAGAAAGGAGAG-GTAGGGAGTGTG | *TACACAAACTTCC-CAATCTCAC | AGGTAGGGAGTGTGG |
| E81 | TGTAAGGTTTAA-GATAGGAGATGTTATT | *TACAAACC-CATATAACTA-CAAAACTCAT | TTTAGTTGGTGA-GATTGA |
| E109 | TAGAAGTTTTATTGGT-TAGGTGTAAGTTA | *CAAAAAAATCTT-TAAACACCAAAAAA-CATC | ATGGATGTGAAGAGG |
| E110 | TTAGGAAG-TATAGGGGTTT-TATAGTTGA | *ACTATAACCTAAC-TAATCCCACTTCA | AGGGGTTT-TATAGTTGAG |
| E112 | *TAGGAAATAAAAAT-GTGAGGGGTAAT | ACTACTACTCTATCC-CTTTTATAACA | ACTCTATCCCTTT-TATAACAAT |
| E114 | GTTTTTGGGTTTTAG-TAATAGTGTATGG | *AACACATATAACTTC-CCACACATC | GTTTTTAGAGAG-TAGTTTAGG |
| E115 | AGTTTGATATTGAAG-GGGTTATTTTGGA | *AACATAATATACTTC-CCTTCTAAAATCT | GGTTATTTTGGAA-GAGGT |
| D87 | GGGTGTTGTTTTA-GAGGTTTTGTATTTTTG | * ATCACCTCATCCTAC-TAAATACTATAAC | AGATTAATTTTT-TAAGTTTTGTTG |
| D20 | GAAATGTTTGAGA-GAATGGGTTTTAG | *ATACAATAACCT-TAACCACTCATAAC | AAATTTTAGTTTAAG-GTTTTATATA |
| A10 | TGTTATAGGTTG-GAGGGTATGAGTT | *ATCCAATCCTTCT-CAACAATCTCTCTC | AGGAGTAGAGAGTTT-TAG |
| B16 | ATTAAGTTATTTGGTG-GTGAAGAAGTAT | *ACCCCTACCTTCCA-CACAC | TTGTGATTAAGAAG-GTATTAGAA |

| DMR | PCR Forward Primer | PCR Reverse Primer | Sequencing Primer |
|---|---|---|---|
| 25 | AAGGGAGAAGGTTTT-TATGGT | *AAATAAAAAATAAC-CACCCCCAAC | ATGGTTAGTTGGA-GATTTAA |
| 224 | GGTAGTGAAGTTAT-GTTTATGAGGATATT | *CCCTTTCCAATACTC-CAACCACTA | AATTTTTAGGTATTATT-TATAGGGT |
| 269 | *TAGGAAATAAAAAT-GTGAGGGGTAAT | ACTACTACTCTATCC-CTTTTATAACA | ACTCTATCCCTTT-TATAACAA |
| 274 | *GGATGTAGAT-GAATTTGGAAATTT-TAGA | ACTAAAACTCCAC-TATTCCTTTCCACAATT | CTATTCCTTTCCA-CAATTAC |
| 278 | GTTTGAATTTTTTGA-GGGTTATTTAAGGTT | *TCCCCCTTTTAAAAC-TATCTCTTCT | TTTGAGTTTAGTTTTT-TATATTGTG |
| 279 | GTTTTTTTAGTA-GAGTTGGGAGTTT | *AAACAAACTAAACC-TAAACAAATATAACT | GGGAGTTTTGTTTTTTT-TAGAT |
| 281 | AG-GATGGGGGAATAAAAT-GATG | *ACCCAAACCTA-TATAATAATAACTTTCC | GGGGAATAAAATGAT-GTG |
| 289 | GGTTGTGGGTTGT-TATTGTAGTTAAG | *AT-CAAATTTTCTCTTTCC-CCATTC | GTTATTGTAGT-TAAGGGGT |
| C7 | GGTAGTTTTGGATATG-TAAGAGTTG | *CATATCCC-CTCTTCTCTT-TATCTTTTTA | GGTTTATAGAGTTGA-GAATTGTG |
| B1 | AGTTTAAGGT-TATTGGGAAGTATAG-TAT | *TCAATTAATCCT-CAAAACAACCCTCC-TAAA | GGAGTTAGTGGT-TAATTTG |
| B17 | TGAGTTAAGAATG-TAGGGTAGATTGT | *CAAAATACACTAA-CATTTCCTTACTAT-CAA | GAGATAGAGTTTTAT-TAAGTAGT |
| B19 | *AATTGGAGAATAG-GATTTTTAGTGTTTATA | TCCTAAATCTCCC-TATAACACCATCAAT | AACAATAATAAAT-ACTTAAACTACA |
| B20 | GTAGAAG-GATTTTGTTGTG-GTATTT | *ATCAAAAAACC-CATCCCTACA | GGTATTTGTTTTGTTGGT |
| B21 | AGTATAAGTGATG-GAGTTTTAGATAAAT | *AATCTTAAACCAATA-CAAATCCATACT | TGATGGAGTTTTA-GATAAATTAG |
| A9 | *ATTTTTAAATTT-TAAATGGGTAA-GAAGT | ACCAACCTAATCTA-CAAAATAAAATCTAAT | ACCAAAAATAACC-CACT |
| A18 | AGGAATTTAAGTG-GTTTGTTTATGGGTATA | *ACCCATTTCCATTATT-TAAAACAAACT | GAAAGATTTATTTAG-TAATTTGTTG |
| A19 | TAGTTGAATTGAG-TATGGAGAAAGAGTA | *AACACAAACTAACC-CTAAAATCCTCAT | TGAATATTTTTTAG-TATAGAGTTGT |
| A24 | *ATTGTTTTTGAGGTT-TAGGGTTAAAG | AACCTCTTACTTCTC-CAACTAC | ACTTCTCCAACTA-CAAAT |
| A28 | GGTTAGGAAGGGATT-TAGTGTTAGTTTA | *ACCAACCACTAACT-CAATCTAA | TGTTGTTTTTTT-TAATTTGTTTATG |
| A30 | AGTTATGTG-TAGGGTATTAATT-TAATGT | *TCTCCCCATACC-TAACTTCAC | AGATGTAT-GTTTTTGAAGTTTA |

| DMR | PCR Forward Primer | PCR Reverse Primer | Sequencing Primer |
|---|---|---|---|
| D40 | TGGGTTTATA-GAGTTGAGAATTGTGT | *CATATCCC-CTCTTCTCTT-TATCTTTTTA | AGAATTGTGTAAGT-TATTAGTTTAA |
| D52 | TTGTAGTGATTTTATG-GAGAAAAAGAGAA | *CAACAATCCAATC-CAAAATTCTAAAC | TTTTTTGGTATTT-TAGTTGTTTTTA |
| D53 | TGTTAGGGGGTTGGT-TATTT | *ACTACCCCTACA-CATATACAAATTATT | TTTTAGAGTGTGGGGA |
| D72 | TGGTAAAATTAGATGT-GTGGTATATGT | *AAACACTAAACCA-CAACCCAATATCTACA | AGAAATTAAT-TATATTGGAGTATT |
| 17 | AGTTTTTTTGGGTGT-GAGAATTAT | *ATCCAAAACCAC-TAAAAAACACAAC | TGGGTGTGAGAAT-TATT |
| 181 | AGGAAGTATTTGA-GATGTTTAGAGTT | *CCACTACAATATATC-CTTAACCCTACT | TTATAGGATTATAG-GATGTGT |
| 60 | *AGATGTAAAA-GAAAGGAAGGTAGT | CAATCCCCCATTCAAT-ACAAAAA | ACAAAAATACC-CTCCC |
| 177 | *AATTTAAGTGAA-GAATTTTGGTTTTATG | AACCCTAAATATTCTC-CTTTACTCAAC | ATTCTCCTTTACT-CAACT |
| 185 | *TGTTTTATTTG-TAGTTGGAGAAG-TAAG | ACCTCTCAAAAATCC-TATCCATAATAT | CACCCCCTCAC-TAATCCTCTA |
| 189 | TGGGAGTTAGTTATTG-GTTTAGTTTGAGG | *AACCCAATACTAATC-CACCTTTACA | TGAGGTTGGTATAG-GAA |
| 220 | AGTTATGTG-TAGGGTATTAATT-TAATGT | *TCTCCCCATACC-TAACTTCACACT | AGATGTAT-GTTTTTGAAGTTTA |
| 230 | TGGGTTTTAAAGTT-TAAGGTTATTGG | *TCCTCAAAACAACC-CTCCTAAA | AGTTAGTGGT-TAATTTGTTAATTTT |
| 280 | AGAAGGGGTTATAG-GAAGTATTTTAGG | *CCCAAAAAATAATC-CCTTCCCTTTTC | GGGTTATAGGAAG-TATTTTAGGA |
| 282 | GAATTGTTTAGGGAG-GATTTTTTTATAGT | *CCAACATTCTA-CAAAATCTACTACTCC | TGTTGTGTAGATA-GATATTTTGAT |
| UC1 | AGGAAAATATATGT-GTTGTTTATTTAGTGT | *AAACCAACTCAAT-ACTCACAATATTCT | ATGTGTTGTTTATT-TAGTGTTT |
| UC2 | AGATGGGAGTTGT-GTTGATGATAAT | *ACCAATCACCAA-CAAAATTTAATAAC | TGGGTTTGTAGGTGTA |

PCR primers were used with a concentration of 250nm. The sequencing primer was used at a concentration of 417nm. All primers were designed using PyroMark Assay Design Software 2.0 (Qiagen) by G Blake. * denotes Biotin tag. UC: Unchanged.

Table 2.4: **Barcoded primers used for MeDIP-seq**.

| Name | Primer Sequence | Barcode | Sequence Obtained |
|---|---|---|---|
| iPCRtag1 | CAAGCAGAAGACGGCATACGA-GATAACGTGATGAGATCGGTCTCG-GCATTCCTGCTGAACCGCTCTTCCGATC | AACGT-GAT | ATCACGTT |
| iPCRtag2 | CAAGCAGAAGACGGCATACGA-GATAAACATCGGAGATCGGTCTCG-GCATTCCTGCTGAACCGCTCTTCCGATC | AAA-CATCG | CGATGTTT |
| iPCRtag3 | CAAGCAGAAGACGGCATACGAGATAT-GCCTAAGAGATCGGTCTCGGCATTCCT-GCTGAACCGCTCTTCCGATC | ATGCCTAA | TTAGGCAT |
| iPCRtag4 | CAAGCAGAAGACGGCATACGA-GATAGTGGTCAGAGATCGGTCTCG-GCATTCCTGCTGAACCGCTCTTCCGATC | AGTG-GTCA | TGACCACT |
| iPCRtag5 | CAAGCAGAAGACGGCATACGAGATAC-CACTGTGAGATCGGTCTCGGCATTCCT-GCTGAACCGCTCTTCCGATC | ACCACTGT | ACAGTGGT |
| iPCRtag6 | CAAGCAGAAGACGGCATACGAGATA-CATTGGCGAGATCGGTCTCGGCATTCCT-GCTGAACCGCTCTTCCGATC | ACATTGGC | GCCAATGT |
| iPCRtag7 | CAAGCAGAAGACGGCATACGAGATCA-GATCTGGAGATCGGTCTCGGCATTCCT-GCTGAACCGCTCTTCCGATC | CAGATCTG | CAGATCTG |
| iPCRtag8 | CAAGCAGAAGACGGCATACGAGAT-CATCAAGTGAGATCGGTCTCGGCATTC-CTGCTGAACCGCTCTTCCGATC | CATCAAGT | ACTTGATG |

Barcoded primer sequences provided by Claire Senner, Babraham Institute, Cambridge, UK.

# Chapter 3

# Analysis of spermatogenesis and fertility in *Mtrr$^{gt}$ mice*

Testes histological analysis and RT-qPCRs for genes involved in folate metabolism were performed in collaboration with an undergraduate student Jessica Hall (Department of Physiology, Development and Neuroscience, Cambridge, UK).

## 3.1 Introduction

In recent decades it has been established that there is a link between folate intake, male fertility and testes function (Boxmeer et al., 2009; Singh and Jaiswal, 2013). Evidence for this comes from both human populations with polymorphisms in genes encoding for folate metabolic enzymes and from mouse models (reviewed in Singh and Jaiswal (2013)). For example, *Mthfr*$^{-/-}$ mice have testes morphology abnormalities and reduced fertility, although the severity of the phenotype depends upon the genetic background of the mouse (Chan et al., 2010). Furthermore, in men, infertility correlated with reduced folate concentrations in serum compared to fertile individuals (Murphy et al., 2011). Similarly, dietary folate deficiency in mice is associated with delayed onset of meiosis during spermatogenesis, although fertility is normal in these males (Lambrot et al., 2013).

Spermatogenesis is a highly coordinated differentiation event that occurs in the seminiferous tubules of the adult testes and results in the formation of spermatozoa. The process is supported by somatic cells including Leydig cells and Sertoli cells. Leydig cells are found in the interstitial space between seminiferous tubules and produce testosterone (Smith and Walker, 2014). Sertoli cells are found within the seminiferous tubules and provide a niche for spermatogenic cells by maintaining the blood testes barrier and producing growth factors (Smith and Walker, 2014). For mature spermatozoa to form, a series of mitotic and meiotic divisions must occur. Sperm precursors called spermatogonia directly contact the basement membrane of the seminiferous tubule epithelium (Griswold, 2016). A spermatogonium will undergo a mitotic division to self-renew and form a primary spermatocyte (Griswold, 2016). Primary spermatocytes move apically towards the lumen of the seminiferous tubule as they meiotically divide to form haploid spermatids. Spermatids will then cytodifferentiate to give rise to spermatozoa, which are released into the tubule lumen and eventually leave the testis for the epididymis where they become functionally mature (Smith and Walker, 2014). Spermatogenesis occurs in waves along the length of the seminiferous tubule to allow for the continuous production of sperm. Defects in spermatogenesis might lead to subfertility or infertility (Anawalt, 2013).

Sperm have highly specialised epigenetic landscape (Reik and Surani, 2015). Sperm specific DNA methylation patterns are established following a wave of epigenetic reprogramming that occurs in the primordial germ cells of the embryo (Reik et al., 2001). DNA methylation is orchestrated by a family of DNMT enzymes, while DNA demethylation reactions require TET enzymes. In male germ cells in rodents, DNMT3C is a non-catalytically active enzyme that is responsible for transposon silencing (Barau et al., 2016). Although DNA methylation is mostly established prenatally, extensive changes to the epigenetic landscape of developing sperm occur in adult males (Oakes et al., 2007; Loukinov et al., 2002). This includes changes in DNA methylation patterns, involving both *de novo* methylation and demethylation, at unique and repetitive loci in spermatogonia and early spermatocytes (Oakes et al., 2007). A key event during spermatogenesis, occurring as haploid speramtids mature into spermatozoa, is the extensive histone replacement by protamines. Nucleosomes are disassembled and replaced firstly by highly basic transition proteins and then by protamines (PRM1 and PRM2) (Brunner et al., 2014). This allows the DNA to become highly compacted and the transcriptionally silent (O'Doherty and McGettigan, 2015). Protamines, both PRM1 and PRM2, can be post-translationally modified (e.g., methylated) (Brunner et al., 2014). Nucleosomes are retained a certain genomic regions including genes important for embryonic development and CpG rich regions lacking DNA methylation (Erkek et al., 2013). Residual nucleosomes often contain the non-canonical histone variant H3.3 which is trimethylated at lysine 4 (Erkek et al., 2013). Non-coding RNAs, mainly tRNA fragments, taken up by sperm in the epidiymis may carry RNA modifications (Sharma et al., 2016; Chen et al., 2015). The RNA methyltransferases DNMT2 (also known as TRDMT1) and NSUN2 are important for these modifications (Tuorto et al., 2012). Intriguingly, *NSun2* is required for normal testes development, *Nsun2*$^{-/-}$ mice have severe defects in the meiotic progression of germ cells (Hussain et al., 2013).

Folate metabolism may influence spermatogensis and fertility through a number of mechanisms. The metabolism of folate is required for DNA synthesis and is intertwined with methionine metabolism to promote the transmission of one-carbon methyl groups required for cellular methylation (Jacob et al.,

1998; Ghandour et al., 2002). Therefore, folate metabolism is important for cell proliferation and widespread epigenetic changes, such as those that occur during spermatogenesis.

A role for MTRR in spermatogenesis has not yet been established in mice. It is important to ascertain whether abnormalities in the differentiation and function of germ cells may act as a contributing factor to the TEI mechanism in the $Mtrr^{gt}$ model. Germ cell abnormalities, or reduced germ cell quality, may reflect underlying genetic defects or oxidative stress in the testicular environment (Ray et al., 2017; Agarwal et al., 2014). Abnormal sperm parameters may perhaps contribute to embryonic phenotypes. For example, in a study of men undergoing *in vitro* fertilization, abnormal sperm morphology was associated with reduced embryo quality (Shu et al., 2010). Therefore I aimed to assess the morphology and function of testes and spermatozoa in the $Mtrr^{gt}$ mouse line.

## 3.2 Results

### 3.2.1 *Mtrr* is broadly expressed in the testes

To determine the spatial localisation pattern of MTRR in adult testes, immunohistochemistry was performed on C57Bl/6 sexually mature adult male testes using an antibody against MTRR. MTRR was broadly expressed in the testes including in all spermatogenic cell types, Leydig cells, and Sertoli cells (Figure 3.1A-C). However, MTRR expression was not apparent in all spermatogonia (Figure 3.1B). This could suggest that protein expression might be dependent upon the cell cycle. Alternatively, a fixation artefact, the low cytoplasm to nuclear ratio of spermatogonia or section thickness may have made observation of MTRR expression challenging. Subcellularly, MTRR protein expression was detected in the nucleus and cytoplasm (Figure 3.1B).

The $Mtrr^{gt}$ allele is a hypomorphic mutation and the degree of genetic knockdown is tissue specific (Padmanabhan et al., 2013). Therefore I examined expression levels of the wildtype *Mtrr* transcript via RT-qPCR in testes from $Mtrr^{+/+}$, $Mtrr^{+/gt}$ and $Mtrr^{gt/gt}$ mice compared to C57Bl/6 control testes. The RT-qPCR primers were designed specifically to recognise exons after the gene-trap insertion, thus RT-qPCR only detects full-length wildtype *Mtrr* transcripts. Wildtype *Mtrr* mRNA expression was significantly decreased in $Mtrr^{+/gt}$ and $Mtrr^{gt/gt}$ testes to 59% and 14% of C57Bl/6 levels, respectively ($p < 0.0001$; Figure 3.1D). $Mtrr^{+/+}$ testes expressed levels of *Mtrr* mRNA similar to C57Bl/6 controls (Figure 3.1D). This data indicates a robust knockdown of the wildtype *Mtrr* transcript in $Mtrr^{gt/gt}$ and $Mtrr^{+/gt}$testes.

Next, I established whether the expression of other genes involved in folate uptake (e.g., *Folr1* and *Slc19a1*) and metabolism (e.g., *Mthfr*) was altered by the $Mtrr^{gt}$ mutation in mouse testis. Using RT-qPCR, the mRNA expression of *Folr1*, *Slc19a1*, and *Mthfr* genes was determined and found to be similar in C57Bl/6, $Mtrr^{+/+}$, $Mtrr^{+/gt}$ and $Mtrr^{gt/gt}$ testes (Figure 3.1E-G).

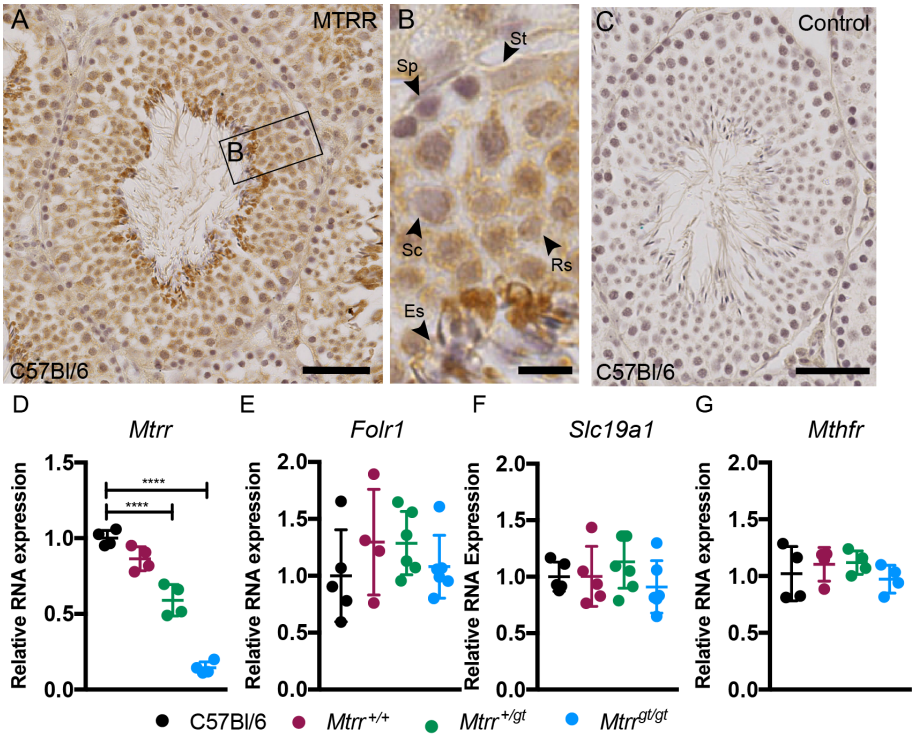



Figure 3.1: **MTRR protein is widely expressed in adult mouse testes. (A-B)** Representative histological section of a seminiferous tubule from an adult C57Bl/6 testis (seminiferous tubule stage VII/VIII) that was immunostained with an antibody against MTRR (brown). Nuclei are purple. Three histological sections from three males were analysed. Scale bars: A, 50 $\mu$m; B, 5 $\mu$m. **(B)** Higher-magnification of boxed region in (A). Arrowheads indicate cell types: St, sertoli cell; Sp, spermatogonia; Sc, spermatocyte; Rs, round spermatid; Es, elongated spermatid. **(C)** Control histological section of C57Bl/6 seminiferous tubule from adult testis immunostained with the secondary antibody only. Scale bar: 50 $\mu$m. **(D-G)** Graphs showing relative mRNA expression of genes encoding for folate metabolic enzymes and transporters in testes from C57Bl/6 control males (black), *Mtrr*$^{+/+}$ males (purple), *Mtrr*$^{+/gt}$ males (green), and *Mtrr*$^{gt/gt}$ males (blue) as determined via RT-qPCR analysis. The following genes were assessed: **(D)** *Mtrr* (wildtype transcript), **(E)** *Folr1*, **(F)** *Slc19a1*, and **(G)** *Mthfr*. Data is plotted as mean $\pm$ sd. Testes from 4-6 males per genotype were assessed. Data is represented as fold change compared to C57Bl/6 controls, which was normalised to 1. A one-way ANOVA statistical test, with Dunnett's multiple comparisons test, was performed on each data set. **** $P<0.0001$.

To verify the knockdown of *Mtrr* in testes, I performed immunohistochemistry against MTRR in *Mtrr*$^{+/+}$, *Mtrr*$^{+/gt}$, *Mtrr*$^{gt/gt}$ and C57Bl/6 control testes. Unexpectedly, immunohistochemistry failed to clearly demonstrate the strong knockdown of *Mtrr* observed at the mRNA level using RT-qPCR in *Mtrr*$^{+/gt}$ and

*Mtrr*$^{gt/gt}$ testes. By immunohistochemistry, the MTRR protein had equivalent intensity and localisation to C57Bl/6 controls in *Mtrr*$^{+/+}$, *Mtrr*$^{+/gt}$ and *Mtrr*$^{gt/gt}$ testes (Figure 3.2A-E). The *Mtrr*$^{gt}$ mutation is suspected to lead to a truncated gene-trap containing mRNA transcript (Elmore et al., 2007). This gene-trap containing transcript is translated leading to a non-functional protein, as confirmed by western- blotting and $\beta$-galactoside staining in *Mtrr*$^{+/gt}$placentas (Elmore et al., 2007). Details of the epitopes of the MTRR protein the MTRR antibody (which is polyclonal) binds were not available. We speculate that the MTRR antibody recognises motifs of this non-functional protein thus leading to MTRR detection at C57Bl/6 levels even in *Mtrr*$^{gt/gt}$ testes. However, a western blot should be performed to verify the antibody specificity.

To confirm that the protein expression of FOLR1, SLC19A1 and MTHFR was unaffected by the *Mtrr* mutation, I performed immunohistochemistry against FOLR1, SLC19A1 and MTHFR in *Mtrr*$^{+/+}$, *Mtrr*$^{+/gt}$, *Mtrr*$^{gt/gt}$ and C57Bl/6 control testes. It was previously shown that MTHFR protein is widely expressed in C57Bl/6 mouse testis including in spermatogonia, spermatids, pachytene and preleptotene spermatocytes and Sertoli cells, but not leptotene or zygotene spermatocytes or interstitial cells (Garner et al., 2013). I broadly confirmed this finding via immunohistochemistry (Figure 3.2F), though some MTHFR expression was detected in the interstitial cells (Figure 3.2F and Ff). SLC19A1 and FOLR1 showed a similar broad expression pattern as MTRR and MTHFR in C57Bl/6 testes (Figure 3.2K, O). Notably, SLC19A1 had limited expression in spermatocytes (Figure 3.2K and Kk) and FOLR1 showed relatively strong expression in Sertoli cells (Figure 3.2O and Oo). There was no apparent difference in FOLR1, SLC19A1 or MTHFR protein expression in *Mtrr*$^{+/+}$, *Mtrr*$^{+/gt}$ or *Mtrr*$^{gt/gt}$ testes compared to C57Bl/6 testes (Figure 3.2). This data suggests that most cells in the testes take up and metabolise folate. Additionally, cells in *Mtrr*$^{+/gt}$ and *Mtrr*$^{gt/gt}$testes do not appear to compensate at either a transcriptional or protein level, through up-regulation of other pathway components, for the metabolic defect caused by the *Mtrr*$^{gt}$ allele.

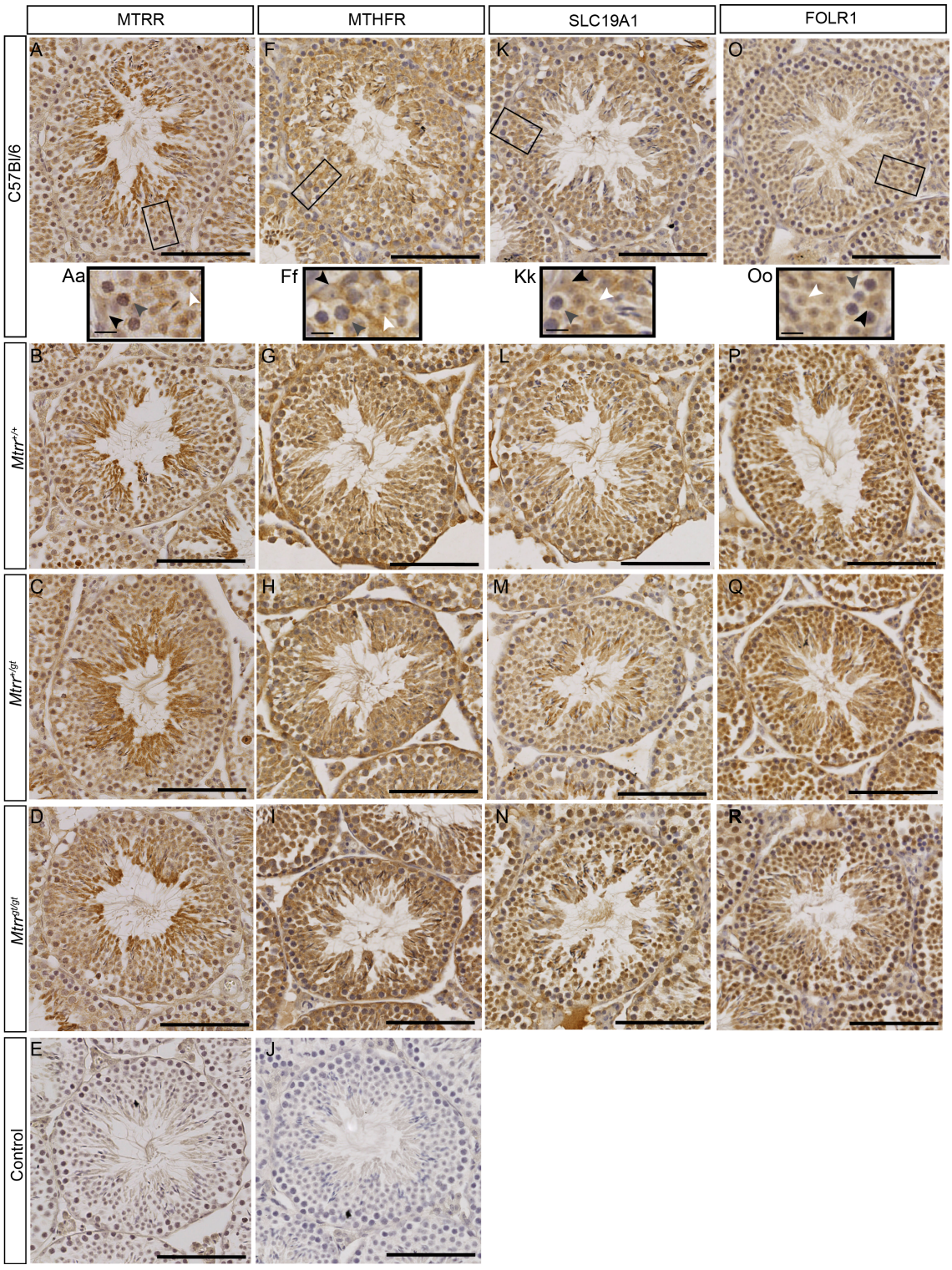

MTRR
MTHFR
SLC19A1
FOLR1
C57Bl/6
Mtrr+/+
Mtrrgt/gt
Mtrrgl/gt
Control
A
Aa
F
Ff
K
Kk
O
Oo
B
G
L
P
C
H
M
Q
D
I
N
R
E
J

Figure 3.2: **The machinery to metabolise folate is widely expressed in the testes. (A-D)** Representative histological sections of seminiferous tubules immunostained with an antibody against MTRR are shown for **(A)** C57Bl/6 with (Aa) a higher-magnification of boxed region in (A), **(B)** *Mtrr*$^{+/+}$, **(C)** *Mtrr*$^{+/gt}$ and **(D)** *Mtrr*$^{gt/gt}$ testes. **(F-I)** Representative histological sections of seminiferous tubules immunostained with an antibody against MTHFR are shown for **(F)** C57Bl/6 with (Ff) a higher-magnification of boxed region in (F), **(G)** *Mtrr*$^{+/+}$, **(H)** *Mtrr*$^{+/gt}$ and **(I)** *Mtrr*$^{gt/gt}$ testes. **(K-N)** Representative histological sections of seminiferous tubules immunostained with an antibody against SLC19A1 are shown for **(K)** C57Bl/6 with (Kk) a higher-magnification of boxed region in (K), **(L)** *Mtrr*$^{+/+}$, **(M)** *Mtrr*$^{+/gt}$ and **(N)** *Mtrr*$^{gt/gt}$ testes. **(O-R)** Representative histological sections of seminiferous tubules immunostained with an antibody against FOLR1 are shown for **(O)** C57Bl/6 with (Oo) a higher-magnification of boxed region in (O), **(P)** *Mtrr*$^{+/+}$, **(Q)** *Mtrr*$^{+/gt}$ and **(R)** *Mtrr*$^{gt/gt}$ testes. Coloured arrowheads in Aa, Ff, Kk and Oo indicate cell types: Sertoli cells (black), spermatocyte (grey) and round spermatid (white). **(E,J)** Control histological section of C57Bl/6 seminiferous tubules from adult testis immunostained with the secondary antibody only. Three histological sections from three males per genotype were analysed for each antibody. Scale bars: 50 $\mu$m, except Aa, Ff, Kk, Oo: 5 $\mu$m.

### 3.2.2 Testes of *Mtrr*$^{gt/gt}$ mice are morphologically more spherical

Prior to assessing whether the *Mtrr*$^{gt}$ mutation affected testes morphology or function, we performed gross analysis of male body weight. Age matched C57Bl/6, *Mtrr*$^{+/+}$, *Mtrr*$^{+/gt}$ and *Mtrr*$^{gt/gt}$ male mice were weighed. *Mtrr*$^{+/+}$ and *Mtrr*$^{gt/gt}$ male weights were equivalent to C57Bl/6 controls (Figure 3.3A). Interestingly, I observed that *Mtrr*$^{+/gt}$ male body weight was significantly higher than C57Bl/6 controls ($p<0.0001$; Figure 3.3A). This data suggests that the *Mtrr*$^{gt}$ allele might have different metabolic repercussions when in heterozygous or homozygous form.

Testes from C57Bl/6, *Mtrr*$^{+/+}$, *Mtrr*$^{+/gt}$ and *Mtrr*$^{gt/gt}$ males were then weighed and processed for histological analysis. The average testis weight of *Mtrr*$^{gt/gt}$ males was significantly less than to C57Bl/6 controls ($p=0.0133$; Figure 3.3B). *Mtrr*$^{+/+}$ and *Mtrr*$^{+/gt}$ testes weights were no different to C57Bl/6 controls (Figure 3.3B). However, when testis weight was calculated as a percentage of body weight, no differences between the *Mtrr* genotypes and C57Bl/6 controls were observed (Figure 3.3C). This data indicates that the decrease in testes weight

in *Mtrr*$^{gt/gt}$ mice was likely small and proportional to body weight, despite the *Mtrr*$^{gt/gt}$body weight not being statistically significantly different from C57Bl/6 controls.

Gross analysis of the testis morphology revealed that compared to the ovoid shape observed in control C57Bl/6, *Mtrr*$^{+/+}$ and *Mtrr*$^{+/gt}$ testes, the *Mtrr*$^{gt/gt}$ testes appeared rounder and closer to a spheroid shape (Figure 3.3D). This observation was supported by a shorter major axis in *Mtrr*$^{gt/gt}$ testes relative to C57Bl/6 controls (p=0.0228; Figure 3.3F-G). The absolute number of seminiferous tubules counted per testes histological section was reduced in *Mtrr*$^{gt/gt}$ testes versus C57Bl/6 controls (Figure 3.3H). This reduction reflects a smaller testes size in *Mtrr*$^{gt/gt}$ males since seminiferous tubule density (tubules per mm$^2$) was equivalent in *Mtrr*$^{gt/gt}$ and C57Bl/6 males (Figure 3.3I). The testes axes lengths, seminiferous tubule number and density of *Mtrr*$^{+/+}$ and *Mtrr*$^{+/gt}$ testes were similar to controls (Figure 3.3F-I). To investigate the seminiferous tubules themselves more specifically, the average cross-sectional area of each tubule and tubule epithelial thickness (at its thickest and thinnest, Figure 3.3E) were determined. All parameters measured were equivalent in C57Bl/6 control in *Mtrr*$^{+/+}$, *Mtrr*$^{+/gt}$ and *Mtrr*$^{gt/gt}$ testes (Figure 3.3J-L). Additionally, the thickness of the tunica albuginea, the fibrous connective tissue encapsulating the testis, was unaffected by *Mtrr* deficiency (Figure 3.3M). Overall, *Mtrr*$^{gt/gt}$ homozygosity in mice alters testis shape and size yielding fewer seminiferous tubules, yet normal gross seminiferous tubule morphology.

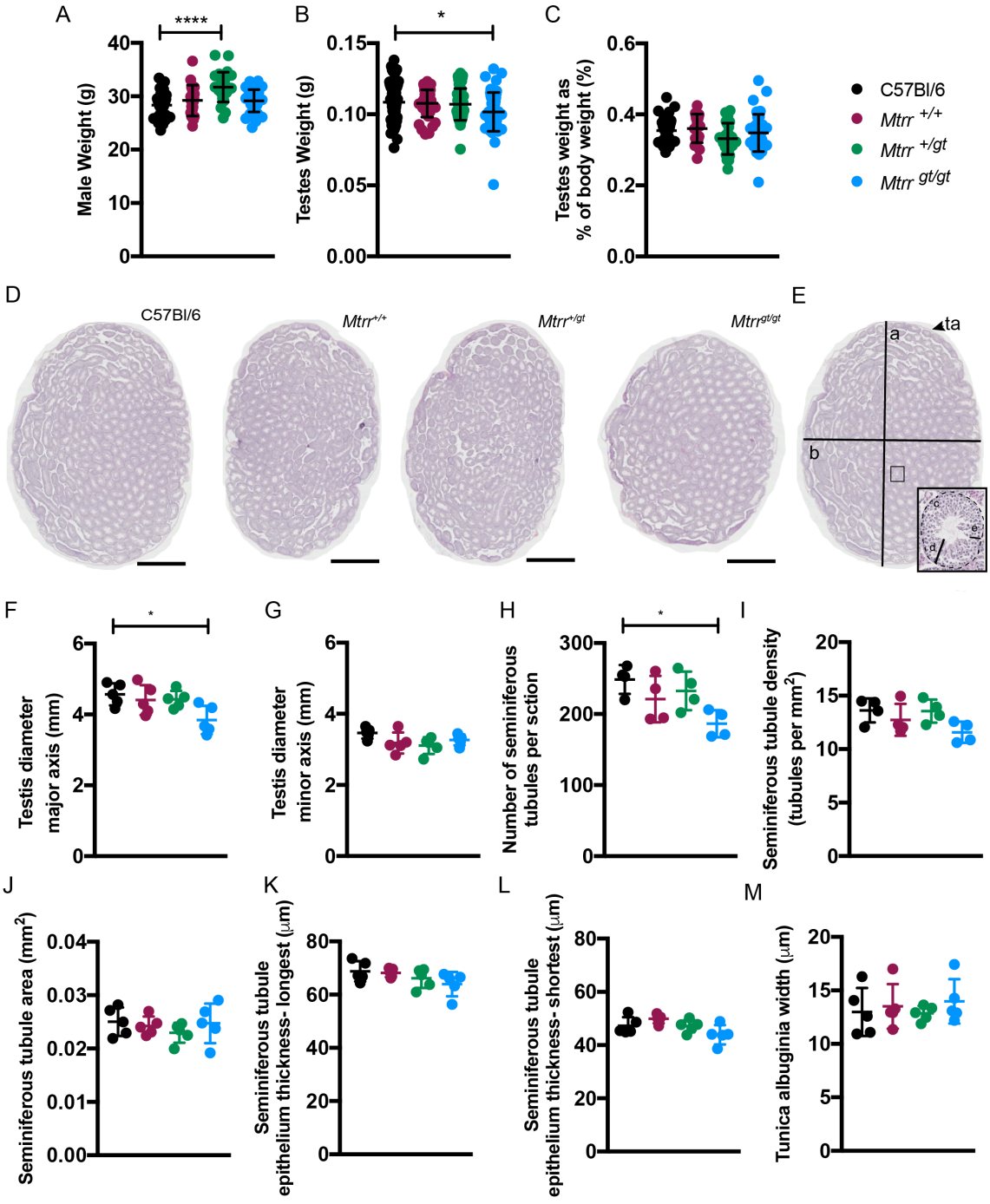

A
B
C
****
*
Male Weight (g)
Testes Weight (g)
Testes weight as % of body weight (%)
C57Bl/6
Mtrr +/+
Mtrr +/gt
Mtrr gt/gt
D
C57Bl/6
Mtrr+/+
Mtrr+/gt
Mtrrgt/gt
E
a
ta
b
c
d
e
F
G
H
I
Testis diameter major axis (mm)
Testis diameter minor axis (mm)
Number of seminiferous tubules per sction
Seminiferous tubule density (tubules per mm2)
J
K
L
M
Seminiferous tubule area (mm2)
Seminiferous tubule epithelium thickness- longest (µm)
Seminiferous tubule epithelium thickness- shortest (µm)
Tunica albuginia width (µm)

Figure 3.3: ***Mtrr$^{gt/gt}$* testes are more spherical in shape with fewer seminiferous tubules. (A-C)** Graphs showing **(A)** average male body weight, **(B)** average testis weight, and **(C)** testes weight as a percentage of body weight for male mice of the following genotypes: C57Bl/6 controls (black), *Mtrr*$^{+/+}$ (purple), *Mtrr*$^{+/gt}$ (green) and *Mtrr*$^{gt/gt}$ (blue). Data is plotted as mean ± sd. N=20-64 males per genotype. **(D)** Representative histological sections of whole testes stained with haematoxylin and eosin. C57Bl/6, *Mtrr*$^{+/+}$, *Mtrr*$^{+/gt}$ and *Mtrr*$^{gt/gt}$ testes are shown. Scale bar: 1 mm. **(E)** Histological section of C57Bl/6 control testis illustrating the parameters measured in F-G and J-L. The box represents a region of higher-magnification shown in the inset. Dotted line, outline of seminiferous tubule; a, major axis of testes diameter; b, minor axis of testes diameter; c, seminiferous tubule area; d, longest seminiferous tubule epithelium width; e, shortest seminiferous tubule epithelium width, ta, tunica abuginea. **(F-M)** Data representing measurements of testes and seminiferous tubule morphology for C57Bl/6 controls (black), *Mtrr*$^{+/+}$ (purple), *Mtrr*$^{+/gt}$ (green) and *Mtrr*$^{gt/gt}$ (blue). Data is plotted as mean ± sd. Testes from at least 4 males were assessed per genotype. Each point represents the average value per male as determined by assessing at least six histological sections per male. Measurements on histological sections of testes included: **(F-G)** testis diameter along the **(F)** major axis and **(G)** minor axis, **(H)** the average number of seminiferous tubules per histological section of testis, **(I)** seminiferous tubules density, **(J)** average seminiferous tubule area, **(K-L)** seminiferous tubule epithelial width determined at the **(K)** longest and **(L)** shortest, and **(M)** the thickness of the tunica albuginea that encapsulates the entire testis. Statistical tests: one-way ANOVA tests with Dunnett's multiple comparisons test. *p<0.05, ****p<0.0001.

### 3.2.3 Testes function and spermatogenesis is unaltered by the *Mtrr$^{gt}$* mutation

Since I observed overall changes in testes shape and the number of seminiferous tubules was reduced in *Mtrr*$^{gt/gt}$ testes, I wanted to assess whether testes function was altered by the *Mtrr*$^{gt}$ mutation. Firstly, I assessed Leydig cell function. Leydig cells are somatic cells localised to the interstitial space between seminiferous tubules. They produce testosterone in the presence of luteinizing hormone (Vasta et al., 2006), which drives spermatogenesis. To determine whether the *Mtrr*$^{gt}$ allele alters Leydig cell function, I measured serum testosterone levels using an ELISA. Although highly variable, the average serum testosterone concentration in *Mtrr*$^{+/+}$, *Mtrr*$^{+/gt}$, and *Mtrr*$^{gt/gt}$ males was statistically similar to C57Bl/6 males (Figure 3.4A). The variability may in part reflect poor quality blood samples. I also quantified the number of Sertoli

cells. Sertoli cells are somatic cells within the seminiferous tubule that provide the niche for spermatogenic stem cell development (Smith and Walker, 2014). These cells were characterised by a triangular shaped nucleus and prominent nucleolus. Using histological sections of testes, the average number of Sertoli cells per seminiferous tubule section was similar in C57Bl/6, *Mtrr*$^{+/+}$, *Mtrr*$^{+/gt}$ and *Mtrr*$^{gt/gt}$ testes (Figure 3.4B). Together, this suggests that supporting cells of the testes are likely unaffected by *Mtrr* deficiency.

To explore whether the *Mtrr* mutation directly impacts spermatogenesis, seminiferous tubules were staged based on the spermatogenic cell types present and the characteristics of the spermatid acrosome (Carrell and Aston, 2013). Stages I-XII cover the process of spermatogenesis from spermatogonium to elongated spermatid. As such, seminiferous tubule staging allows key processes of spermatogenesis such as acrosome formation, progression of the waves of spermatogenesis, and maturation of the germinal epithelium to be assessed (Carrell and Aston, 2013). Overall, normal progression of spermatogenesis was apparent in C57Bl/6, *Mtrr*$^{+/+}$, *Mtrr*$^{+/gt}$ and *Mtrr*$^{gt/gt}$ seminiferous tubules. No significant difference in the relative proportions or overt morphology of cells at each spermatogenic stage was observed compared to controls (Figure 3.4C-D). Therefore, spermatogenesis appears to progress in a normal manner along the seminiferous tubule in *Mtrr*$^{gt/gt}$ homozygous mice regardless of the size and shape change of the testis.

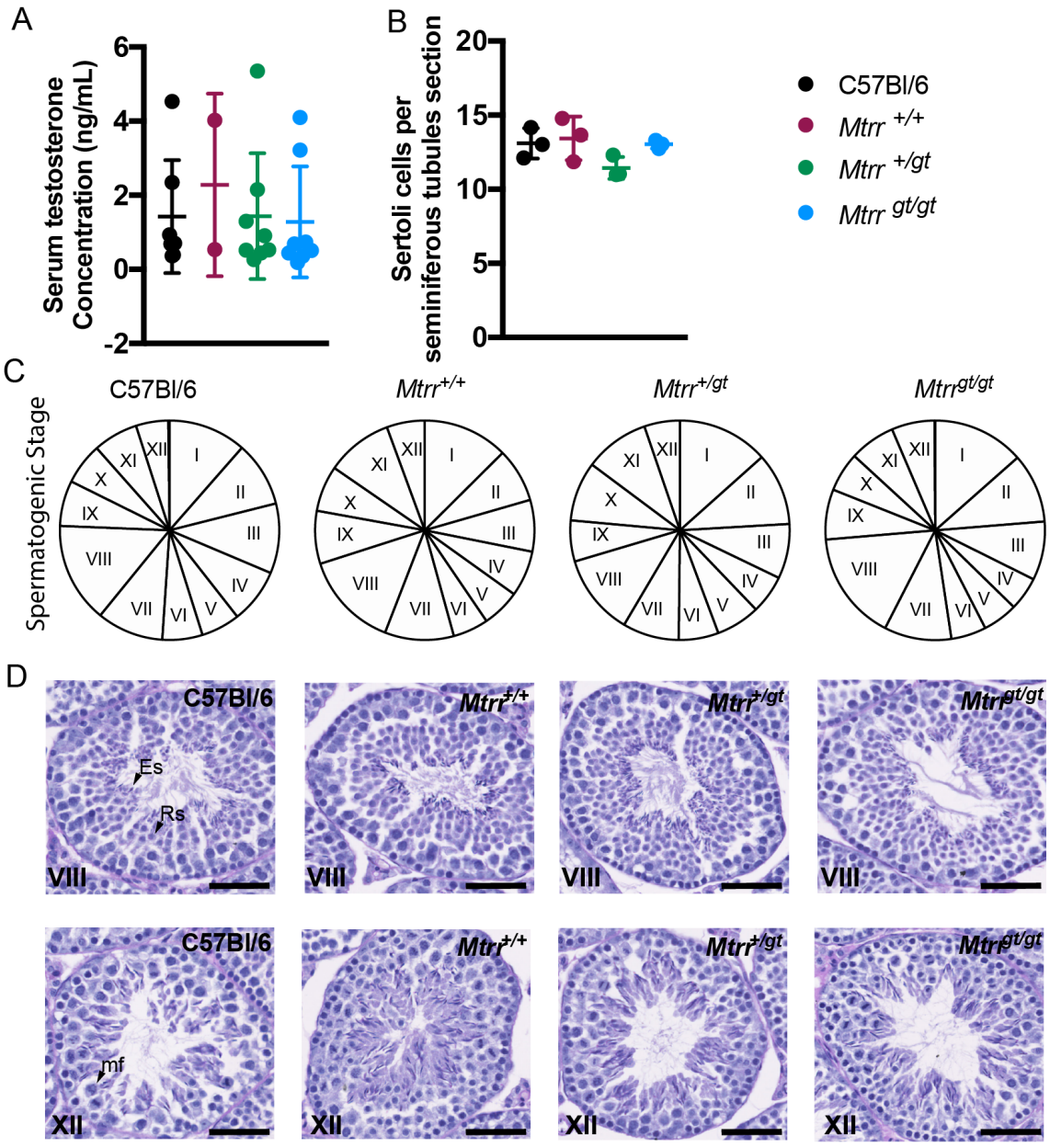

A
Serum testosterone Concentration (ng/mL)
6
4
2
0
-2
B
Sertoli cells per seminiferous tubules section
20
15
10
5
0
C57Bl/6
Mtrr +/+
Mtrr +/gt
Mtrr gt/gt
C
Spermatogenic Stage
C57Bl/6
Mtrr+/+
Mtrr+/gt
Mtrrgt/gt
I
II
III
IV
V
VI
VII
VIII
IX
X
XI
XII
D
C57Bl/6
Mtrr+/+
Mtrr+/gt
Mtrrgt/gt
Es
Rs
VIII
mf
XII

Figure 3.4: ***Mtrr* deficiency does not alter testes function**. **(A)** Serum testosterone concentrations as determined by ELISA in C57Bl/6 (black), $Mtrr^{+/+}$ (purple), $Mtrr^{+/gt}$ (green) and $Mtrr^{gt/gt}$ (blue). Data is plotted as mean ± sd (n=2-8 mice per genotype). **(B)** Graph showing the average number of Sertoli cells in the seminiferous tubules of C57Bl/6 (black), $Mtrr^{+/+}$ (purple), $Mtrr^{+/gt}$ (green) and $Mtrr^{gt/gt}$ (blue) testes. Data is plotted as mean ± sd (n=3 males per genotype). Each data point represents the average Sertoli cell number in seminiferous tubules within at least three histological sections per male. **(C)** The proportion of seminiferous tubules categorised as one of the twelve stages of spermatogenesis (stage I-XII) in C57Bl/6, $Mtrr^{+/+}$, $Mtrr^{+/gt}$ and $Mtrr^{gt/gt}$ testes. An average of 220 tubules were staged per section. Six histological sections were assessed per testes. One testis was assessed per male and four mice were assessed per genotype. Statistical analyses performed: (A-B) one-way ANOVA tests, (C) two-way ANOVA test, both with Dunnett's multiple comparisons test. **(D)** Representative images of seminiferous tubules for each genotype at key stages of spermatogenesis. Stages VIII (top panels) and XII (bottom panels) are shown. Key cell types used to classify the stage are labelled with arrowheads: Es, elongated spermatid; Rs, Round spermatid; mf, meiotic figures. Histological sections were stained with PAS stain. Scale bars: 50$\mu$m

### 3.2.4 Sperm parameters and fertility are normal in $Mtrr^{gt/gt}$ males

*Mtrr* deficient mice are fertile. This correlates with the normal spermatogenesis observed in *Mtrr* deficient testes. However, whether spermatozoa from $Mtrr^{gt/gt}$ males were capable of maturing and, ultimately, fertilizing an oocyte at a similar rate as controls had not previously been systematically assessed. Therefore, I first analysed mature spermatozoa extracted from the cauda epididymis. Sperm were counted using a haemocytometer, and normalised by cauda epididymis weight. Initial studies performed sperm counts on males over a range of ages from 8 to 29 weeks. In this data set, the sperm counts of $Mtrr^{+/+}$ and $Mtrr^{+/gt}$ males were elevated with respect to C57Bl/6 controls (p=0.002 and p=0.0092, respectively, Figure 3.5A). $Mtrr^{gt/gt}$ male sperm counts were equivalent to C57Bl/6 controls (Figure 3.5A). Conversely, sperm viability, assessed using supravital staining, was reduced in $Mtrr^{gt/gt}$ males versus C57Bl/6 controls (p=0.0003, Figure 3.5B). Sperm viability was unaffected in $Mtrr^{+/+}$ and $Mtrr^{+/gt}$ males (Figure 3.5B). No differences in sperm morphology were noted between C57Bl/6 controls and any *Mtrr* sperm (Figure 3.5C). However, I believe the differences in sperm counts and viability might have reflected

differences in the average age of the males assessed in each group (Average male age of C57Bl/6: 17 weeks, $Mtrr^{gt/gt}$: 19 weeks, $Mtrr^{+/gt}$: 16 weeks, $Mtrr^{+/+}$: 12 weeks). Indeed, a negative correlation was observed when sperm count was plotted against male age (Figure 3.5D). Therefore I reanalysed our data, considering only 16-20 week old males. In these 16-20 week old males, similar numbers of sperm were observed in C57Bl/6, $Mtrr^{+/+}$, $Mtrr^{+/gt}$ and $Mtrr^{gt/gt}$ groups ($p>0.11$; Figure 3.5E). Additional sperm counts should be performed to improved the validity of this data. Sperm viability analysis also showed comparable percentages of viable sperm were present in each genotype analysed when only 16-20 week old males were considered (96.9 to 97.4% viable sperm; $p>0.94$; Figure 3.5F). In addition, an equivalent proportion of spermatozoa (at least 90%) from C57Bl/6, $Mtrr^{+/+}$, $Mtrr^{+/gt}$ and $Mtrr^{gt/gt}$ males showed a normal hooked morphology (Figure 3.5G). The frequency of abnormal sperm morphologies including hookless ($\sim$1%), headless ($\sim$2-5%), or amorphous ($\sim$5-6%) sperm was also similar between genotypes (Figure 3.5G). This data provides further evidence to support the conclusion that sperm quantity and quality decrease with age, including in the *Mtrr* deficient mice. Overall, this data supports a finding that abnormal folate metabolism caused by the *Mtrr* allele does not affect sperm viability, morphology or number.

To assess fertility in $Mtrr^{gt/gt}$ males, I retrospectively analysed the ability of C57Bl/6 and $Mtrr^{gt/gt}$ males to produce a copulatory plug and the timespan between the establishment of a mating pair to the detection of the plug. I found that 100% of C57Bl/6 males (52/52 males) and 95.2% of $Mtrr^{gt/gt}$ males (40/42 males) produced a copulatory plug within seven days. $Mtrr^{gt/gt}$ male mice took an average of $2.7 \pm 0.2$ days to copulate with a female mouse, which was slightly longer than the $2.1 \pm 0.2$ days ($p=0.03$) for C57Bl/6 control males (Figure 3.5H). As this was a retrospective analysis, an important caveat to this experiment is that the female genotype was not taken into account. $Mtrr^{gt/gt}$ males were more often mated to $Mtrr^{gt/gt}$ females, which is likely a confounding factor when determining the plug rate. When only matings with C57Bl/6 or $Mtrr^{+/+}$ females were considered, there was no apparent difference in time taken to copulate between C57Bl/6 and $Mtrr^{gt/gt}$ males ($p=0.84$, Figure 3.5I). To get an estimation of the rate at which $Mtrr^{gt}$ sperm were able to fertilise oocytes,

I retrospectively assessed the ability of *Mtrr*$^{gt/gt}$ males to generate pregnancies after a copulatory plug was detected. This was determined by the presence of implantation sites in the uterus or pups littered. Pregnancies were generated at a slightly, but not significantly, reduced rate between *Mtrr*$^{gt/gt}$ males (87.5%, N=40) and C57Bl/6 males (98.1%, N=52; p=0.08, Fisher's exact test) (Figure 3.5J) regardless of female *Mtrr* genotype. Therefore, male fertility is likely unaffected by *Mtrr*$^{gt/gt}$ homozygosity.

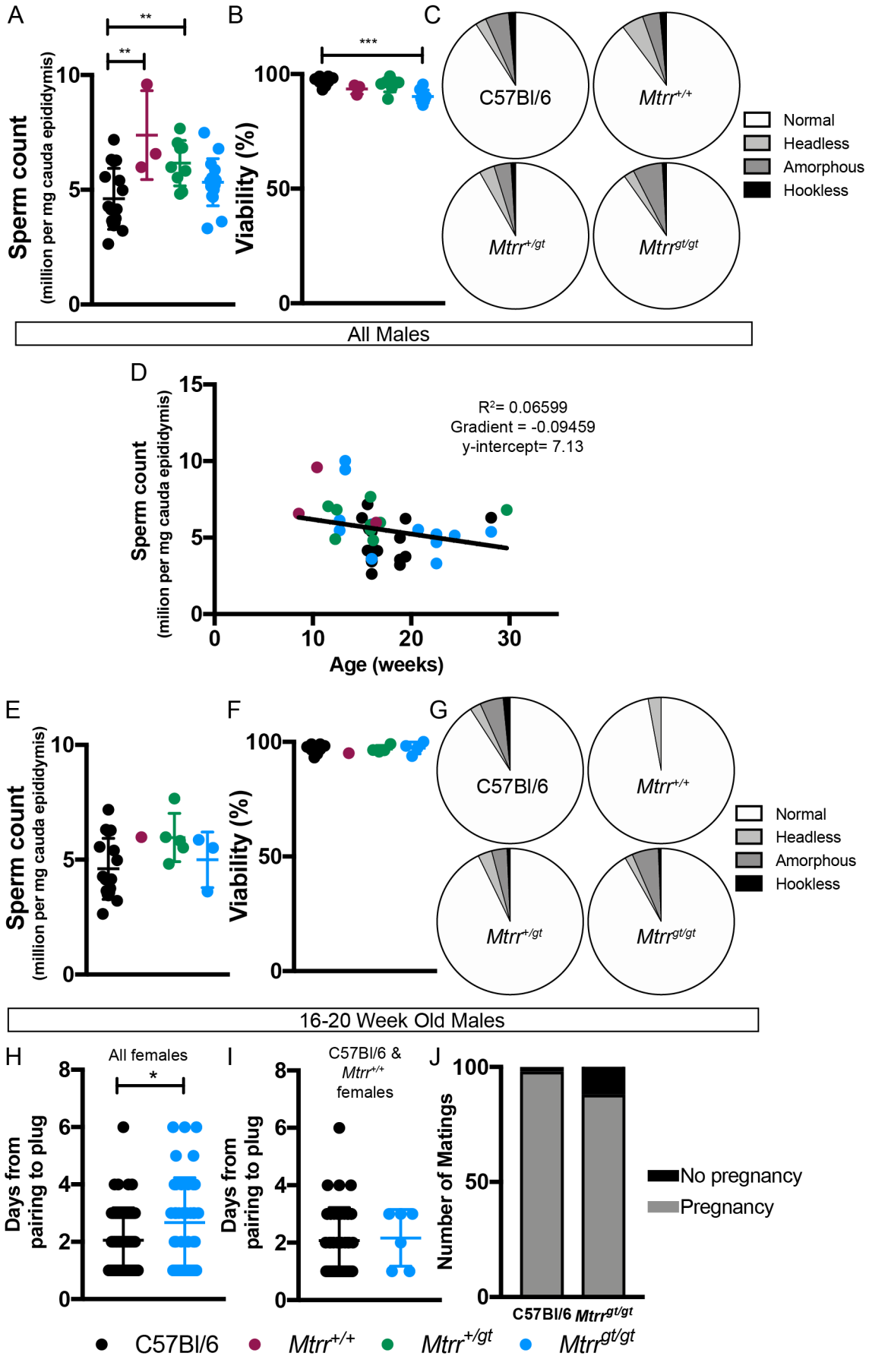

A
B
C
Sperm count
(million per mg cauda epididymis)
Viability (%)
C57Bl/6
Mtrr+/+
Mtrr+/gt
Mtrrgt/gt
Normal
Headless
Amorphous
Hookless
All Males
D
Sperm count
(milion per mg cauda epididymis)
R²= 0.06599
Gradient = -0.09459
y-intercept= 7.13
Age (weeks)
E
F
G
16-20 Week Old Males
H
All females
Days from pairing to plug
I
C57Bl/6 & Mtrr+/+ females
J
Number of Matings
No pregnancy
Pregnancy
C57Bl/6
Mtrr+/+
Mtrr+/gt
Mtrrgt/gt

Figure 3.5: **Sperm count, sperm viability, and fertility are normal in *Mtrr*$^{gt/gt}$ males.** Graphical data is shown for C57Bl/6 (black), *Mtrr*$^{+/+}$ (purple), *Mtrr*$^{+/gt}$ (green) and *Mtrr*$^{gt/gt}$ (blue) males. **(A-C)** Data is shown for males ranging from 8-29 weeks old. **(A)** Data showing the average sperm count normalised to cauda epididymis weight. Data is plotted as mean ± sd. N=3-17 males per genotype. **(B)** Graph depicting the percentage of total sperm that was viable as determined in eosin/nigrosin smears. Data is plotted as mean ± sd. N=3-8 males per genotype. **(C)** Data showing the proportion of sperm with normal morphology (white) or abnormal morphology including headless sperm (light grey), amorphous sperm (dark grey) or hookless sperm (black) in C57Bl/6, *Mtrr*$^{+/+}$, *Mtrr*$^{+/gt}$ and *Mtrr*$^{gt/gt}$ males. N=3-8 males per genotype. **(D)** A plot of sperm count per milligram of cauda epididymis against male age in weeks. Line of best fit depicted, y-intercept: 7.13, slope: -0.09459. N=41 males. **(E-G)** Data is shown for males ranging from 16-20 weeks old only. **(E)** Data showing the average sperm count normalised to cauda epididymis weight. Data is plotted as mean ± sd. N=1-17 males per genotype. **(F)** Graph depicting the percentage of total sperm that was viable as determined in eosin/nigrosin smears. Data is plotted as mean ± sd. N=1-8 males per genotype. **(G)** Data showing the proportion of sperm with normal morphology (white) or abnormal morphology including headless sperm (light grey), amorphous sperm (dark grey) or hookless sperm (black) in C57Bl/6, *Mtrr*$^{+/+}$, *Mtrr*$^{+/gt}$ and *Mtrr*$^{gt/gt}$ males. N=1-8 males per genotype. **(H-I)** Graphs indicating a retrospective analysis of the average number of days between the establishment of a mating pair and detection of the copulatory plug for C57Bl/6 and *Mtrr*$^{gt/gt}$ males. Data is presented as mean ± sd. **(H)** Males were paired with C57Bl/6, *Mtrr*$^{+/+}$, *Mtrr*$^{+/gt}$ or *Mtrr*$^{gt/gt}$ female mice. N=52 C57Bl/6 males and 42 *Mtrr*$^{gt/gt}$ males. **(E)** Only matings with C57Bl/6 and *Mtrr*$^{+/+}$ females were considered. N=45 C57Bl/6 males and 6 *Mtrr*$^{gt/gt}$males. **(J)** The percentage of copulatory plugs generated by C57Bl/6 or *Mtrr*$^{gt/gt}$ males that resulted in pregnancy. Pregnancy was determined by the appearance of implantation sites in the uterus or the presence of a litter at birth. N=40-52 litters. Statistical analyses included: (A-B, E-F) one-way ANOVA tests, (H-I) unpaired independent t-tests, (J) Fishers exact test. *$p<0.05$, **$p<0.0005$, ***$p<0.0005$.

### 3.2.5 Expression of *Dnmt2* is reduced in the epididymis of *Mtrr* mice

Dynamic changes in epigenetic landscape occur as sperm differentiate and mature, both during spermatogenesis and as they pass along the epididymis (Oakes et al., 2007; Loukinov et al., 2002). Folate metabolism is required to transmit one-carbon methyl groups for DNA methylation (Jacob et al., 1998) and *Mtrr* deficiency causes dysregulation of DNA methylation patterns (Pad-

manabhan et al., 2013). Therefore, I aimed to understand whether *Mtrr* deficiency altered the expression levels of the genes encoding for the DNA methylation machinery, including the DNMT and TET enzymes, in the testis or epididymis. Disrupted DNMT or TET expression could contribute to defects in the establishment and/or maintenance of DNA methylation profiles in sperm. Additionally, sperm are known to take up ncRNAs during transit through the epididymis (Sharma et al., 2016), and these ncRNAs may carry RNA modifications (Chen et al., 2015). Therefore, I also examined the gene expression of the RNA methylation machinery, namely *Dnmt2* and *Nsun2*.

Using RT-qPCR analysis, I determined that mRNA expression of *Dnmt1*, *Dnmt3a*, *Dnmt3b*, and *Dnmt3c* genes in $Mtrr^{+/+}$, $Mtrr^{+/gt}$, and $Mtrr^{gt/gt}$ testes and epididymides were comparable to C57Bl/6 controls (Figure 3.6A-D, H-K). Likewise, *Mtrr* deficiency did not alter the expression *Tet1*, *Tet2* and *Tet3* genes in either testes or epididymides (Figure 3.6E-G, L-N). This data suggest that despite widespread DNA methylation dysregulation in the $Mtrr^{gt}$ model, gene expression of the DNA methylation machinery is unchanged in either the testes or epididymis.

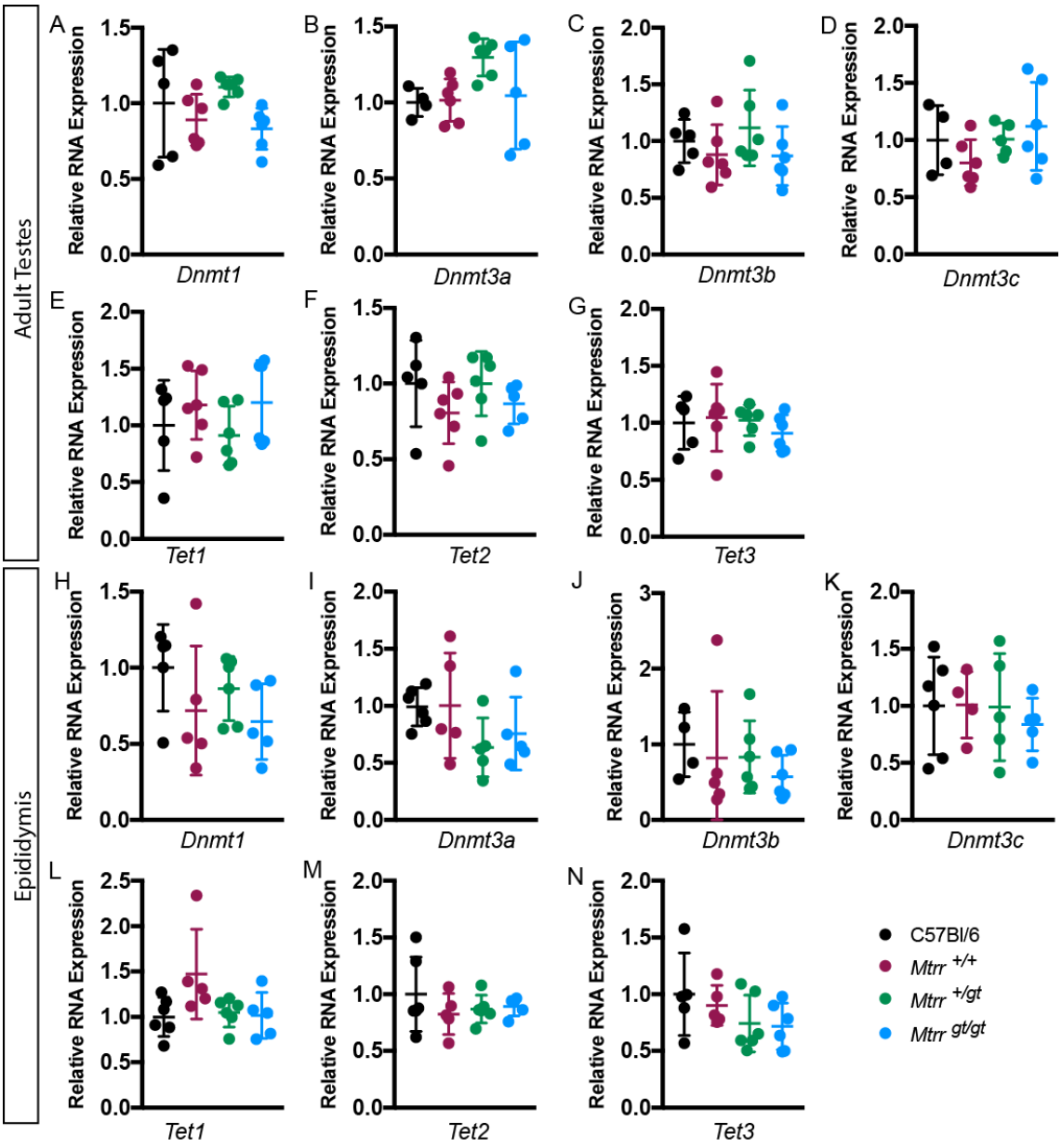



Figure 3.6: **DNA methylation machinery expression is comparable to controls in *Mtrr*$^{+/+}$, *Mtrr*$^{+/gt}$, and *Mtrr*$^{gt/gt}$ testes and epididymides.** Graphical data is shown for C57Bl/6 (black), *Mtrr*$^{+/+}$ (purple), *Mtrr*$^{+/gt}$ (green) and *Mtrr*$^{gt/gt}$ (blue) males in testis (A-G) and epididymis (H-N). Data is plotted as mean ± standard deviation. Graphs show RT-qPCR analysis of the expression of **(A,H)** *Dnmt1*, **(B,I)** *Dnmt3a*, **(C,J)** *Dnmt3b*, **(D,K)** *Dnmt3c*, **(E,L)** *Tet1*, **(F,M)** *Tet2* and **(G,N)** *Tet3*. Data is represented as fold change compared to C57Bl/6 controls, which are normalised to 1. N=4-6 individuals per genotype. Statistical analysis performed was one-way ANOVA.

Again using RT-qPCR analysis, I determined that mRNA expression of *Dnmt2* and *NSun2* genes was equivalent to C57Bl/6 controls in $Mtrr^{+/+}$, $Mtrr^{+/gt}$, and $Mtrr^{gt/gt}$ testes (Figure 3.7A-B). However, the mRNA expression of *Dnmt2* was significantly reduced in $Mtrr^{+/+}$, $Mtrr^{+/gt}$, and $Mtrr^{gt/gt}$ epididymides versus C57Bl/6 controls ((p<0.0024, Figure 3.7C). Interestingly, the expression of *Nsun2* was equivalent to C57Bl/6 controls in all *Mtrr* genotype epididymides (Figure 3.7D). This data shows that that disruption of folate metabolism caused by the $Mtrr^{gt}$ mutation, or parental exposure to a $Mtrr^{gt}$allele, leads to altered *Dnmt2* expression in the epididymis. This raises the exciting possibility that sperm RNA modifications may be perturbed.

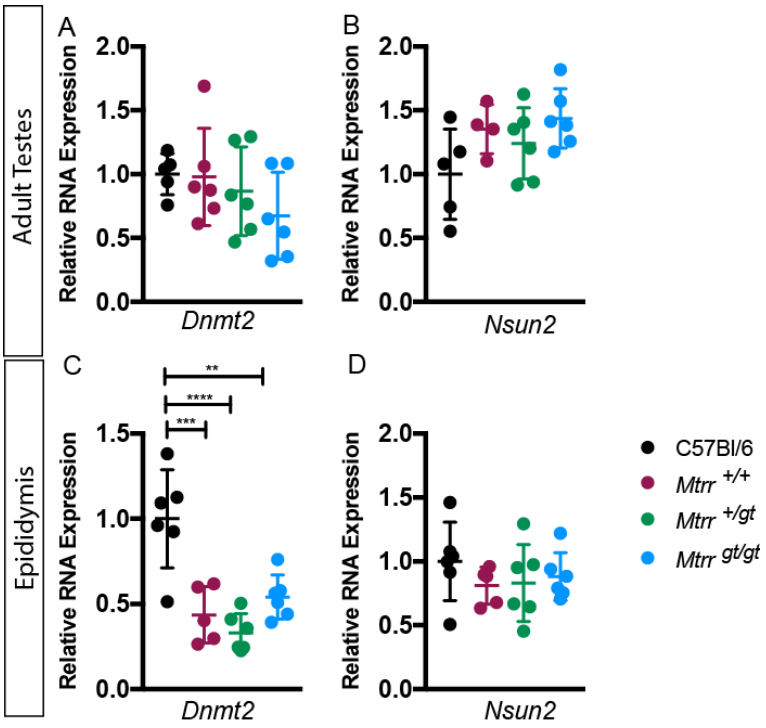



Figure 3.7: ***Dnmt2* expression is reduced in $Mtrr^{+/+}$, $Mtrr^{+/gt}$, and $Mtrr^{gt/gt}$ epididymides versus C57Bl/6 controls**. Graphical data is shown for C57Bl/6 (black), $Mtrr^{+/+}$ (purple), $Mtrr^{+/gt}$ (green) and $Mtrr^{gt/gt}$ (blue) males in (A-B) testis and (C-D) epididymis. Data is plotted as mean ± standard deviation. Graphs show RT-qPCR analysis of the expression of **(A,C)** *Dnmt2*, and **(B,D)** *Nsun2*. Data is represented as fold change compared to C57Bl/6 controls, which are normalised to 1. N=4-6 individuals per genotype. Statistical analysis performed was one-way ANOVA, with Dunnett's multiple comparisons test. ** p<0.005, *** p<0.0005.

## 3.3 Discussion

Understanding the effect of *Mtrr* deficiency on spermatogenesis and testes function is important in the context of TEI within the *Mtrr*$^{gt}$ mouse model. Here I have established that MTRR protein was widely expressed throughout the adult testis. Other proteins involved in folate metabolism and uptake, MTHFR, FOLR1 and SLC19A1, had similar expression patterns suggesting that folate metabolism occurs broadly in the mouse testis. The *Mtrr*$^{gt}$ mutation resulted only in a mild testes phenotype in *Mtrr*$^{gt/gt}$ males. *Mtrr*$^{gt/gt}$ testes were more spherical in shape with fewer seminiferous tubules compared to C57Bl/6 controls. However, this gross morphological phenotype did not appear to have functional consequences since serum testosterone levels, spermatogenesis, sperm counts and viability, and fertility were normal in *Mtrr*$^{gt/gt}$ males. *Mtrr*$^{+/gt}$ and *Mtrr*$^{+/+}$ males had normal testis and sperm morphology and function. However, the expression of the RNA methyltransferase *Dnmt2* was specifically reduced in *Mtrr*$^{+/+}$, *Mtrr*$^{+/gt}$ and *Mtrr*$^{gt/gt}$ epididymides, despite normal expression of all other DNA methylation machinery enzymes in both testes and epididymis. Overall, since no morphological or functional abnormalities are apparent in germ cells of *Mtrr*$^{gt/gt}$, or more importantly *Mtrr*$^{+/gt}$ males, I conclude that defects in spermatogenesis or sperm function are unlikely a contributing factor to the transgenerational inheritance of developmental abnormalities observed in this model. However, reduced *Dnmt2* expression in the epididymis in both males with, or parentally exposed to, the *Mtrr*$^{gt}$ allele suggests RNA modifications could play a role in TEI in this model.

However, these conclusions are drawn with an appreciation that the methods used to assess testes morphology had limitations. Histological analysis may have been affected by technical variations in processing such as the angle and position of testis within the wax block during sectioning, the position of the section within the testes, tissue shrinkage, fixation artefacts and dehydration. As seminiferous tubules are convoluted, counting seminiferous tubule cross-sections may not give a true reflection of tubule number. Performing stereology on serial sections may give a more accurate representation of testes morphology. Although the cells within the histological sections of seminiferous tubules appeared to have a normal morphology and were represented in normal num-

bers, this could be quantified more thoroughly using immunohistochemical techniques with antibodies against markers of specific cell types (e.g androgen-binding protein (ABP) as a Sertoli cell marker (Hansson et al., 1975)).

The scope of our analysis on sperm counts and functionality was also somewhat limited. We performed sperm counts on sperm isolated from the cauda epididymis only. Determination of total epididymal sperm count, as a measure of sperm output, and the number of homogenization-resistant spermatids in a unit mass of testicular tissue, as a measure of sperm reserve, would provide a better estimate of sperm production efficiency and also epididymal sperm reserves (Peirce and Breed, 2001). This would give a better indication of spermatogenic efficiency. More thorough and in-depth analysis of sperm functionality may reveal subtle phenotypes, undetectable using our methods but suggested by the reduction in successful pregnancies following copulation observed in $Mtrr^{gt/gt}$males (Figure 3.5 J). *In vitro* fertilization techniques using $Mtrr^{gt/gt}$ (or C57Bl/6 control) sperm and C57Bl/6 oocytes could be used to determine more accurately if there is reduced fertilisation capacity in $Mtrr^{gt/gt}$ males. *In vitro* tests of sperm penetration, such as the Hemizona Assay, could also be performed (Carrell and Aston, 2013). To better characterise sperm parameters, more advance technologies, such as computer aided sperm analysis (CASA), could be utilised. This would allow parameters such as sperm motility, known to be important for fertility (Larsen et al., 2000), to be measured. Our conclusion that Leydig cell function was normal, as determined by a serum testosterone ELISA, should be interpreted with an understanding that mild haemolysis occurred in some blood samples, reducing the purity of serum. Blood collection via an alternative method to cardiac puncture after cervical dislocation, such as tail vein bleed, may provide a better quality blood sample. Additionally, testosterone levels should be measured in more individuals to account for biological variability, particularly $Mtrr^{+/+}$ males as only two individuals were assessed in the current analysis. Furthermore, an important aspect not considered in this study is seminal fluid. It has the potential to influence the female reproductive tract and has been reported to convey information regarding parental diet that influences offspring phenotype (Watkins et al., 2018). Therefore, it should be evaluated in the $Mtrr^{gt}$ model.

The testes phenotype observed in *Mtrr*$^{gt/gt}$ males is relatively mild. This may be because the *Mtrr*$^{gt}$allele is a hypomorphic mutation. In fact, wildtype *Mtrr* transcript is present in *Mtrr*$^{gt/gt}$ homozygous tissue, although at very low levels in the testes (Figure 3.1D; Padmanabhan et al. (2013); Elmore et al. (2007)). How this occurs in the absence of a wildtype *Mtrr* allele is unclear, though splicing out of the gene-trap that causes the *Mtrr*$^{gt}$ mutation is suspected (Padmanabhan et al., 2013; Elmore et al., 2007). In *Mtrr*$^{gt/gt}$ testis, the level of wildtype *Mtrr* mRNA expression is only 14% of that in C57Bl/6 testis. This was lower than all of the other *Mtrr*$^{gt/gt}$ tissue types that have been tested (e.g., heart, uterus, brain, kidney, liver), which have expression ranging from 19.3-35.7% of control levels (Padmanabhan et al., 2013). This suggests splicing may occur less efficiently in testes than other tissues. Despite the much reduced expression of *Mtrr*, *Mtrr*$^{gt/gt}$ testes function was normal and generated mature and functioning sperm. Perhaps a compensatory mechanism exists in testes to protect sperm against the metabolic insult of *Mtrr* deficiency. This could be similar to the normalisation of folate concentrations in fetal blood that occurs during maternal folate deficiency in an attempt to maintain normal fetal growth (Ek, 1980). It will be necessary to measure folate and Ado-Met concentrations in *Mtrr*$^{gt/gt}$ testes and seminal fluid to test this hypothesis. Interestingly, folate levels in seminal plasma of men are greater than in blood plasma (Wallock et al., 2001). However, no transcriptional compensation by other genes involved in folate transport and metabolism occurs in *Mtrr*-deficient testes. A complete knockout of the *Mtrr* gene might cause a more severe testes or fertility phenotype.

Diminished cell proliferation may account for the reduced size and abnormal shape of *Mtrr*$^{gt/gt}$ testes (Figure 3.3). It is known that disruption of folate metabolism leads to altered nucleotide pools (Bistulfi et al., 2010) and hypomethylation of cellular substrates (Wasson et al., 2006; Dobosy et al., 2008; Waterland et al., 2006). Though it is poorly understood, determination of testes size likely occurs during highly proliferative phases of fetal and perinatal development (Sharpe, 2006; Svingen and Koopman, 2013). Therefore, investigating the effects of *Mtrr* deficiency on testes development in embryos and in early post-natal mice will help to explore the mechanism behind the small, round testes phenotype.

The mild testes phenotype and normal sperm parameters observed in the *Mtrr*$^{gt}$ model are consistent with a model of dietary folate deficiency in C57Bl/6 mice, where no effect of diet was reported on spermatogenesis or sperm counts (Lambrot et al., 2013). Based on *Mthfr* knockout mouse strains, genetic background effects seem to be important when considering folate metabolism and testes phenotype. *Mthfr*$^{-/-}$ BALB/c mice display oligospermia and infertility (Kelly et al., 2005; Chan et al., 2010), whereas *Mthfr*$^{-/-}$ C57Bl/6 mice have a low sperm count but are fertile (Chan et al., 2010). *Mtrr*$^{gt/gt}$ males (which have a C57Bl/6 genetic background) display a testes phenotype most similar to, but less severe than, C57Bl/6 mice that are *Mthfr*$^{-/-}$ mutant including reduced testes weight with normal fertility (Chan et al., 2010). The *Mtrr*$^{gt}$ allele has not been bred into an alternative genetic background. To do this would allow us to explore the implications of the genetic background of *Mtrr*-deficiency on testis function.

A dose-specific metabolic effect of the *Mtrr*$^{gt}$ allele was further evident when we assessed male body weight. We observed that *Mtrr*$^{+/gt}$ not *Mtrr*$^{gt/gt}$ males had increased body weight compared to C57Bl/6 control males. Interestingly, this finding is correlated with plasma total homocysteine concentrations. As expected, *Mtrr*$^{gt/gt}$ males display hyperhomocysteinemia with respect to C57Bl/6 controls (Padmanabhan et al., 2013). However, plasma total homocysteine levels are reduced in *Mtrr*$^{+/gt}$ males compared to C57Bl/6 males (Padmanabhan et al., 2013). The reason behind this reduction and its relationship to body weight is unclear and should be explored further. Intriguingly, the increased body weight in heterozygous males appears to be late onset (16-20 weeks of age) since *Mtrr*$^{+/gt}$ males at 11 weeks have normal body weights (Elmore et al., 2007). However, this finding might be based on the control used: Elmore et al. used *Mtrr*$^{+/+}$ males whereas we used C57Bl/6 males in this study since ancestral *Mtrr* deficiency was shown to affect *Mtrr*$^{+/+}$ mice (Padmanabhan et al., 2013). Similar to Elmore et al., when body weight of *Mtrr*$^{+/gt}$ males at 16-20 weeks was compared to *Mtrr*$^{+/+}$ males in our study, no significant difference was observed.

We observed that mRNA expression of the DNA methylation machinery, the DNMT and TET enzymes responsible for DNA methylation and demethy-

lation respectively, was not altered in the testes or epididymides of the *Mtrr* model. This however does not exclude the possibility that abnormal DNA methylation patterns may be established in the sperm. Consistent with this hypothesis are folate-deficient C57Bl/6 males that are fertile with normal sperm counts and yet the sperm have altered epigenomes that are associated with an increased risk of congenital malformations in their offspring (Lambrot et al., 2013). Epigenetic reprogramming, including widespread DNA demethylation, occurs in the developing germ cells as they migrate to and occupy the genital ridges (Reik and Surani, 2015). The male germ cells then enter mitotic arrest and reestablishment of DNA methylation patterns occur at the prospermatogonia stage at around E15 (Reik et al., 2001). There is evidence that epigenetic reprogramming can occur during spermatogenesis in adult testes (Oakes et al., 2007; Loukinov et al., 2002). Indeed, sperm DNA methylation patterns seem to be susceptible to environmental insults only received in adulthood (e.g. high fat diet (Wei et al., 2014)). Dysregulation of the DNA methylation machinery during embryonic stages of testes and germ cell development, rather than in the adult testes, may be more likely to lead to altered epigenetic marks within the mature sperm. Assessment of DNA methylation machinery in developing *Mtrr* testes would help determine if this were the case.

Remarkably we do see that the mRNA expression of the RNA methyltransferase *Dnmt2* is reduced in the epididymis of $Mtrr^{+/+}$, $Mtrr^{+/gt}$ and $Mtrr^{gt/gt}$ males. DNMT2 expression is required for tRNA fragment mediated inheritance of high fat diet induced phenotypes (Zhang et al., 2018; Chen et al., 2015). RNA modifications, particularly those on tRNAs, should therefore be explored in the *Mtrr* model. DNMT2 has also been implicated in the *Kit* paramutant phenotype, with inheritance of the paramutant phenotype not occurring in $Dnmt2^{-/-}$ mice (Kiani et al., 2013). Paramutation has been excluded as a mechanism in the *Mtrr* model, as expression of $Mtrr^{+/+}$ mRNA was normal in embryos and placentas derived from $Mtrr^{+/gt}$ parents (Padmanabhan et al., 2013). RNA methylation, mediated by DNMT2, has been implicated in tRNA stability, secondary structure and biological properties (Tuorto et al., 2012; Zhang et al., 2018), which perhaps provides insight into how RNA modifications could mediate phenotypic inheritance. Additional RNA modifying enzymes, such

as DEAD box polypeptide 1, which has also been implicated in non-genetic inheritance (Hildebrandt et al., 2015), should be assessed in the *Mtrr* model.

The status of *Mtrr*$^{+/gt}$ sperm is important in the context of TEI, as *Mtrr*$^{+/gt}$ males represent the maternal grandfather in the transgenerational inheritance pedigrees. Here we show that *Mtrr*$^{+/gt}$ heterozygosity does not affect testes morphology or sperm parameters. This suggests that it is unlikely that defects in testes function or spermatogenesis contribute to the mechanism behind TEI. A similar investigation probing folate metabolism in the ovary and exploring potential abnormal ovarian and oocyte morphology and function in the *Mtrr*$^{gt}$ model, specifically in *Mtrr*$^{+/+}$F1 generation females, will be necessary to completely rule out a gonadal or germ cell morphological contribution to the inheritance of phenotypes. However, the data presented here strengthens the hypothesis that a factor intrinsic to sperm, but independent of their structure, such as the genome or epigenome, is responsible for transgenerational phenotypic inheritance.

# Chapter 4

# Assessing genetic stability in *Mtrr*$^{gt}$ mice

Bioinformatics analysis in this chapter was performed in collaboration with Dr Russell Hamilton (Department of Genetics & Centre for Trophoblast Research, Cambridge, UK) and Dr Xiaohui Zhao (Centre for Trophoblast Research, Cambridge, UK). Validation of variants identified by whole genome sequencing was performed in collaboration the undergraduate students Daphne Chia and Panayiotis Laouris (Department of Physiology, Development and Neuroscience, Cambridge, UK).

## 4.1 Introduction

The folate and methionine cycles play vital roles in DNA synthesis and repair (Bistulfi et al., 2010; Blount et al., 1997; Duthie et al., 2000b). Folate is required for the *de novo* synthesis of the nucleotide thymidine. Thymidylate synthase catalyses the conversion of deoxyuridine monophosphate (dUMP) to deoxythymidine monophosphate (dTMP) by reductive methylation using a one-carbon unit from 5-10-methylenetetrahydrofolate (Avendao and Menndez, 2008) (Chapter 1, Figure 1.5). Dietary folate deficiency decreases dTMP synthesis, increasing the dUMP/dTMP ratio and resulting in dUTP (uracil) misincorporation into DNA (Bistulfi et al., 2010; Blount et al., 1997). Uracil misincorporation initiates DNA repair processes, such as base excision repair. Errors during DNA repair can lead to chromosome breaks and chromosomal aberrations (Blount et al., 1997; Cabelof et al., 2004; Duthie et al., 2000a; Duthie, 2011). With continued folate deficiency, permanent imbalances in the nucleotide pools can lead to repeated uracil misincorporation and a catastrophic repair cycle (Duthie et al., 2002).

There are numerous studies exploring the role of folate in DNA stability, particularity with regard to carcinogenesis (reviewed in Duthie (2011)). Folate deficient splenctomised human patients had increased micronucleated erythrocytes on blood smears compared to non-folate deficient patients. Micronuclei are normally removed by the spleen and are indicative of chromosomal damage (Blount et al., 1997). When given folate supplementation, these patients showed reduced uracil levels in their DNA (Blount et al., 1997). In fertile men, but not subfertile men, low folate concentrations in seminal plasma was associated with elevated DNA fragmentation index, as determined by the sperm chromatin structure assay (Boxmeer et al., 2009). Similarly, folate-deficient BALB/c mice showed increased sperm DNA fragmentation index and increased mutation frequency at expanded simple tandem repeats (Swayne et al., 2012). In cultured human lymphocytes, the detected levels of DNA strand breakage and uracil misincorporation correlated with the duration and severity of exposure to folate-deficient culture conditions (Duthie and Hawdon, 1998). Furthermore, a study of human lymphocytes used detailed analysis of nuclear morphology and observed increased nuclear abnormalities in folate-deficient

conditions, in part resulting from telomere-end fusion (Bull et al., 2012). Further investigations in another human cell line showed folate deficiency compromised telomere homoeostasis with telomere length fluctuations and increased presence of uracil in telomeres (Bull et al., 2014). This data suggests that some loci may be particularly sensitive to genetic instability resulting from folate deficiency.

However, there is also evidence to suggest that folate deficiency may not always lead to genetic instability. Colonocytes, harvested from folate-deficient Sprague-Dawley rats, showed impaired DNA excision repair rates but normal mismatch repair (Choi et al., 1998). Additionally, there was no instability at five mircosatellite loci compared to folate sufficient controls, even in the presence of a colonic carcinogen (Choi et al., 1998). In tumour susceptible $Apc^{Min}$ mice, folate deficiency was not associated with increased tumour number or size as might have been expected (Trasler et al., 2003). Intriguingly, in $Apc^{Min}$ mice with reduced *Dnmt1* expression (due to heterozygous mutation in the *Dnmt1* catalytic domain), folate deficiency led to fewer but larger tumours compared to mice on folate sufficient diets (Trasler et al., 2003). BALB/c mice fed a diet lacking choline and folic acid showed no increase in mutation rate at expanded simple repeats compared to chow fed controls (Voutounou et al., 2012). The mutation rate was also not increased in the F1 generation offspring (Voutounou et al., 2012). Altogether, these data suggest that folate deficiency may only influence genetic stability in certain circumstances. Discrepancies between studies likely result from differing severities, durations and timing of folate deficiency, either in the culture medium or diet. Varying susceptibilities to DNA damage in different cell types or in animals with different genetic backgrounds may also play a role.

Folate metabolism is also required for cellular methylation reactions (Jacob et al., 1998; Ghandour et al., 2002). Reduced S-adenosylmethionine (SAM) levels, as a result of abnormal folate metabolism or dietary folate deficiency, are associated with DNA hypomethylation (Padmanabhan et al., 2013; Wasson et al., 2006; Waterland et al., 2006; Dobosy et al., 2008). Therefore, folate metabolism may influence DNA stability through epigenetic mechanisms. DNA methylation at cytosine residues is mutagenic due to spontaneous deamination to

thymidine, creating a T-G mismatch (Holliday and Grigg, 1993). DNA methylation patterns can also influence where genetic mutations occur in a broader context. For example in breast cancer cells, differentially methylated loci between cancer and non-cancerous cells were associated with DNA breakpoints (Eric Tang et al., 2012). DNA methylation also plays a vital role in suppressing the activity of potentially deleterious transposable elements. Loss of methylation at transposable elements has been show to lead to genomic instability in cancer (Daskalos et al., 2009). Transposition has huge mutagenic capacity at both the excision and insertion sites (Wicker et al., 2016). Indeed, this phenomenon has been harnessed for mutation generation in the Sleeping Beauty Transposon system (Largaespada, 2009). Overall, folate deficiency or defective folate metabolism may lead to genetic instability, either directly through its role in thymidine synthesis or indirectly through altered DNA methylation patterns.

Epigenetic inheritance remains a controversial topic. Those sceptical of the concept of epigenetic inheritance have long suggested that genetic mutations rather than epimutations are responsible for the inheritance of phenotypes reported. As such, genetic stability has been considered in several models of TEI. The *Agouti viable yellow* ($A^{vy}$) metastable epiallele displays variable and heritable methylation of an IAP element upstream of the *Agouti* gene (Chapter 1, Figure 1.4). Whole genome sequencing of two $A^{vy}$ littermates, one with a yellow coat and one a pseudoagouti coat, was used to demonstrate that genetic differences were unlikely to account for the different coat phenotypes of the sequenced individuals (Oey et al., 2015). Transgenerational inheritance of a wide spectrum of diseases induced by *in utero* exposure of rats to the fungicide vinclozolin is associated with sperm epimutations (Anway et al., 2006; Schuster et al., 2016; Beck et al., 2017; Nilsson et al., 2018). However, vinclozolin is a known mutagen (Hrelia et al., 1996). Chromosomal genomic hybridisation analysis was used to demonstrate that there was an increase in copy number variants (amplification or deletion of repeat elements) in the F3 generation sperm, but not F1 generation sperm, in this model (Skinner et al., 2015). This may suggest the genetic changes are secondary to the epigenetic changes (McCarrey et al., 2016). Additionally, exposing pregnant Big Blue

Rats, which carry a lacI mutation-reporter transgene, to vinclozolin led to a trend towards increased mutation frequency in the F3 generation offspring of vinclozolin exposed rats versus untreated controls (McCarrey et al., 2016). Genetic background is also important in the transgenerational inheritance of adult disease, including male infertility, kidney and prostate disease and immune abnormalities, in vinclozolin exposed mice (Guerrero-Bosagna et al., 2012). Transgenerational inheritance of disease following vinclozolin exposure occurs in CD-1 outbred mice but not 129 inbred mice suggesting a potential genetic influence on the phenotypic inheritance (Guerrero-Bosagna et al., 2012). Furthermore, the length of telomeres, the heterochromatic tandem repeats that protect chromosome ends from degradation (Calado and Dumitriu, 2013), can be responsive to the environment and inherited. Female rats fed a low protein diet have a reduction in telomere length associated with premature reproductive aging in the F2 female offspring (Aiken et al., 2016). Altogether, this data suggests genetic factors may play a role in some epigenetic inheritance models.

Since the folate cycle is necessary for DNA synthesis and repair, there is a possibility that the developmental phenotypes observed in the *Mtrr*$^{gt}$model might result from genetic mutations rather than through epigenetic inheritance. There is some circumstantial evidence that genetic factors are not responsible for the phenotypes observed in the *Mtrr*$^{gt}$model. The phenotypes observed in the F2 generation of the *Mtrr* maternal grandparental pedigrees or in the *Mtrr*$^{gt/gt}$pedigree occur in non-Mendelian ratios (Padmanabhan et al., 2013). This suggests they do not result from inheritance of a single mutation. Furthermore, the severity or frequency of phenotypes, tumour formation frequency or lethality has not increased over time in the *Mtrr*$^{gt/gt}$pedigree as might be expected with the gradual accumulation of mutation burden caused by general genetic instability. Additionally, efforts to minimise genetic background effects were taken. The *Mtrr*$^{gt}$mouse line was generated using 129/P2 embryonic stem cells (ESCs), but the *Mtrr*$^{gt}$allele was maintained by backcrossing it into a C57Bl/6 genetic background for at least 8 generations (Elmore et al., 2007; Padmanabhan et al., 2013). However, an in-depth analysis of genetic stability has not been performed to date. Therefore, in this chapter, I explore genetic stability in the *Mtrr*$^{gt}$ model.

## 4.2 Results

In order to address whether abnormal folate metabolism in the *Mtrr*$^{gt}$ model influences genetic stability I used whole genome sequencing (WGS). This approach was adopted for its base pair resolution, allowing identification of a spectrum of genetic changes that might be expected with genetic instability (e.g., deletions, rearrangements, single nucleotide polymorphisms and telomere length changes) (Le Scouarnec and Gribble, 2012; Pikor et al., 2013; Sims et al., 2014). This approach also allowed potential off-target mutations present in the ESCs used to generated the *Mtrr*$^{gt}$model or mutations present in the C57Bl/6 mice used for backcrossing to be identified.

Six *Mtrr*$^{gt/gt}$embryos at E10.5 with a range of congenital malformations, including heart and neural tube defects, and two phenotypically normal C57Bl/6 control embryos at E10.5 were selected for whole genome sequencing analysis. *Mtrr*$^{gt/gt}$embryos were sequenced as they have the most severe disruption of folate metabolism. Studying embryos with congenital malformations at E10.5 increases the likelihood of detecting significant genetic changes that are embryonic lethal as *Mtrr*$^{gt/gt}$embryos with severe abnormalities do not survive past E14.5 (X. Anderson and E. Watson, unpublished data).

### 4.2.1 The whole genome sequencing data was high quality and aligned well to the reference genome

The quality of whole genome sequencing data was assessed using FastQC to ensure the data was of suitable quality for downstream analysis. Approximately 3.5 x$10^8$ paired-end reads were sequenced for each genome (Table 4.1). All samples had mean phred quality scores of greater than 30 throughout the read length indicating that over 99.9% of bases were likely correctly called (Figure 4.1A). GC content analysis allowed potential biases in the libraries to be identified. My sequencing libraries had an average GC content of 41%, and GC content was approximately normally distributed (Figure 4.1B). This suggested the libraries did not show GC bias. The slightly lower than expected (50%) GC content may result from difficulties mapping short reads to CG repetitive regions. Reads were mapped to the reference genome using BowTie2 and

Table 4.1: **Number of reads obtained from whole genome sequencing of C57Bl/6 and *Mtrr*$^{gt/gt}$ embryos at E10.5.**

| Sample | Genotype | Sex | Phenotype | Total Reads (bp) |
|---|---|---|---|---|
| C57Bl/6-1 | C57Bl/6 | F | Normal | 316,450,854 |
| C57Bl/6-2 | C57Bl/6 | M | Normal | 485,175,952 |
| *Mtrr*$^{gt/gt}$-1 | *Mtrr*$^{gt/gt}$ | M | Abnormal | 305,799,426 |
| *Mtrr*$^{gt/gt}$-2 | *Mtrr*$^{gt/gt}$ | M | Abnormal | 378,194,017 |
| *Mtrr*$^{gt/gt}$-3 | *Mtrr*$^{gt/gt}$ | F | Abnormal | 373,211,179 |
| *Mtrr*$^{gt/gt}$-4 | *Mtrr*$^{gt/gt}$ | F | Abnormal | 306,795,803 |
| *Mtrr*$^{gt/gt}$-5 | *Mtrr*$^{gt/gt}$ | M | Abnormal | 305,364,476 |
| *Mtrr*$^{gt/gt}$-6 | *Mtrr*$^{gt/gt}$ | M | Abnormal | 330,535,970 |

"Abnormal" denotes that the embryo had a congenital malformation such as a heart or neural tube defect. M: male, F: female

alignment rates assessed. Sequence alignment rates were good, with more than 98.5% of reads mapping in all samples (Figure 4.1C-D). Approximately 60% of paired-end reads mapped uniquely to the C57Bl/6 reference genome, with some discordant and multimapping reads (Figure 4.1D). Genome coverage was high across all samples, with at least 40% of the genome having greater than 30x coverage in all but sample *Mtrr*$^{gt/gt}$-4, for which only 20x coverage was achieved (Figure 4.1E). The paired-end reads spanned on average 300bp, indicating good genome coverage (Figure 4.1F). Importantly, across all quality metrics performed, there was consistency between all samples. Overall, the data was of sufficient quality to proceed with variant calling.

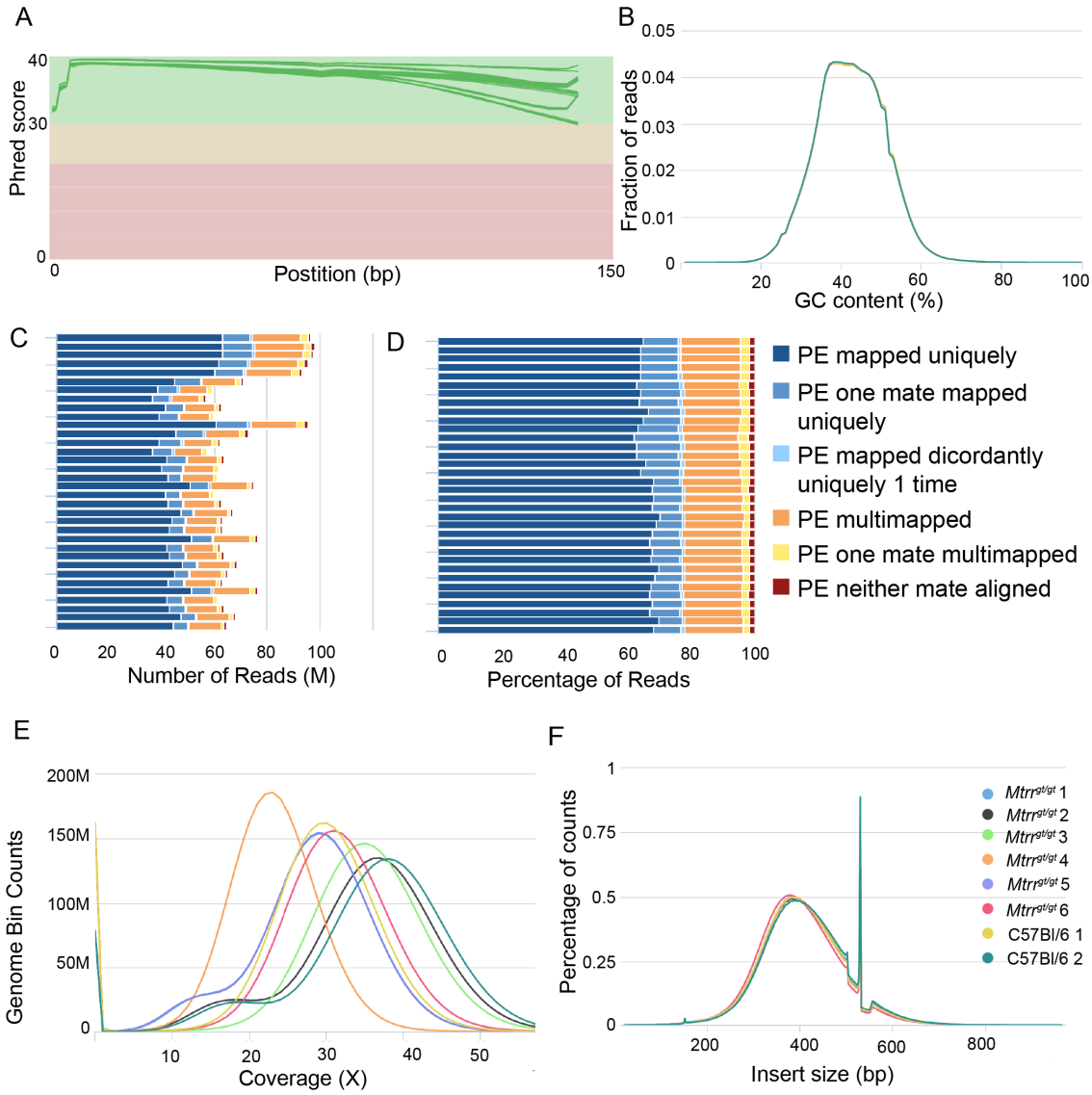



Figure 4.1: **Quality metrics and alignment of next generation whole genome sequencing data. (A)** The Phred quality scores are plotted against read position in base pairs (bp) along the read length. Each green line represents average quality scores for an individual sample. **(B)** The fraction of reads with given GC content (%) is plotted for each sample: *Mtrr*$^{gt/gt}$-1 (pale blue), *Mtrr*$^{gt/gt}$-2 (grey), *Mtrr*$^{gt/gt}$-3 (green), *Mtrr*$^{gt/gt}$-4 (orange), *Mtrr*$^{gt/gt}$-5 (purple), *Mtrr*$^{gt/gt}$-6 (pink), C57Bl/6-1 (yellow) and C57Bl/6-2 (teal). **(C-D)** Graphs plotting **(C)** the number of reads **(D)** the percentage of reads that aligned to the reference genome. Blue: paired-end (PE) reads with at least one mate uniquely aligned, orange: both mates multimapped, yellow: one mate multimapped, red: neither mate aligned. Data is shown for all samples, but each is presented with multiple bars. **(E-F)** Data is plotted for *Mtrr*$^{gt/gt}$-1 (pale blue), *Mtrr*$^{gt/gt}$-2 (grey), *Mtrr*$^{gt/gt}$-3 (green), *Mtrr*$^{gt/gt}$-4 (orange), *Mtrr*$^{gt/gt}$-5 (purple), *Mtrr*$^{gt/gt}$-6 (pink), C57Bl/6-1 (yellow) and C57Bl/6-2 (teal) showing **(E)** genome coverage and **(F)** insert size (the distance between paired-end reads).

### 4.2.2 The frequency of structural variants was not increased in *Mtrr*$^{gt/gt}$ embryos with congenital malformations

To begin my assessment of genetic instability in the *Mtrr*$^{gt}$ model, we identified medium and large scale genetic changes, such as insertions and deletions (indels) and structural variants (greater than 1 kb insertions, deletions, inversions and translocations (Freeman et al., 2006)), not present in the C57Bl/6 reference genome using Manta (Chen et al., 2016b) (Figure 4.2). All variants identified by Manta will be referred to as structural variants (SVs) for simplicity. SVs called by Manta are classified as either 'diploid' or 'candidate'. The diploid SVs identified have been scored and a probability assigned to each SV. Candidate SVs represent unscored SVs, with only a minimal amount of evidence required for a SV to be considered a candidate. Candidate SVs were much more numerous than diploid SVs on all chromosomes and in all samples (Figure 4.2).

Firstly, I considered the frequency of SVs on each chromosome (Figure 4.2). The number of SVs, both diploid and candidate, was similar between C57Bl/6 controls and *Mtrr*$^{gt/gt}$ embryos on all chromosomes, except chromosome 13 (Figure 4.2). The number of SVs identified on chromosome 13 was far greater in *Mtrr*$^{gt/gt}$ embryos than in C57Bl/6 embryos ($p<0.0001$, unpaired t-test with Welch's correction). Additionally, in *Mtrr*$^{gt/gt}$ embryos, there was a far greater number of SVs on chromosome 13 compared to the number of SVs identified on other chromosomes ($p<0.0001$, two-way ANOVA with Sidak's multiple comparisons test). Visualisation of the SV data on the genome browser highlighted that the structural variants on chromosome 13 in *Mtrr*$^{gt/gt}$ embryos cluster around the *Mtrr* locus (*Mtrr*, chr13:68560780-68582121). This was not unexpected given the large gene-trap insertion into intron 9 of the *Mtrr* gene in the *Mtrr*$^{gt}$ model. Failure to align reads correctly in this region could lead to the identification of SVs.


Figure 4.2: **SVs identified using Manta in C57Bl/6 and *Mtrr*$^{gt/gt}$embryos.** SVs identified using Manta are plotted per chromosome for C57Bl/6 (blue shades) and *Mtrr*$^{gt/gt}$embryos (green shades). SV class count was plotted for each structural variant class: INV: inversion, INS: insertion, DUP: duplication, DEL: deletion, BND: Break points. Diploid (scored) (black font) and candidate (unscored) (red font) SVs are shown.

When the region surrounding the *Mtrr* locus was masked (masked region chr13:60424810-70750000), the number of SVs, both diploid and candidate, on chromosome 13 in *Mtrr*$^{gt/gt}$embryos became equivalent to C57Bl/6 controls

(Figure 4.3). Rescaling of the data following the masking of the *Mtrr* locus made apparent that the frequency of SVs was elevated on chromosomes 1 and 4 compared to the all other chromosome in both C57Bl/6 and $Mtrr^{gt/gt}$embryos compared to the other autosomes ($p<0.05$ - $p<0.0001$, two-way ANOVA with Sidak's multiple comparisons test, Figure 4.3). The frequency of SVs on the X and Y chromosomes was also elevated in comparison to some but not all autosomes ($p<0.0001$ - $p>0.9999$, two-way ANOVA with Sidak's multiple comparisons test, Figure 4.3). An explanation for the increased SV frequency on these chromosomes is unknown and may warrant further investigations.


Figure 4.3: **SVs identified using Manta in C57Bl/6 and *Mtrr*$^{gt/gt}$embryos, with the *Mtrr* locus masked.** SVs identified using Manta are plotted per chromosome for C57Bl/6 (blue shades) and *Mtrr*$^{gt/gt}$embryos (green shades). The *Mtrr* locus on chromosome 13 has been masked, such that SVs present at that locus do no appear in the counts. SV class count was plotted for each structural variant class: INV: inversion, INS: insertion, DUP: duplication, DEL: deletion, BND: Break points. Diploid (scored) (black font) and candidate (unscored) (red font) SVs are shown.

In order to minimise the impact of false positive SV calls, further analysis of SVs was performed on a refined SV dataset in which only diploid SVs were

considered. The *Mtrr* locus was masked in this dataset. Variants at the *Mtrr* locus are examined in more detail in Section 4.2.5. Considering only this refined SV data set, the total number of SVs was not increased in $Mtrr^{gt/gt}$embryos versus C57Bl/6 controls (p=0.6886, unpaired t-test with Welch's correction, Figure 4.4A). I simplified the SV categorisation into structural variants and indels. There was no difference in the mean number of variants in either category between C57Bl/6 control and $Mtrr^{gt/gt}$embryos ($p>0.75$, two-way ANOVA with Sidak's multiple comparisons test, Figure 4.4B).

To explore the potential deleteriousness of SVs identified, I determined their genomic location using SnpEff (Cingolani et al., 2012). The majority of SVs were located in intronic regions (approximately 40%), followed by intergenic ($\sim$ 30%), downstream ($\sim$ 10%) and upstream ($\sim$ 10%) regions (Figure 4.4C). Less than 5% of SVs were located in exons or transcripts (Figure 4.4C). SnpEff was also used to estimate the putative impact/deleteriousness of variants. Over 95% of SVs identified in both C57Bl/6 and $Mtrr^{gt/gt}$embryos had an impact classified as 'modifier'. 'Modifiers' represent non-coding variants or variants affecting non-coding genes or where there is no evidence of impact (Figure 4.4D). A small percentage ($< 5\%$) of high impact variants were detected (Figure 4.4D). High impact variants are likely to severely affect protein function causing truncation, loss of function or triggering nonsense-mediated decay. However, this tool carries the caveat that the classification of variants cannot accurately predict the impact of the variant in terms of phenotype. Importantly, the genomic locations (e.g. intronic, intergenic) and impacts estimated by SnpEff of the SVs identified in C57Bl/6 and $Mtrr^{gt/gt}$embryos were equivalent. Overall, the majority of SVs in $Mtrr^{gt/gt}$embryos were located in non-coding regions and were predicted to be unlikely to have deleterious consequences.

Despite equivalent frequency and impact of SVs in $Mtrr^{gt/gt}$and C57Bl/6 control embryos, it is possible that SVs may still contribute to the inheritance of phenotypes in the $Mtrr^{gt}$model. A single or group of SVs present in $Mtrr^{gt/gt}$mice (absent from C57Bl/6 controls) could result in the embryonic phenotypes observed. In order to explore this, we performed an intersectional analysis comparing the SVs identified in each individual to the SVs identified in all other individuals. Strikingly, the overall level of intersection was extremely

low. In both C57Bl/6 and *Mtrr*$^{gt/gt}$embryos, the majority of SVs identified were unique to that individual (Figure 4.4E). This suggests these SVs were likely generated *de novo* and were not inherited. A set of 22 SVs were common across all individuals (Figure 4.4E). These likely represent a diversion of our C57Bl/6 colony, maintained in Cambridge, UK, and the reference C57Bl/6 strain. While, backcrossing of the *Mtrr*$^{gt}$line was performed into a C57Bl/6 colony at the University of Calgary, Canada, some backcrossing into the C57Bl/6 UK colony used as controls has since occurred. Only four SVs were common to all *Mtrr*$^{gt/gt}$embryos, but not present in C57Bl/6 embryos (Figure 4.4E., Table 4.2). These SVs are likely mutations established within the *Mtrr*$^{gt}$mouse line. Mutations in the genes closest to or within these SVs have not been associated with any of the phenotypes observed in the *Mtrr*$^{gt}$model. However, *Mtrr*$^{gt/gt}$-specific SVs should be explored further to exclude any contribution to the phenotypes observed in the *Mtrr*$^{gt}$model (Section 4.2.7).

Table 4.2: **Structural variants common to all *Mtrr*$^{gt/gt}$embryos.**

| Chr | Location | Type | Context | Closest Gene |
|---|---|---|---|---|
| 1 | 80828728 | Deletion | Intergenic | *Dock10* |
| 11 | 32122791 | Deletion | Intergenic | *Nsg2,Il9r* |
| 15 | 83975384 | Deletion | Intron | *Efcab6* |
| 19 | 36911361 | Duplication | - | *Fgfbp3, Btaf1, Cpeb3, March5, Ide, Kif11* |

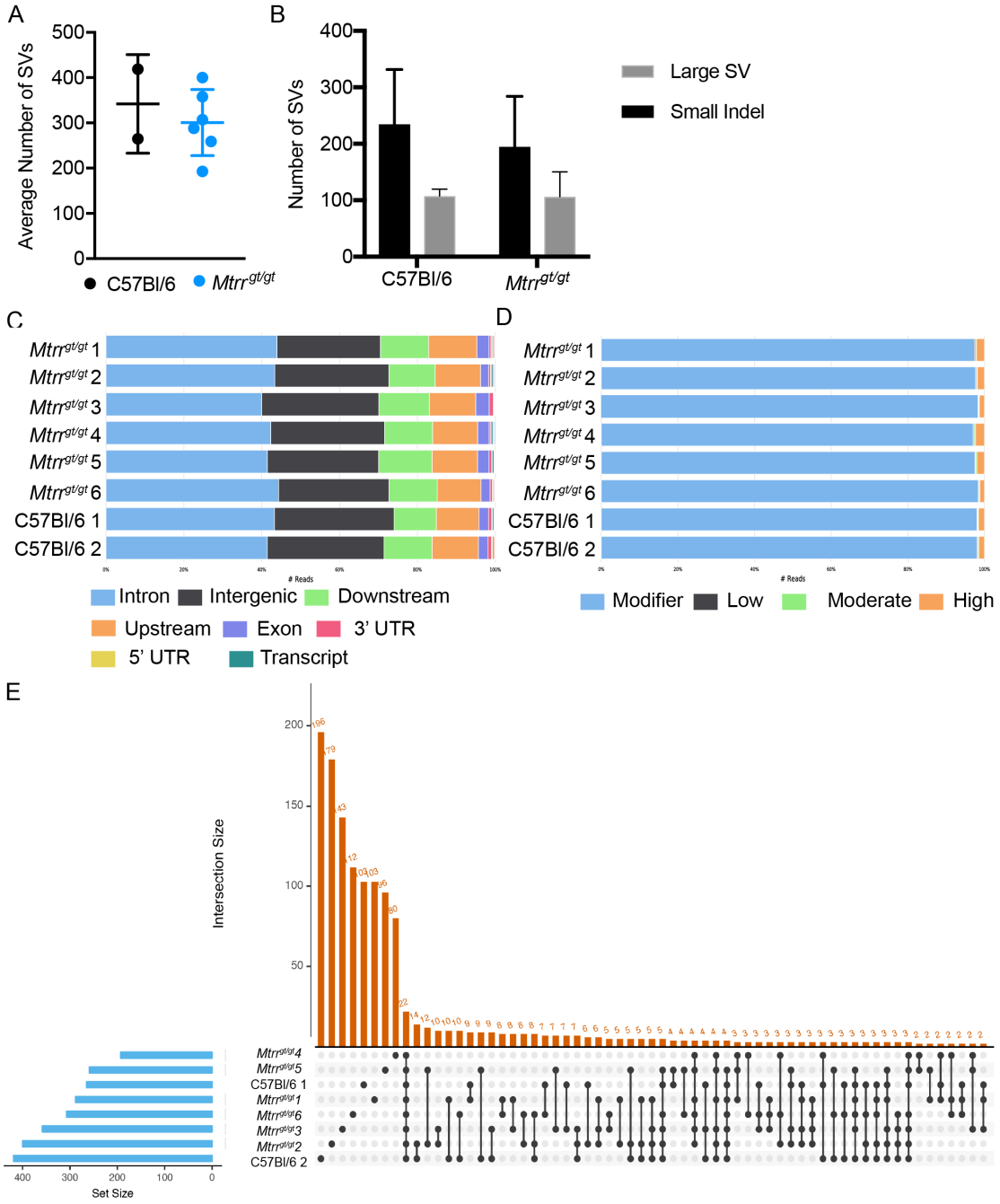

A
Average Number of SVs
500
400
300
200
100
0
C57Bl/6
Mtrr gt/gt
B
Number of SVs
400
300
200
100
0
C57Bl/6
Mtrr gt/gt
Large SV
Small Indel
C
Mtrr gt/gt 1
Mtrr gt/gt 2
Mtrr gt/gt 3
Mtrr gt/gt 4
Mtrr gt/gt 5
Mtrr gt/gt 6
C57Bl/6 1
C57Bl/6 2
# Reads
Intron
Intergenic
Downstream
Upstream
Exon
3' UTR
5' UTR
Transcript
D
Mtrr gt/gt 1
Mtrr gt/gt 2
Mtrr gt/gt 3
Mtrr gt/gt 4
Mtrr gt/gt 5
Mtrr gt/gt 6
C57Bl/6 1
C57Bl/6 2
# Reads
Modifier
Low
Moderate
High
E
Intersection Size
200
150
100
50
Mtrr gt/gt 4
Mtrr gt/gt 5
C57Bl/6 1
Mtrr gt/gt 1
Mtrr gt/gt 6
Mtrr gt/gt 3
Mtrr gt/gt 2
C57Bl/6 2
400
300
200
100
0
Set Size

Figure 4.4: **The frequency and impact of structural variants was not increased in *Mtrr*$^{gt/gt}$ embryos**. **(A)** The total number of SVs is shown for C57Bl/6 (black) and *Mtrr*$^{gt/gt}$(blue) embryos. Data is shown as mean ± standard deviation (sd). **(B)** The mean number of SVs classed as either large SVs (grey) or small indels (black) is plotted for C57Bl/6 and *Mtrr*$^{gt/gt}$embryos. Data is shown as mean ± sd. **(C)** The percentage of SVs by genomic region: intron (blue), intergenic (grey), downstream (green), upstream (orange), exon (purple), 3' UTR (pink), 5'UTR (yellow) and transcripts (teal), is plotted for all C57Bl/6 and *Mtrr*$^{gt/gt}$embryos. **(D)** The percentage of SVs classed as either modifier (blue), low impact (grey), moderate impact (green) or high impact (orange) as determined using SnpEff for all C57Bl/6 and *Mtrr*$^{gt/gt}$embryos. **(F)** An UpSet plot is used to show intersection between the SVs found in individuals samples. Total SV counts per sample are plotted horizontally (blue bars) and intersections are plotted vertically (orange bars).

### 4.2.3 The frequency of single nucleotide polymorphisms was not increased in *Mtrr*$^{gt/gt}$ embryos with congenital malformations

Next, I identified single nucleotide polymorphisms (SNPs) and small indels ($<$ 150bp) using GATK. Variants identified using GATK are here after referred to as SNPs for simplicity. Variant calling using GATK was performed using combined C57Bl/6 and *Mtrr*$^{gt/gt}$sequencing data, rather than performing variant calling for each sample individually. This gave greater power for accurate variant calling as information about alleles and their frequencies could be compared across samples. In total, 887,244 SNPs were initially identified across all embryos. Extensive filtering was then performed. Removal of biased and low quality SNPs reduced the total by over 97% to 29,700 SNPs across all embryos. A second filtering step was then performed primarily to remove SNPs found at highly repetitive regions (Oey et al., 2015). This reduced the total number of SNPs to 9494 SNPs. The number of SNPs identified per sample after filtering was on average approximately 5000 SNPs. Overall, there was no significant difference in the total number of SNPs identified between *Mtrr*$^{gt/gt}$and C57Bl/6 control embryos (p=0.7181, unpaired t-test with Welch's correction, Figure 4.5A).

I next considered the transition/transversion (Ts/Tv) ratio and the homozygous/heterozygous (Hom/Het) ratio. These provide a quality control mea-

surement to assess the validity of variant calling. The Ts/Tv ratio is a ratio of the number of transition mutations (purine to purine or pyrimidine to pyrimidine) over the number of transversions (purine to pyrimidine and vice-versa). In a biological context, the expected Ts/Tv ratio is approximately 2 (Wang et al., 2015). The mean Ts/Tv ratio in my data was 0.96. Furthermore, the Ts/Tv ratio was higher in *Mtrr*$^{gt/gt}$embryos versus the C57Bl/6 control embryos (C57Bl/6 Ts/Tv: 0.86 $\pm$ 0.01, *Mtrr*$^{gt/gt}$Ts/Tv: 0.99 $\pm$ 0.01, p=0.0.162, unpaired t-test with Welch's correction, Figure 4.5B). A low Ts/Tv ratio may indicate false positive SNP calling. The Hom/Het ratio, the ratio of homozygous to heterozygous mutations, is also expected to be approximately 2 (Wang et al., 2015). The mean Hom/Het ratio in my data was 1.07. There was no significant difference in the Hom/Het ratio between *Mtrr*$^{gt/gt}$and C57Bl/6 embryos (C57Bl/6 Hom/Het: 0.83 $\pm$ 0.14, *Mtrr*$^{gt/gt}$Hom/Het: 1.15 $\pm$ 0.05, p=0.2253, unpaired t-test with Welch's correction, Figure 4.5C). As the Ts/Tv and Hom/Het ratios were below the expected values, I questioned the validity of my variant calling. However, the Ts/Tv ratio and Hom/Het ratio vary significantly by genome region, gene functionality and ancestry and may be limited in their scope as a quality control indicator in a SNP variant calling context (Wang et al., 2015). Therefore, I proceeded cautiously with further SNP analysis.

I examined the type of variants identified using GATK, classifying them into three categories: SNPs, small insertions and small deletions. There was no significant difference in the mean number of variants in each of the categories between *Mtrr*$^{gt/gt}$embryos and C57Bl/6 controls ($p>0.0567$, two-way ANOVA, Figure 4.5 D). To explore the potential impact of the SNPs identified, I assessed their chromosomal and genomic locations. SNPs were found distributed across all chromosomes, although there was a significant increase in the number of SNPs on chromosomes 13 and X ($p<0.0001$, Figure 4.5 E,F). The frequency of SNPs identified on chromosome 13 in *Mtrr*$^{gt/gt}$embryos was significantly higher than in C57Bl/6 controls (*Mtrr*$^{gt/gt}$: 783.5 SNPs, C57Bl/6: 252.5 SNPs, $p<0.0001$, unpaired t-test with Welch's correction). The increase in SNP frequency on chromosome 13 was likely associated with the presence of the *Mtrr* mutation on chromosome 13. There was no significant difference in the number of SNPs identified on the X chromosome between *Mtrr*$^{gt/gt}$and C57Bl/6 embryos. The

reason for the increased number of SNPs on the X chromosome is unknown and should be explored further. Most SNPs were located in intronic (C57Bl/6: 49.5%, *Mtrr*$^{gt/gt}$: 47.8%), intergenic (C57Bl/6: 29.4%, *Mtrr*$^{gt/gt}$: 32.0%), upstream (C57Bl/6: 10.3%, *Mtrr*$^{gt/gt}$: 10.4%) or downstream (C57Bl/6: 8.2%, *Mtrr*$^{gt/gt}$: 8.9%) regions (Figure 4.5 G,H). Less than 1% of SNPs were in exons (C57Bl/6: 0.52%, *Mtrr*$^{gt/gt}$: 0.46%), 3'UTRs (C57Bl/6: 0.24%, *Mtrr*$^{gt/gt}$: 0.24%), 5'UTRs (C57Bl/6: 0.18%, *Mtrr*$^{gt/gt}$: 0.18%) and transcripts (C57Bl/6: 0.03%, *Mtrr*$^{gt/gt}$: 0.03%) (Figure 4.5 G,H). The low percentage of SNPs arising in exons may have contributed to the low Ts/Tv ratio observed (Figure 4.5 B), as exonic SNPs are associated with higher Ts/Tv ratios (Wang et al., 2015). Together, this data suggested the majority of SNPs identified were unlikely to have highly deleterious impacts.

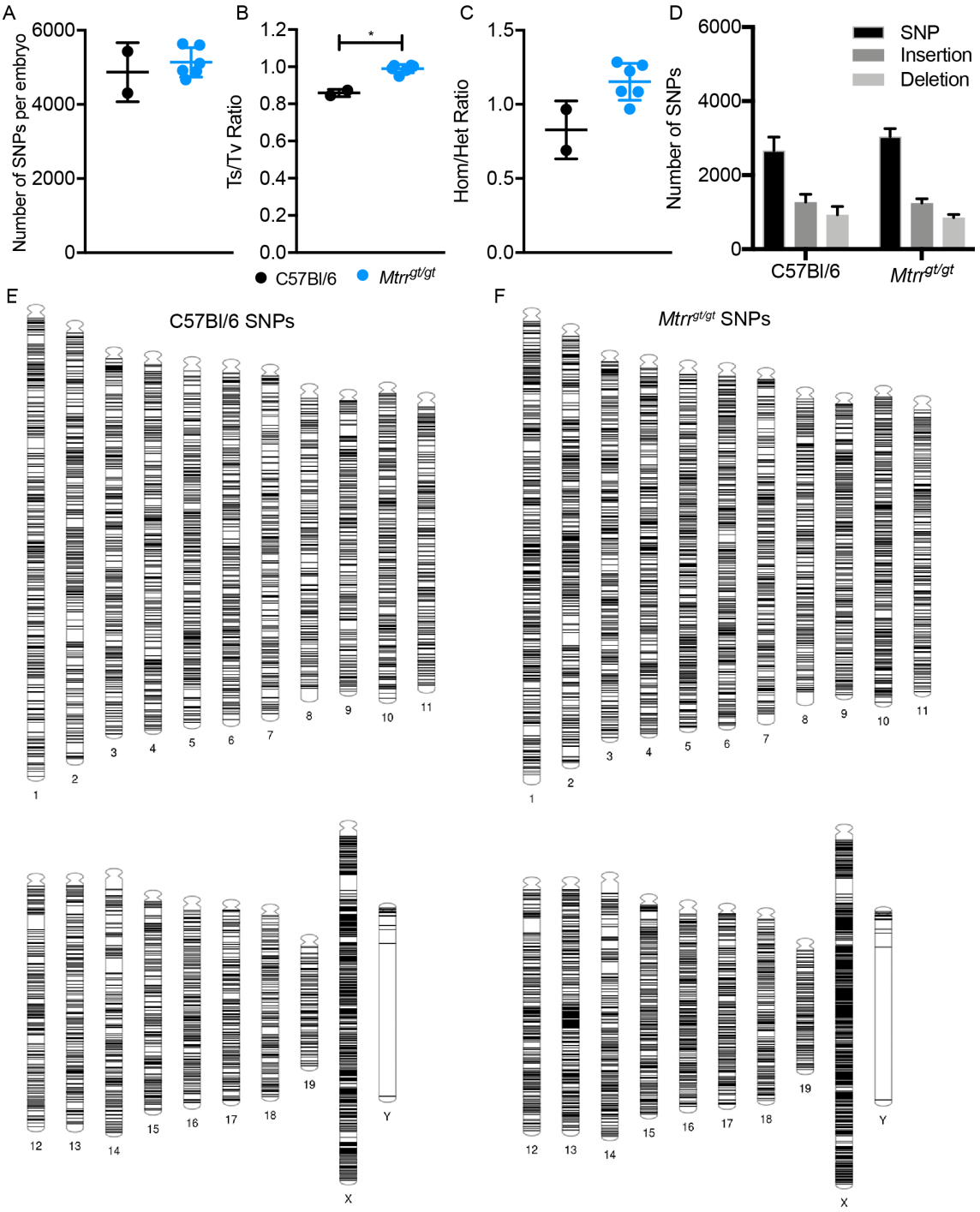

A
Number of SNPs per embryo
6000
4000
2000
0
B
Ts/Tv Ratio
1.2
1.0
0.8
0.6
0.4
0.2
0.0
*
C57Bl/6
Mtrr gt/gt
C
Hom/Het Ratio
1.5
1.0
0.5
0.0
D
Number of SNPs
6000
4000
2000
0
SNP
Insertion
Deletion
C57Bl/6
Mtrr gt/gt
E
C57Bl/6 SNPs
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
X
Y
F
Mtrr gt/gt SNPs
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
X
Y
G
C57Bl/6 SNPs
H
Mtrr gt/gt SNPs
Intergenic
Upstream
Downstream
Exon
3' UTR
5' UTR
Splice Site Region
Transcript
Intron

Figure 4.5: **Single nucleotide polymorphisms and small indels identified in C57Bl/6 and *Mtrr*$^{gt/gt}$embryos (A-C)** Data is plotted for C57Bl/6 embryos (black) and *Mtrr*$^{gt/gt}$embryos (blue). Data is presented as mean ± sd. **(A)** The total number of SNPs identified in C57Bl/6 and *Mtrr*$^{gt/gt}$embryos using GATK. **(B)** The transition/transversion (Ts/Tv) ratio is plotted for C57Bl/6 and *Mtrr*$^{gt/gt}$embryos. **(C)** The ratio of homozygous to heterozygous mutations (Hom/Het) ratio is plotted. **(D)** Graphical data showing the average number of SNPs (black), small insertions (dark grey) and small deletions (light grey) for C57Bl/6 and *Mtrr*$^{gt/gt}$embryos. Data is plotted as mean ± sd. **(E-F)** Phenograms depict the chromosomal location of **(E)** all SNPs identified in C57Bl/6 embryos and **(F)** all SNPs identified in *Mtrr*$^{gt/gt}$embryos. **(G-H)** The type of variant: non-coding exon transcript variant (blue), downstream gene variants (pink), intron variant (green), 3'UTR variant (yellow), transcript variant (dark green), intergenic variant (red), upstream gene variant (black), 5'UTR variant (pale blue) and splice site variant (orange) is displayed for **(G)** all C57Bl/6 SNPs and **(H)** all *Mtrr*$^{gt/gt}$SNPs. (A-C) Unpaired t-tests with Welch's correction were performed, * $p < 0.05$.

To explore similarities and differences in the SNP populations identified in individual samples, I firstly performed a principle component analysis (PCA). This highlighted that the C57Bl/6 and *Mtrr*$^{gt/gt}$embryos cluster separately and are most different along principle component (PC) 1 (Figure 4.6A). Both the *Mtrr*$^{gt/gt}$samples and C57Bl/6 samples were split further into two separate clusters along PC2 (Figure 4.6A). The clustering is seen clearly on the cluster dendrogram (Figure 4.6B). The division along PC2 reflects the sex of the embryos, with males in the top cluster and females in the bottom cluster. I speculate that this may result from SNPs on the allosomes. The division of *Mtrr*$^{gt/gt}$and C57Bl/6 samples along PC2 was not associated with underlying phenotypic differences or littermate effects. This data suggested that there may be distinct SNP populations in C57Bl/6 and *Mtrr*$^{gt/gt}$embryos, but with subgroups within them.

Therefore, I performed intersectional analysis on the SNPs identified across all individuals. This revealed that the vast majority of SNPs identified were common across both C57Bl/6 and *Mtrr*$^{gt/gt}$embryos (Figure 4.6C). Very few SNPs were unique to each individual (Figure 4.6C). The *Mtrr*$^{gt}$line was primarily backcrossed to a C57Bl/6 colony at the University of Calgary, Canada. The *Mtrr*$^{gt}$colony was moved to the UK and a new C57Bl/6 control colony

was bought in. Some backcrossing into the C57Bl/6 UK colony has since occurred. This may account for some of the SNPs common between C57Bl/6 and *Mtrr*$^{gt/gt}$embryos. Only 21 SNPs were unique and common to the *Mtrr*$^{gt/gt}$embryos, i.e. present in all *Mtrr*$^{gt/gt}$embryos but absent in C57Bl/6 embryos (Figure 4.6C, Table 4.3). These SNPs were not located within any genes associated with the phenotypes observed in the *Mtrr*$^{gt}$model (Table 4.3). A SNP on the X chromosome is closest to the *Hccs* gene. The phenotypes of *Hccs*$^{-/-}$ mice are similar to those observed in *Mtrr*$^{gt}$mice (Padmanabhan et al., 2013; Prakash et al., 2002; Drenckhahn et al., 2008), however the SNP is over 50kb upstream of *Hccs*, therefore may be unlikely to influence *Hccs* expression. Less than two SNPs were unique and common to the C57Bl/6 embryos. This value was low and may reflect the extensive filtering performed on my SNP dataset.

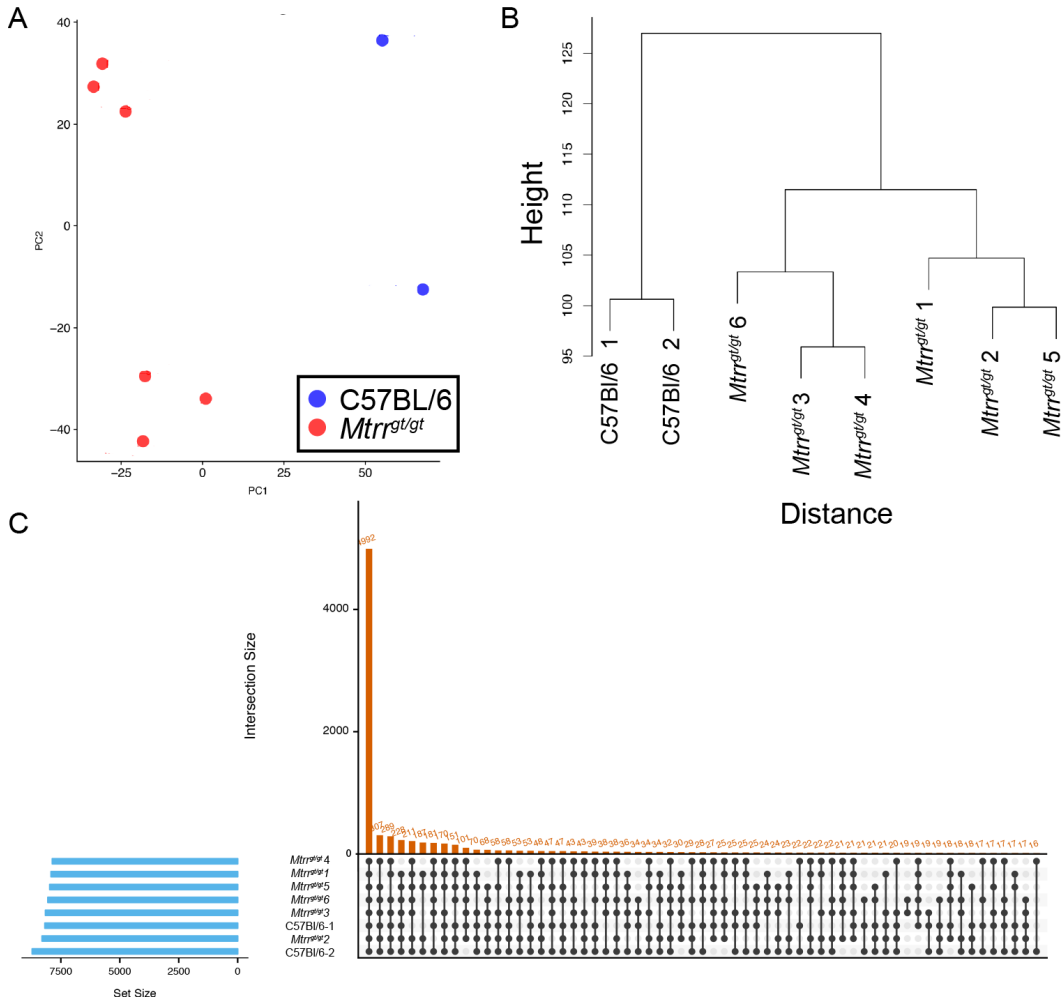



Figure 4.6: **PCA-clustering and intersectional analysis of SNPs in *Mtrr*$^{gt/gt}$ and C57Bl/6 embryos. (A)** Principle component analysis performed using data for SNPs identified in C57Bl/6 (dark blue) and *Mtrr*$^{gt/gt}$ (red) embryos. **(B)** A cluster dendrogram shows hierarchical clustering of C57Bl/6 and *Mtrr*$^{gt/gt}$ embryos based on SNP data. **(C)** An UpSet plot is used to show intersections between SNPs found in individuals samples. Total SNP counts per sample are plotted horizontally (blue bars) and intersections are plotted vertically (orange bars).

Table 4.3: **SNPs common to all *Mtrr*$^{gt/gt}$ embryos.**

| Chr | Location | Reference Allele | *Mtrr*$^{gt/gt}$ Allele | Context | Closest Gene | KO Phenotype | KO Phenotype Reference |
|---|---|---|---|---|---|---|---|
| 1 | 23478171 | TC | T | Intergenic | *Ogfrl1* | - | - |
| 1 | 130763137 | A | C | Intergenic | *AA986860* | - | - |
| 13 | 58081242 | C | G | Intron | *Klhl3* | Abnormal kidney morphology, increased chlorine and potassium in blood | Sasaki et al. (2017) |
| 13 | 81473683 | T | C | Intron | *Gpr98* (*Adgrv1*) | Abnormal auditory hair bundle development, deafness | McGee et al. (2006) |
| 13 | 81473686 | T | C | Intron | *Gpr98* | As above | As above |
| 13 | 81473693 | C | T | Intron | *Gpr98* | As above | As above |
| 13 | 81473695 | C | G | Intron | *Gpr98* | As above | As above |
| 13 | 81473744 | C | T | Intron | *Gpr98* | As above | As above |
| 13 | 81473746 | C | G | Intron | *Gpr98* | As above | As above |
| 13 | 81473756 | C | T | Intron | *Gpr98* | As above | As above |
| 14 | 119813417 | ACCCC | ACCCCC | Intron | *Hs6st3* | - | - |
| 16 | 85907608 | GC | C | Intergenic | *Adamts5* | Decreased osetoarthritis susceptibility | Glasson et al. (2005) |
| 16 | 89162602 | G | A | Intergenic | *Krtap20-2* | - | - |
| 18 | 87587984 | T | C | Intergenic | *Cbln2* | - | - |

| Chr | Location | Reference Allele | *Mtrr*$^{gt/gt}$ Allele | Context | Closest Gene | KO Phenotype | KO Phenotype Reference |
|---|---|---|---|---|---|---|---|
| 6 | 114014496 | T | TCTCCC-CTCCC | Intron | *Atp2b2* | Gait, balance, inner ear, cerebellum and cochlear abnormalities | Kozel et al. (1998) |
| 6 | 121526732 | G | T | Intergenic | *Iqsec3* | Bone structure abnormalities and increased body fat | Smith et al. (2018) |
| 8 | 98943351 | A | G | Intergenic | *Gm15679* | - | - |
| 8 | 104303432 | C | G | Intron | *Cmtm1* | - | - |
| X | 68533748 | TC | C | Intergenic | *Gm10474* (*Fmr1os*) | - | - |
| X | 71514412 | C | $C(A)_{45}$ | Intergenic | *Cd99l2* | Abnormal leukocyte extravasation | Seelige et al. (2013) |
| X | 169366447 | TC | T | Intergenic | *Hccs* | Embryonic lethality, embryonic growth retardation, abnormal gastrulation, abnormal cardiac physiology | Prakash et al. (2002); Drenckhahn et al. (2008) |

- denotes that no knock-out (KO) exists or no phenotypes have been reported.

### 4.2.4 Attempts to validate the whole genome sequencing data suggest a high frequency of false-positive calls

In order to examine the reliability of my data and ascertain a false discovery rate for variant calls, I attempted to validate a subset of variants by Sanger sequencing. Validations were limited to SNPs and small indels. In order to verify that the filtering process (see methods, Chapter 2) performed was not excluding true positive variants, I undertook validations of some variants removed by filtering (n=14), in addition to variants present in the filtered SNP data (n=4). I struggled to design primers that gave a unique PCR product, which greatly limited the scope of my validations. This largely reflected the location of many of the SNPs in repetitive regions of the genome. Sanger sequencing validations were performed on DNA from a subset of the embryos used for whole genome sequencing analysis: *Mtrr*$^{gt/gt}$-4, *Mtrr*$^{gt/gt}$-5, *Mtrr*$^{gt/gt}$-6, C57Bl/6-1 and C57Bl/6-2. Only one of eighteen variants tested validated successfully (Table 4.4). The variant that validated was T to C conversion on Chr7: 30544362, identified by whole genome sequencing in samples *Mtrr*$^{gt/gt}$-4, *Mtrr*$^{gt/gt}$-5 and C57Bl/6 -2. However, this validation was only partial as a small insertion at this locus in sample *Mtrr*$^{gt/gt}$-6 failed to validate (Table 4.4). All other SNPs or small indels failed to validate (Table 4.4). This data suggested there were numerous false positive variant calls within my data. Given the small number of variants that underwent validation, it was not possible to calculate a meaningful false positive rate. The high level of false positives was unexpected given the filtering steps to remove low quality variants and the level of coverage achieved, but may have been indicated by the low Ts/Tv ratio. The majority of SNPs that I attempted to validate (13 of 18) were within or on the margin of repetitive regions. This increases the chance of sequencing errors, both during Sanger sequencing and whole genome sequencing (for both my data and the reference genome), due to enzyme slippage. Additionally, the variant that validated was removed in the filtering process. This suggests that the filtering process may be excluding true-positive variants. No additional SNPs were identified in the PCR products that underwent Sanger sequencing for SNP validation. This suggested a low false negative rate.

Table 4.4: **Variants that underwent validation by Sanger sequencing in C57Bl/6 and *Mtrr*$^{gt/gt}$ E10.5 embryos.**

| SNP | Chr | Location | Reference | C57Bl/6-1 | C57Bl/6-2 | *Mtrr*$^{gt/gt}$-4 | *Mtrr*$^{gt/gt}$-5 | *Mtrr*$^{gt/gt}$-6 | Validated? |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 7 | 18963777 | TTC | TTC | TTC | CTC | T/TTC | T/TTC | Not Validated |
| 2 | 1 | 81498220 | TAA | TAA | TAA | T/TAA | T/TAA | TAA | Not Validated |
| 3 | 2 | 36359099 | C | C | C | T/C | C | C | Not Validated |
| 4 | 2 | 36359103 | A | A | A | G/A | A | A | Not Validated |
| 5 | 2 | 36359114 | C | C | C | T/C | C | C | Not Validated |
| 6 | 2 | 36359133 | C | C | C | T/C | C | C | Not Validated |
| 7 | 6 | 13698031 | C | G/C | C | C | C | C | Not Validated |
| 8 | 7 | 30544362 | TAC | TAC | CAC/TAC | CAC/TAC | CAC/TAC | $TAC(AC)_7$/TAC | Validated |
| 9 | 7 | 30553705 | T | T | C/T | C/T | T | C/T | Not Validated |
| 10 | 7 | 30553722 | G | G | C/G | G | G | G | Not Validated |
| 11 | 7 | 92402649 | C | C | $C(T)_{21}$ | $C/C(T)_{21}$ | C | $C(T)_6$ | Not Validated |
| 12 | 7 | 92402699 | TAG | $TAG(T)_{14}$/T | TAG | TAG | TAT | TAG | Not Validated |
| 13 | 7 | 92402704 | TAA-GAGCTCT-CACTGC | Ref | C | Ref | T/Ref | Ref | Not Validated |
| 14 | 7 | 92402721 | G | G | G | G | T/G | G | Not Validated |
| 15 | 10 | 100097326 | T | T | C/T | C/C | C/T | C/T | Not Validated |
| 16 | 5 | 30156701 | C | C | CT/C | CT/CT | CT/C | C | Not Validated |
| 17 | 7 | 104586593 | G | C/G | - | C/G | C/G | C/G | Not Validated |
| 18 | 10 | 82833148 | A | G/A | G/A | G/A | G/G | G/A | Not Validated |

A single allele is shown in the case of homozygosity, both alleles are shown in the case of heterozygosity. Ref: reference allele for the locus.

### 4.2.5 The *Mtrr* locus has a distinct genetic background to the rest of the genome in *Mtrr*$^{gt/gt}$mice

The numbers of SVs and SNPs were increased on chromosome 13 with respect to the other autosomes (Figure 4.2, Figure 4.5 F). This correlated with the presence of the *Mtrr* gene on chromosome 13 (chr13: 68560780-68582121). The embryonic stem cells (ESCs) containing the gene-trap insertion used to generate the *Mtrr*$^{gt}$mouse line were of a 129/P2 genetic background. However, upon germline transmission, the *Mtrr*$^{gt}$allele was backcrossed into a C57Bl/6 genetic background for at least 8 generations (Elmore et al., 2007; Padmanabhan et al., 2013). Therefore while the majority of the genome is C57Bl/6 genetic background, the *Mtrr* gene and immediately flanking regions are likely to be 129/P2 derived.

I explored the extent of the 129/P2 genetic background near the *Mtrr* gene further. The increased frequency of SNPs in *Mtrr*$^{gt/gt}$embryos appeared to be distinctly asymmetric about the *Mtrr* locus (Figure 4.7A). I divided the genomic region depicted in Figure 4.7 (chr13:39,060,780-99,060,780) into 1 Mb windows. I used the Sanger Mouse Genomes Project (Yalcin et al., 2011; Keane et al., 2011) to obtain details of all SNPs that are present between the 129P2_OlaHsd (129/P2) genome and C57Bl/6 reference genome (mm10) in each 1 Mb window (referred to as "A") (Figure 4.7 B). I also calculated the total number of SNPs that I had identified in C57Bl/6 and *Mtrr*$^{gt/gt}$embryos using whole genome sequencing within each 1 Mb window (referred to as "B") (Figure 4.7 B). I then intersected these two data sets (A and B) for each window. This gave us the number of 129/P2 SNPs identified in each window in C57Bl/6 and *Mtrr*$^{gt/gt}$embryos (referred to as "C") (Figure 4.7 B). I calculated this as a percentage of the total number of SNPs for the region as a whole (e.g. all 1 Mb windows considered near the *Mtrr* gene). I then plotted this percentage against the genomic location of each 1 Mb window (Figure 4.7 C). A clearly defined region of approximately 20 Mb around the *Mtrr* gene has increased 129/P2 SNP frequency compared to the surrounding genome. I defined this region as the *Mtrr locus* (Chr13: 59,060,780-79,060,780). This suggests an approximately 20 Mb region around the *Mtrr* gene is of 129/P2 genetic background.

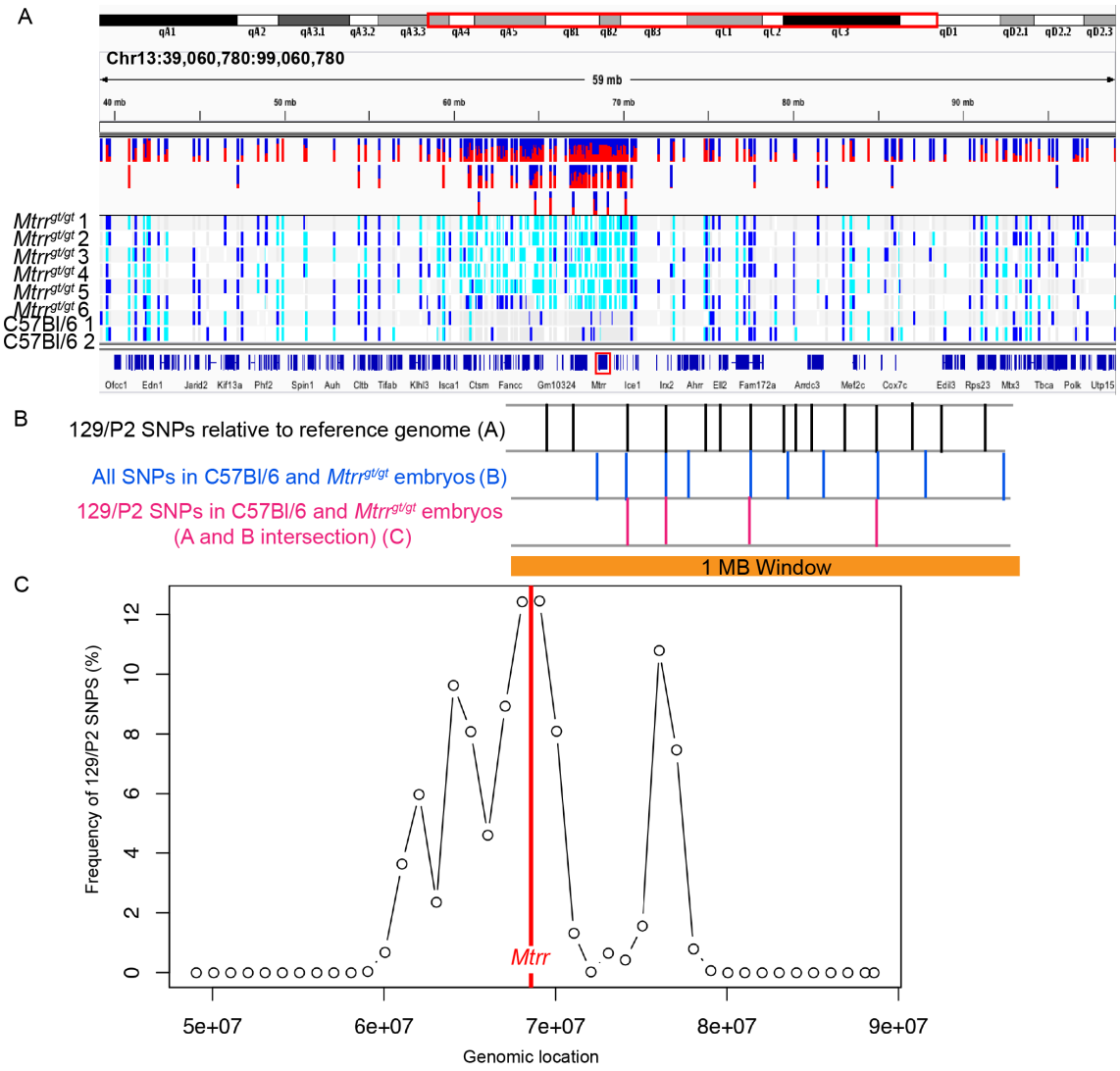



Figure 4.7: **The *Mtrr* locus is of 129/P2 genetic background in *Mtrr*$^{gt/gt}$mice.** **(A)** An IGV snapshot shows the SNPs present in a region of chromosome 13 (39,060,780-99,060,780) surrounding the the *Mtrr* gene (red box) in *Mtrr*$^{gt/gt}$and C57Bl/6 embryos. **(B)** The SNPs that are present between the 129P2_OlaHsd genome and C57Bl/6 reference genome (mm10) in the given window (”A”) are depicted in black. The SNPs that I had identified in C57Bl/6 and *Mtrr*$^{gt/gt}$embryos using whole genome sequencing in the given window (”B”) are shown in blue. The intersection of A and B, 129/P2 SNPs identified in C57Bl/6 and *Mtrr*$^{gt/gt}$embryos (”C”), in the given window are shown in pink. **(C)** For 1 Mb genomic windows either side of the *Mtrr* gene (red line), the number of 129/P2 SNPs identified in *Mtrr*$^{gt/gt}$and C57Bl/6 embryos (”C”) as a percentage of the total number of SNPs identified in the region as a whole (e.g. all 1 Mb windows considered near the *Mtrr* gene) is plotted.

Table 4.5: **The number of 129/P2 SNPs near the *Mtrr* gene**

| **Window** | **129/P2 SNPs relative to C57Bl/6 reference genome (A)** | **SNPs identified in C57Bl/6 and *Mtrr*$^{gt/gt}$embryos (B)** | **129/P2 SNPs in C57Bl/6 and *Mtrr*$^{gt/gt}$embryos (C)** |
|---|---|---|---|
| *Mtrr* locus | 86565 | 942 | 348 |
| *Mtrr* upstream | 35688 | 68 | 9 |
| *Mtrr* downstream | 29448 | 47 | 1 |

Refer to Figure 4.7 for a diagrammatic depiction of A, B and C.

Interestingly, the number of 129/P2 SNPs relative to the C57Bl/6 reference genome (A) was considerably higher at the *Mtrr* locus compared to the average number in the regions flanking the *Mtrr* locus, referred to as the *Mtrr* downstream (chr13: 39,060,780-59,060,780) and *Mtrr* upstream (chr13: 79,06,0780-99,060,780) windows (Table 4.5). This suggested that the region near *Mtrr* may be a genomic region that differs substantially between the two strains, regardless of the *Mtrr*$^{gt}$mutation. Given the presence of 129/P2 origin DNA flanking the *Mtrr* gene, there is a possibility that deleterious mutations present in the 129/P2 strain may be present in the *Mtrr*$^{gt}$model. These are referred to as "passenger mutations" (Vanden Berghe et al., 2015). Of the 86565 variants between the 129/P2 and C57Bl/6 reference genome at the *Mtrr* locus, 14977 fall in regulatory or protein coding regions and of these 2306 are missense mutations. I identified 52 protein coding or regulatory 129/P2 variants at the *Mtrr* locus in my whole genome sequencing analysis dataset. This included four missense mutations (Table 4.6). One of these is predicted to lead to nonsense mediated decay of the *Mterf3* trasncript. The *Mterf3* knockout phenotypes are similar to those seen in the *Mtrr*$^{gt}$model (Padmanabhan et al., 2013; Park et al., 2007). However, *Mterf3* is an essential gene and loss is embryonic lethal (Park et al., 2007). As the 129/P2 strain is viable it suggests that this may not be a true variant. Indeed, few of the 129/P2 variants reported by the Sanger Mouse Genomes Project have been experimentally validated. Despite this, the *Mterf3*

variant and the other protein coding or regulatory 129/P2 variants identified in *Mtrr*$^{gt/gt}$embryos should be explored further in the *Mtrr*$^{gt}$model.

Table 4.6: **129/P2 missense mutations present in *Mtrr*$^{gt/gt}$embryos.**

| Chr | Location | C57Bl/6 Allele | 129/P2 Allele | Gene | Impact | KO phenotype |
|---|---|---|---|---|---|---|
| 13 | 66921762 | T | C | *Mterf3* | NMD transcript variant | Mitochondria defects, enlarged heart, embryonic growth restriction, embryonic lethality (Park et al., 2007) |
| 13 | 67086603 | C | T | *Gm28557* | NMD transcript variant | Unknown |
| 13 | 67086621 | A | C | *Gm28557* | NMD transcript variant | Unknown |
| 13 | 67473412 | G | T | *Zfp874b* | 3′ UTR variant | Unknown |

Chr: chrosmome, KO: knock-out, NMD: Nonsense-mediated decay.

Importantly, a large number of 129/P2 specific SNPs (A) were absent from our sequencing data (Table 4.5). This was unexpected if the region flanking the *Mtrr* gene is indeed 129/P2 derived. Therefore, I considered if the filtering process I applied to the SNPs following GATK identification was removing these variants from our data. Indeed, 60426 of 86565 variants between the 129/P2 and C57Bl/6J strains were present in the unfiltered C57Bl/6 and *Mtrr*$^{gt/gt}$embryo SNP dataset. Deeper analysis revealed these variants were being removed during the first filtering step which removes low quality and biased SNP calls. It will be essential to perform Sanger sequencing to determine if these SNPs were indeed present in the *Mtrr*$^{gt/gt}$embryos to determine if the filtering process was removing true variants. Additionally, there was a large number of SNPs at the *Mtrr* locus (594 of 942) that are not found in 129/P2 mice. Those SNPs at the *Mtrr* locus not of 129/P2 origin should be explored further.

There are also SVs that exist in the 129/P2 genome relative to the C57Bl/6 reference genome at the *Mtrr* locus. These include 472 structural variants and

2268 insertions/deletions. I identified a high frequency of SVs at the *Mtrr* locus. These included mainly candidate SVs (insertions and deletions) and diploid deletions. It will be important to compare the SVs identified by WGS with those present in the 129/P2 genome as I have done for SNPs. This will help determine if the SVs at the *Mtrr* locus reflect the 129/P2 genetic background, if they are *de novo* mutations in the $Mtrr^{gt}$ mouse line or if they result from failure to correctly align gene-trap containing reads back to the reference genome.

### 4.2.6 Variants identified at the *Mtrr* locus have limited effects on gene expression

Following the identification of a high SNP frequency at the *Mtrr* locus I wanted to explore the biological impact of these variants. I investigated if the alternate genetic background and high variant frequency at the *Mtrr* locus led to altered expression of genes within the *Mtrr* locus. I selected eight genes (*Uqcrb, Cts8, Tpbpa, Gas1, Ptch1, Hsd17b3, Srd5a1* and *Nsun2*) to assess. These genes mostly lie downstream of *Mtrr* and the expression of more genes upstream of *Mtrr* should be assessed. Unfortunately I was unable to optimise qPCR primers to assess *Mterf3* expression which was identified to have a missense variant in the 129/P2 strain (Table 4.6). However, the genes *Ptch1, Hsd17b3, Srd5a1* and *Nsun2* have variants in the 129/P2 strain and these were present in $Mtrr^{gt/gt}$embryos in the unfiltered SNP dataset, allowing us to explore if potential passenger mutations had functional relevance.

Using RT-qPCR, I assessed mRNA expression of genes in placentas and embryos at E10.5 and adult testes, quantifying each gene in tissue in which their expression had previously been reported (Figure 4.8 A). I analysed the mRNA expression levels of *Uqcrb, Cts8, Tpbpa*, and *Nsun2* in $Mtrr^{gt/gt}$placentas, from conceptuses with embryonic congenital malformations and conceptuses that were phenotypically normal at E10.5, and compared them to C57Bl/6 control placentas at E10.5. The expression of *Uqcrb, Cts8* and *Nsun2* in $Mtrr^{gt/gt}$ placentas was equivalent to C57Bl/6 controls (Figure 4.8 B-E). However, expression of *Tpbpa* was reduced specifically in $Mtrr^{gt/gt}$placentas from conceptuses with congenital malformations versus C57Bl/6 controls (Figure 4.8 D). However, $Mtrr^{gt/gt}$placentas are small due to an underdevelopment of all placental layers including the ectoplacental cone where *Tpbpa* is expressed (N. Padmanabhan and E. Watson, unpublished data). Therefore, the reduction in *Tpbpa* mRNA expression may reflect poor placenta development rather than a genetic background effect. I next examined the expression of *Hsd17b3, Srd5a1* and *Nsun2* in C57Bl/6, $Mtrr^{+/+}$, $Mtrr^{+/gt}$and $Mtrr^{gt/gt}$adult testes. Expression of *Hsd17b3, Srd5a1* and *Nsun2* was equivalent to C57Bl/6 controls in all *Mtrr* genotypes, although *Hsd17b3* expression was highly variable in $Mtrr^{+/gt}$and $Mtrr^{gt/gt}$males (Figure 4.8 F-H). I measured the expression *Gas1, Nsun2* and *Ptch1* in C57Bl/6

and *Mtrr*$^{gt/gt}$phenotypically normal embryos at E10.5. To specifically examine the effect of the 129/P2 genetic background at the *Mtrr* locus, I also assessed the expression of *Gas1* and *Ptch1* in 129/P2 embryos at E10.5. Expression of *Gas1* but not *Ptch1* was significantly decreased in *Mtrr*$^{gt/gt}$embryos versus C57Bl/6 controls (Figure 4.8 I,J). Expression of *Gas1* and *Ptch1* was equivalent in C57Bl/6 and 129/P2 embryos. This would indicate a genetic background effect was not responsible for the reduced *Gas1* expression in *Mtrr*$^{gt/gt}$embryos, despite the presence of a 129/P2 specific 3'UTR variant approximately 500bp upstream from the *Gas1* transcription start site. However, given the variability in expression of *Gas1* in both 129/P2 and C57Bl/6 embryos, more biological replicates will need to be examined to fully understand the effect of genetic background on *Gas1* expression in the *Mtrr*$^{gt}$model. The expression of all genes at the *Mtrr* locus should be assessed in 129/P2 tissues.

Overall, mRNA expression of genes at the *Mtrr* locus was generally normal. The presence of 129/P2 passenger mutations identified in *Ptch1, Hsd17b3, Srd5a1* and *Nsun2* in the *Mtrr*$^{gt/gt}$embryos in the unfiltered SNP dataset did not result in altered gene expression in the tissues investigated. However, analysis of protein expression will be required to confirm that the 129/P2 genetic background does not influence these genes post-transcriptionally. Therefore from the gene expression data presented I cannot conclude that the 129/P2 region flanking the *Mtrr* gene has no functional consequences in the *Mtrr*$^{gt}$model.

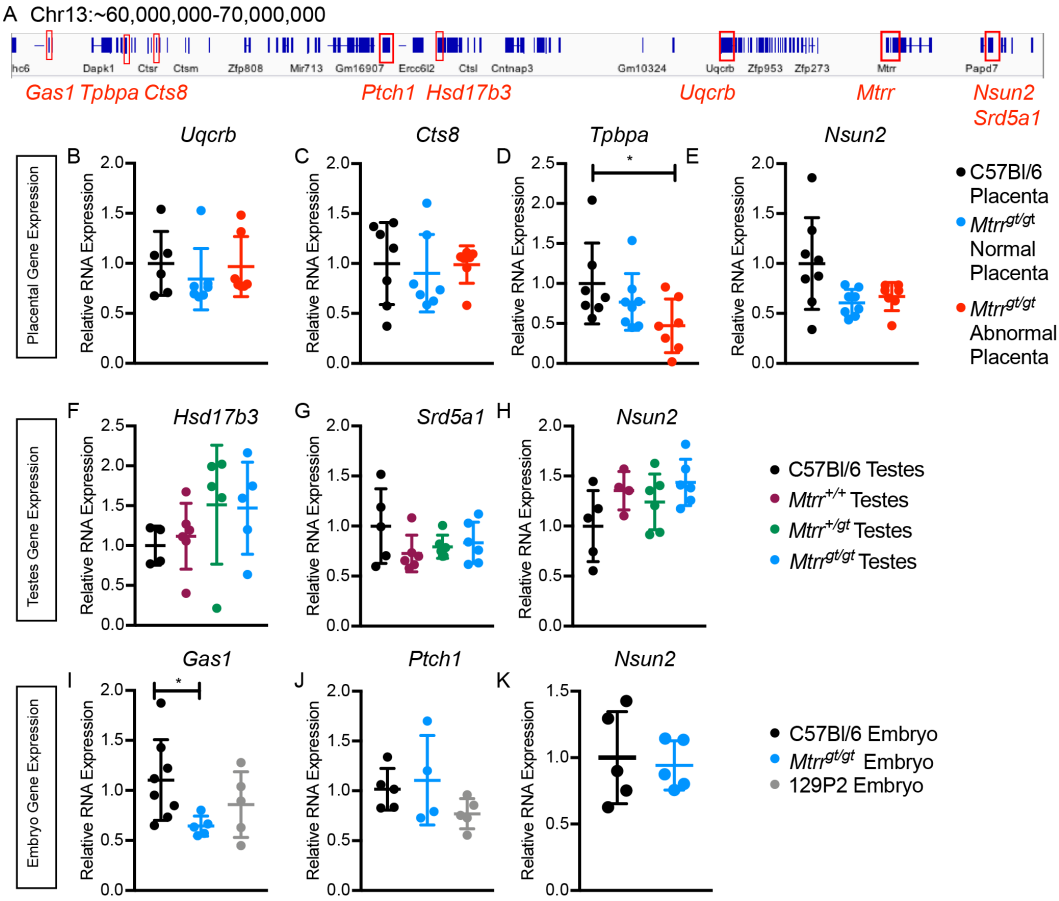



Figure 4.8: **The expression of some genes at the *Mtrr* locus was altered**. **(A)** A region around the *Mtrr* gene (chr13:60162000-70162000) is depicted, with the genes whose expression was assessed, and *Mtrr*, highlighted in red boxes. **(B-E)** Graphical data is shown for C57Bl/6 placentas at E10.5 (black), $Mtrr^{gt/gt}$ placentas from phenotypically normal conceptuses at E10.5 (normal) (blue) $Mtrr^{gt/gt}$ placentas from conceptuses with embryonic congenital malformations at E10.5 (abnormal) (red). Graphs show RT-qPCR analysis of the expression of **(B)** *Uqcrb*, **(C)** *Cts8*, **(D)** *Tpbpa*, **(E)** *Nsun2*. **(F-H)** Graphical data is shown for C57Bl/6 (black), $Mtrr^{+/+}$ (purple), $Mtrr^{+/gt}$ (green) and $Mtrr^{gt/gt}$ (blue ) male testes. Graphs show RT-qPCR analysis of the mRNA expresion of **(F)** *Hsd17b3*, **(G)** *Srd5a1* and **(H)** *Nsun2*. **(I-K)** Graphical data is shown for C57Bl/6 embryos at E10.5 (black), $Mtrr^{gt/gt}$ phenotypically normal embryos at E10.5 (blue) and 129/P2 embryos at E10.5 (grey). Graphs show RT-qPCR analysis of the expression of **(I)** *Gas1*, **(J)** *Ptch1* and **(K)** *Nsun2*. All data is plotted as mean ± standard deviation. Data is represented as fold change compared to C57Bl/6 controls, which are normalised to 1. N=4-8 individuals per genotype. Statistical analysis performed (A-J) one-way ANOVA with Dunnett's multiple comparisons test, (K) independent unpaired t-test. * $p < 0.05$.

### 4.2.7 A tandem duplication on chromsome 19 impacts gene expression

A small number of SVs were present in all *Mtrr*$^{gt}$embryos but absent from the C57Bl/6 controls (Table 4.2). I wanted to explore if these variants may be phenotypically relevant in the *Mtrr*$^{gt}$model. I explored expression of genes near a tandem duplication located on chromosome 19 (chr19: 36911361-37379467) only present in *Mtrr*$^{gt/gt}$embryos and absent from C57Bl/6 controls (Figure 4.9 A). Using RT-qPCR I examined the mRNA expression of *Fgfbp3, Btaf1, March5, Ide* and *Kif11* in *Mtrr*$^{gt/gt}$phenotypically normal embryos at E10.5 compared to C57Bl/6 controls. mRNA expression of *Fgfbp3, March5 and Kif11* was equivalent to C57Bl/6 controls in *Mtrr*$^{gt/gt}$embryos (Figure 4.9 B,D,F). However, the expression of *Btaf1* and *Ide* was elevated in *Mtrr*$^{gt/gt}$embryos with respect to C57Bl/6 controls ($p<0.0001$, Figure 4.9 C,E). *Btaf1* plays a role in controlling embryonic growth and brain development (Wansleeben et al., 2011) and *Ide* is involved in regulating insulin levels and glucose tolerance (Farris et al., 2003). Growth phenotypes, specifically growth enhancement and growth restriction are both present in the *Mtrr*$^{gt}$model (Padmanabhan et al., 2013), therefore I considered if the misexpression of *Btaf1* and *Ide* may be phenotypically relevant in the *Mtrr*$^{gt}$model.

I sought to assess if this tandem duplication on chromosome 19 could contribute to the transgenerational inheritance observed in the *Mtrr*$^{gt}$model. Therefore, I assessed the expression of *Btaf1* and *Ide* in tissues of the F1, F2 and F3 generations of the maternal grandfather pedigree (Figure 4.9 G). The expression of *Btaf1* was not disrupted in F1, F2 or F3 generation embryos at E10.5, even those with congenital malformations, F1 or F2 trophoblast at E10.5 or F1 adult liver compared to respective C57Bl/6 control tissues (Figure 4.9 G). *Ide* expression was also normal in the F1 and F2 generation tissues tested (Chapter 6, Figure 6.4 H,P). Whether the tandem duplication on chromosome 19 is present in wildtype tissue of the F1, F2 and F3 generations needs to be confirmed. The normalisation of *Btaf1* and *Ide* gene expression in F1, F2 and F3 generation tissues may indicate that it is not present. The expression of *Fgfbp3, March5* and *Kif11* should also be assessed in offspring tissues. Altogether, this suggests that the tandem duplication on chromosome 19 examined here

is unlikely to be responsible for the transgenerational inheritance of congenital malformations seen in the *Mtrr*$^{gt}$model. However, functional analysis needs to be extended to all SVs common and unique to *Mtrr*$^{gt/gt}$embryos before I can conclusively rule out a potential genetic contribution to the inheritance of phenotypes in the *Mtrr*$^{gt}$model.

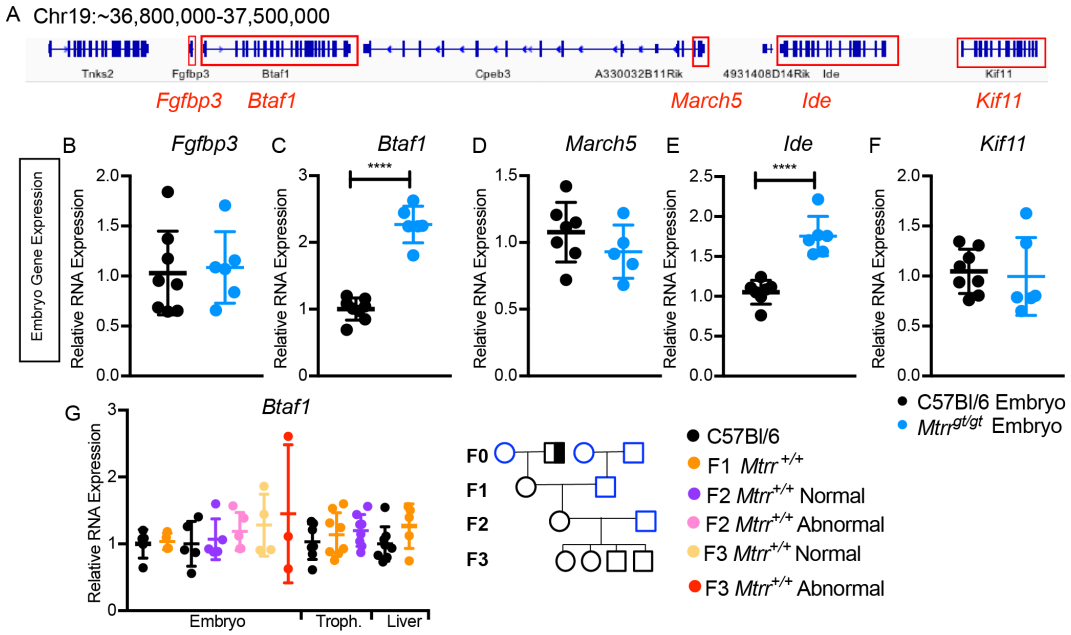



Figure 4.9: **A tandem duplication on chromosome 19 was associated with elevated *Btaf1* and *Ide* expression**. **(A)** The locus on chromosome 19 (chr19: 36800000-37500000) at which there was a tandem duplication in *Mtrr*$^{gt/gt}$embryos is depicted, with the genes whose expression was assessed highlighted in red boxes. **(B-F)** Graphical data is shown for C57Bl/6 embryos at E10.5 (black) and *Mtrr*$^{gt/gt}$ phenotypically normal embryos at E10.5 (blue). Graphs show RT-qPCR analysis of the expression of **(B)** *Fgfbp3*, **(C)** *Btaf1*, **(D)** *March5*, **(E)** *Ide* and **(F)** *Kif11*. **(G)** Graphical data is shown for C57Bl/6 tissues (black), *Mtrr*$^{+/+}$ F1 embryos and trophoblast at E10.5 and adult liver (orange), *Mtrr*$^{+/+}$ F2 phenotypically normal embryos and trophoblast at E10.5 (purple), *Mtrr*$^{+/+}$ F2 embryos with congenital malformations at E10.5 (pink), *Mtrr*$^{+/+}$ F3 phenotypically normal embryos at E10.5 (pale orange) and F3 embryos with congenital malformations at E10.5 (red). The maternal grandfather pedigree is also depicted. All data is plotted as mean ± standard deviation. Data is represented as fold change compared to C57Bl/6 controls, which are normalised to 1. N=4-8 individuals per genotype. Statistical analysis performed (B-F) independent unpaired t-test, (G) one-way ANOVA with Dunnett's multiple comparisons tests, **** $p < 0.0001$.

### 4.2.8 Analysis of telomere length in *Mtrr*$^{gt/gt}$embryos

It was previously shown that telomeres are responsive to environmental stimuli and that changes in telomere length may be inherited intergenerationally (Aiken et al., 2016). Telomeres are also known to be vulnerable to folate deficiency (Bull et al., 2012, 2014). Therefore, I explored telomere length in the *Mtrr*$^{gt}$model. Telomore length can be estimated from whole genome sequencing data by quantifying the number of telomere motif sequences present (Ding et al., 2014). The number of telomere motif copies (TTAGGG or CCCTAA), above a defined threshold, is converted to an estimated telomere length by multiplying by a constant for genome length divided by number of telomere ends and the total number of sequence reads (Ding et al., 2014). I analysed our whole genome sequencing data using TelSeq and identified a small but statistically significant decrease in the average telomere length in *Mtrr*$^{gt/gt}$embryos versus C57Bl/6 controls (p=0.01244, Welch's Two Sample t-test, Figure 4.10 A). C57Bl/6 embryos had a calculated telomere length of 19.86 kb, which is towards the lower limit of previous estimates (20-65kb) (Kipling and Cooke, 1990). This may reflect the fact that Telseq is limited by read count thresholds (Nersisyan and Arakelyan, 2015).

I therefore sought to validate the reduced telomere length using an alternative method. Relative telomere length can be estimated using qPCR. The ratio of telomere repeat copy number to single copy gene number is proportional to the average telomere length (Cawthon, 2002). The factor by which each sample differs from a reference sample is calculated (Cawthon, 2002). I measured the relative telomere length of C57Bl/6 embryos at E10.5 and *Mtrr*$^{gt/gt}$embryos at E10.5 that were phenotypically normal or had congenital malformations. Phenotypically normal *Mtrr*$^{gt/gt}$embryos were included in case the shortened telomere length estimated for *Mtrr*$^{gt/gt}$embryos from the sequencing data was due to cell death in the embryos with congenital malformations (Herrera et al., 1999). All samples were independent of those analysed by whole genome sequencing. Using this approach I did not observe a significant decrease in relative telomere length in *Mtrr*$^{gt/gt}$embryos, regardless of phenotype, versus C57Bl/6 controls ($p>0.3340$, Figure 4.10 B). There was a trend towards reduced telomere length in the embryos with congenital malformations. Telomere

length should be assessed in more *Mtrr*$^{gt/gt}$embryos to provide more conclusive evidence. However, with the data available, I conclude that telomere length was not reduced in *Mtrr*$^{gt/gt}$embryos.

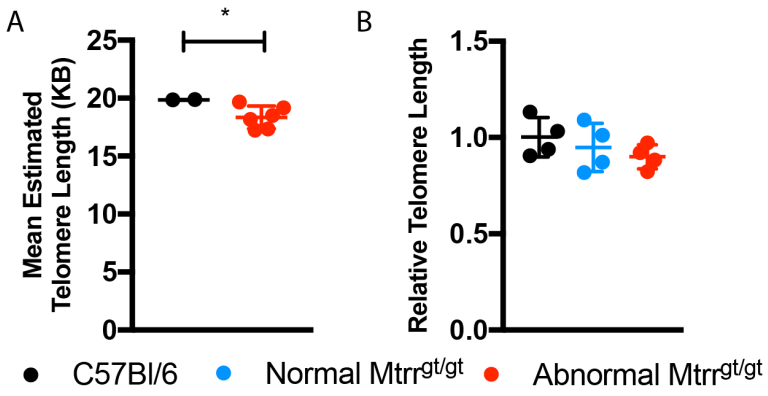



Figure 4.10: **Telomere length was not verifiably reduced in *Mtrr*$^{gt/gt}$embryos versus C57Bl/6 controls (A)** Data for telomere lengths, calculated using TelSeq from the whole genome sequencing data, for C57Bl/6 (black) and *Mtrr*$^{gt/gt}$(red) embryos. **(B)** Data for relative telomere lengths, calculated using qPCR, for C57Bl/6 embryos (black) and *Mtrr*$^{gt/gt}$phenotypically normal embryos (normal) (blue) and *Mtrr*$^{gt/gt}$embryos with congenital malformations (abnormal) (red) at E10.5. Data is represented as relative telomere length compared to C57Bl/6 controls, which are normalised to 1. N=4 per group. Data is plotted as mean ± sd. (A) Unpaired t-test with Welch's correction, (B) one-way ANOVA with Sidak's multiple comparisons test performed, * $p < 0.05$.

### 4.2.9 Transposon expression is not increased in *Mtrr*$^{gt/gt}$embryos

One of the key functions of DNA methylation is to repress transposable elements and maintain genomic stability (Crichton et al., 2014). It is known that DNA methylation is dysregulated in the *Mtrr*$^{gt}$model (Padmanabhan et al., 2013). Therefore I wanted to evaluate if this dysregulation of DNA methylation influenced transposon silencing and consequently impacted genetic stability. The expression of IAP-GAG, IAP-3LTR, LINE-1-ORF2, LINE-1-5UTR and SINE-B1 was assessed using RT-qPCR (Kim et al., 2014). This analysis was performed in *Mtrr*$^{gt/gt}$adult liver tissue and *Mtrr*$^{gt/gt}$ placentas at E10.5 from phenotypically normal conceptuses and from conceptuses with congenital malformations, compared to age and sex matched C57Bl/6 control tissues. Both of these tissues showed global DNA hypomethylation (Padmanabhan et al.,

2013). In adult liver, the mRNA expression of all the transposons assessed was equivalent in *Mtrr*$^{gt/gt}$male livers versus C57Bl/6 controls ($p>0.05$, Figure 4.11 A-E). There was a trend towards upregulation of LINE-1-ORF2 ($p=0.0596$, Figure 4.11 D) with expression particularly variable in *Mtrr*$^{gt/gt}$liver samples. In *Mtrr*$^{gt/gt}$placentas at E10.5 there was decreased expression of SINE-B1 relative to C57Bl/6 placentas ($p<0.05$, Figure 4.11 H). The expression of IAP and LINE-1 elements was similar to C57Bl/6 controls (Figure 4.11 F-G,I-J).

Altogether, this data suggests that in the tissues assessed there is no loss of transposon silencing in the *Mtrr*$^{gt}$model. Indeed, there appears to be reduced SINE-B1 expression in *Mtrr*$^{gt/gt}$placentas, despite global hypomethylation in this tissue. This unexpected finding should be explored further particularly with regard to the known placental phenotypes in the *Mtrr*$^{gt}$model (Padmanabhan et al., 2013). It would also be beneficial to explore transposable element expression in *Mtrr*$^{gt/gt}$embryos to examine a potential contribution to embryonic phenotypes.

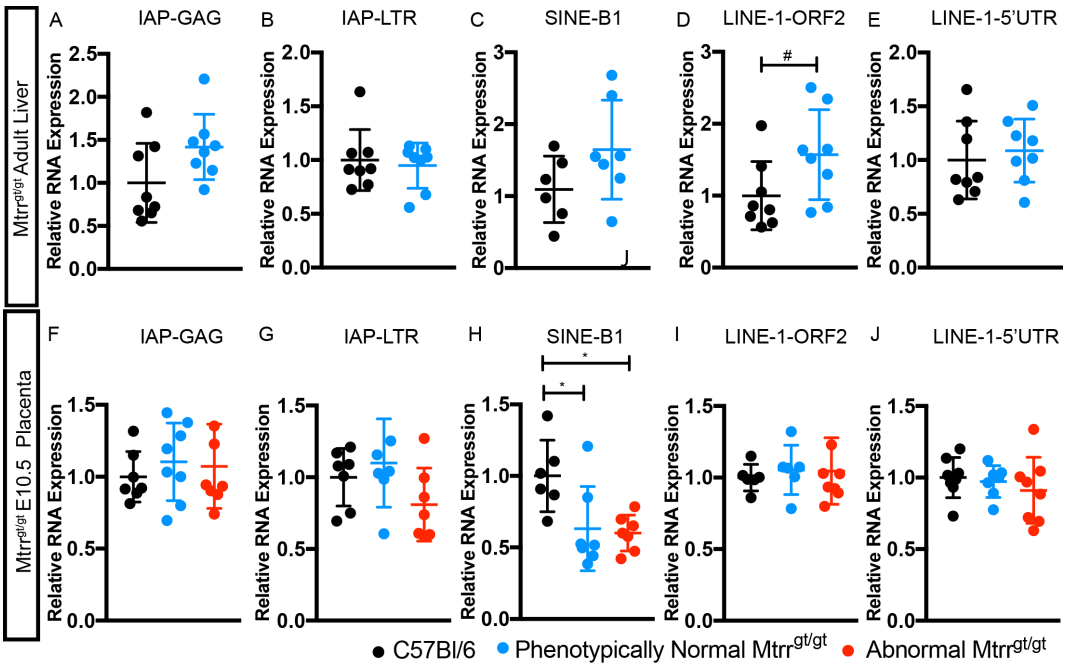



Figure 4.11: **Transposable elements are not upregulated in *Mtrr*$^{gt/gt}$tissues.** **(A-E)** Graphical data is shown for C57Bl/6 (black) and *Mtrr*$^{gt/gt}$ (blue) adult male liver. Graphs show RT-qPCR analysis of the mRNA expression of **(A)** IAP-GAG, **(B)** IAP-LTR, **(C)** SINE-B1, **(D)** LINE-1-ORF2, **(E)** LINE-1-5′ UTR. **(F-J)** Graphical data is shown for C57Bl/6 placentas (black), *Mtrr*$^{gt/gt}$ phenotypically normal placentas (normal, blue) and *Mtrr*$^{gt/gt}$ placentas associated with embryonic congenital malformations (abnormal, red) at E10.5. Graphs show RT-qPCR analysis of the mRNA expression of **(F)** IAP-GAG, **(G)** IAP-LTR, **(H)** SINE-B1, **(I)** LINE-1-ORF2, **(J)** LINE-1-5′ UTR. Data is plotted as mean ± standard deviation. Data is represented as fold change compared to C57Bl/6 controls, which are normalised to 1. N=7-8 individuals per genotype. Statistical analysis performed (A-E) independent unpaired t-test, (F-J) one-way ANOVA with Dunnett's multiple comparisons test. ♯ $p<0.1$, * $p< 0.05$.

## 4.3 Discussion

Assessing the effect of *Mtrr* deficiency on genetic stability is important in order to validate the *Mtrr*$^{gt}$model for the study of TEI. Here I used whole genome sequencing to show that there was a similar frequency of structural variants and single nucleotide polymorphisms in *Mtrr*$^{gt/gt}$embryos with congenital malformations versus C57Bl/6 controls. However, there was an increased frequency of variants on chromosome 13 in *Mtrr*$^{gt/gt}$embryos associated with the *Mtrr* locus. SNP analysis suggested that the region around the *Mtrr* gene was likely to be a 129/P2-derived chromosomal segment, carried over from the ESCs in which the *Mtrr*$^{gt}$mutation was generated. Importantly this 129/P2 region had a number of potential passenger mutations, a few of which were present in *Mtrr*$^{gt/gt}$embryos. However, the 129/P2 genetic background around the *Mtrr* gene did not contribute to altered expression of genes in this region. Beyond the *Mtrr* locus, I identified a small number of variants common to *Mtrr*$^{gt/gt}$embryos but absent from C57Bl/6 control embryos. These are likely established variants within the *Mtrr*$^{gt/gt}$mouse line. Analysis of one such variant, a tandem duplication on chromosome 19, suggested that these variants might influence the expression of some genes in *Mtrr*$^{gt/gt}$tissues. Additionally, my data suggested that telomere length was normal in *Mtrr*$^{gt/gt}$embryos and highlighted that transposon silencing was not altered by the *Mtrr*$^{gt}$mutation, despite disrupted DNA methylation. Overall, I conclude that the genome of *Mtrr*$^{gt/gt}$embryos is relatively stable and therefore genetic instability is unlikely a factor contributing to transgenerational inheritance of congenital malformations observed in *Mtrr*$^{gt}$mice. However, I cannot completely exclude a genetic cause for the transgenerational inheritance observed in the *Mtrr*$^{gt}$model.

My conclusions are made cautiously. I struggled to validate SNPs identified by WGS with Sanger sequencing. This suggests a high false discovery rate within my data. False positive variant calling could have resulted from errors introduced at all stages of the sequencing process including during PCR library amplification, incorrect base calls during sequencing, misalignment during mapping and insufficient read depth (Farrer et al., 2013; Ribeiro et al., 2015). I took extensive steps to minimise false positive variant discovery, including extensive quality control procedures, stringent read alignment and filtering

to remove low quality variants. Indeed GATK, which was used to call SNPs and small indels, performs local realignment of reads in order to reduce false positive SNP calls. Over 20% of the variants I attempted to validate (4 of 18) were only identified in sample *Mtrr*$^{gt/gt}$-4. This sample had low genome coverage (an average of 20x versus at least 30x achieved for all other samples), which may have led to more false positive variant calls. Greater coverage is generally associated with higher quality variant calling, with higher coverage required to detected heterozygous variants than homozygous variants (Sims et al., 2014). A coverage of 30-35x times is generally considered sufficient for accurate variant calling (Ajay et al., 2011; Wu et al., 2017), although upwards of 60x coverage may be recommended for indel detection (Fang et al., 2014). I suggest that it is unlikely that a single factor led to the high false discovery rate observed in our data. Performing the variant calling again, for both SVs and SNPs, with highly stringent parameters may lead to a variant dataset that more accurately reflects the genetic variability of the *Mtrr*$^{gt}$model. Nevertheless, the scope of the validation I performed was limited (I attempted to validate less than 0.2% of the total SNPs identified). Therefore it will be necessary to perform further validations by Sanger sequencing before final conclusions about the accuracy of the variant calling can be drawn. Very few studies validate SNPs identified using whole genome sequencing by an alternative method. If my study is representative, this suggests many other datasets may also contain false positive variant calls.

I performed intersectional analysis on both the SVs and SNPs identified in this study. This enabled me to compare the variants identified in each individual to those identified in all other individuals. The results were starkly contrasting between the SV and SNP datasets. The vast majority of SVs were unique to each individual, perhaps suggesting they are *de novo* mutations in that individual. It would be interesting to assess if variants identified in *Mtrr*$^{gt/gt}$embryos are present in the tissues of the parents of the individual in which the variant was identified. This would also allow a germline mutation rate to be determined. Contrastingly, the majority of SNPs were common across all individuals. These may reflect variants within our C57Bl/6 colony that differ from the reference genome, introduced into the *Mtrr*$^{gt}$line by periodic

backcrossing. I was surprised by the lack of intersection between individuals for SVs and high level of intersection for SNPs. The C57Bl/6 colony and *Mtrr*$^{gt}$colony have been divergent for a number of years but are extensively inbred within each colony. I speculate that the high proportion of SNPs common across all individuals (both *Mtrr*$^{gt/gt}$and C57Bl/6 embryos) represents an artefact of the way in which the SNP calling was performed. Data for all individuals was brought together into a combined/merged data file prior to SNP calling in order to increase SNP calling power. However, I suspect this biased the SNP calling in favour of SNPs that were common across individuals. To test this hypothesis, SNP calling was also performed for each individual. However, while the number of SNPs unique to each individual rose (from $< 16$, to $> 100$), SNPs common to all individuals still represented the vast majority of SNPs called. SV identification was performed on an individual sample basis, which might increase the likelihood of detecting SVs unique to each individual. Further investigation will be required to resolve if this is a data analysis artefact or a true reflection of the intersection of SNPs and SVs in C57Bl/6 and *Mtrr*$^{gt/gt}$mice.

In this study I identified an average of approximately 11,000 variants (6000 SVs and 5000 SNPs) per individual for both C57Bl/6 and *Mtrr*$^{gt/gt}$embryos. This figure represents the most refined data set (SVs: dipoid only, *Mtrr* locus masked, SNPs: extensively filtered). I believe this figure is of an appropriate order. The mutation rate in the mammalian genome is approimately $2.2x10^{9}$ per base pair per year (Kumar and Subramanian, 2002). Less than 1000 SNPs were identified when comparing two $A^{vy}$ littermates (Oey et al., 2015). The mice used in my study were not all littermates, but the populations from which they are derived are extensively inbred. Genomic diversity is known to exist within inbred strains, for example copy number variants are known to be present in C57Bl/6 mice (Watkins-Chow and Pavan, 2008; Mahajan et al., 2016). A comparison of two closely related C757Bl/6 strains: C57Bl/6J and C57Bl/6N identified approximately 10,000 putative variants (Simon et al., 2013). The C57Bl/6J and C57Bl/6N have been maintained separately for 220 generations (Simon et al., 2013). This figure I feel is closest available comparator for my study of *Mtrr*$^{gt}$versus C57Bl/6 mice.

SNP analysis highlighted that an approximately 20 Mb region around the *Mtrr* gene is likely to be 129/P2 derived. The extent of the 129/P2 region at the *Mtrr* locus highlights the limitations of backcrossing. Through genotyping for the gene-trapped allele, we are consistently selecting for this region. There is a low probability of meiotic homologous recombination occurring near a targeted gene resulting in large regions flanking the gene remaining of 129/P2 origin. Indeed, when backcrossing eight generations into C57Bl/6, there is a 0.91 probability that a 1 centiMorgan (cM) region flanking the target gene on both sides is still of donor (129/P2) origin (Lusis et al., 2007). Furthermore, it would be interesting to determine if any regions outside the *Mtrr* locus were 129/P2 genetic background (perhaps with an unknown linkage to the *Mtrr* locus) by mapping our sequencing reads to the 129/P2 genome.

Genomic regions surrounding a transgene are known as the passenger genome (Vanden Berghe et al., 2015). The 'genetic contamination' of the passenger genome, and any mutations present within it, may influence the phenotypes of backcrossed mice (Fontaine and Davis, 2016; Vanden Berghe et al., 2015). For example *Casp1*$^{-/-}$ mice have strong protection against a lethal lipopolysaccharide (LPS) challenge, however this was found to be due to an inactivating passenger mutation in the nearby *Casp11* gene (Kayagaki et al., 2011). Similarly, mice with a mutation in the *Plat* gene, which causes neurological abnormalities, had a 20 Mb region flanking the *Plat* gene that was of 129 origin in otherwise C57Bl/6 mice (Szabo et al., 2016). RNA-Seq identified high levels of differential expression of genes near the *Plat* locus. Independently derived *Plat*$^{-/-}$ mice in a pure C57Bl/6 background did not show anomalous clustering of differentially expressed genes at the *Plat* locus (Szabo et al., 2016), although whether this altered the phenotype observed in these mice was not described. In this study, I identify that 52 129/P2 specific protein coding or regulatory variants are present at the *Mtrr* locus in *Mtrr*$^{gt/gt}$embryos. It is assumed that these co-segregate with the *Mtrr*$^{gt}$allele. However, extensive dysregulation of genes at the *Mtrr* locus was not observed using RT-qPCR, although only a small subset of genes were assessed. *Gas1* was downregulated in the *Mtrr*$^{gt/gt}$embryos versus C57Bl/6 controls. Misexpression of this gene may be phenotypically relevant in the *Mtrr*$^{gt}$model. GAS1 can cause cell cycle arrest (Del Sal et al.,

1992). *Gas1* knockout mice have a range of eye (Lee et al., 2001), digit and limb and cerebellum development phenotypes (Liu et al., 2002). Anecdotally, some *Mtrr*$^{gt}$mice do present with anophthalmia, although the frequency of this phenotype needs to be characterised. Furthermore, RNA-Seq data from *Mtrr*$^{gt/gt}$placentas at E10.5 did show some clustering of differential expressed genes around the *Mtrr* locus, with 15 out of 70 differentially expressed genes located within the *Mtrr* locus (K. Menelaou and E. Watson, unpublished data). These genes are mainly involved in placenta function and therefore may contribute to the placental phenotypes observed in the *Mtrr*$^{gt}$model (K. Menelaou and E. Watson, unpublished data). It will be vital to determine if mutations in the 129/P2 region around *Mtrr* gene contribute in any way to the phenotypes reported in the *Mtrr*$^{gt}$model.

In order to fundamentally address whether there is a genetic contribution to the transgenerational inheritance of phenotypes the *Mtrr*$^{gt}$model, it will be necessary to assess the genomes of wildtype individuals derived from a *Mtrr*$^{+/gt}$parent or grandparent. The transgenerational inheritance of phenotypes observed in the *Mtrr*$^{gt}$model should be independent of the passenger mutations at the *Mtrr* locus. Beyond the F0 generation, all individuals in the transgenerational pedigree are *Mtrr*$^{+/+}$females and C57Bl/6 males. *Mtrr*$^{+/+}$females will not carry the co-segregating 129/P2 region at the *Mtrr* locus. However, it will be necessary to evaluate if variants identified as polymorphic within the *Mtrr*$^{gt/gt}$embryos (such as the tandem duplication on chromosome 19) are present in *Mtrr*$^{+/+}$offspring. Strikingly in the *Mtrr*$^{gt}$model, there is a large degree of variability between individuals. Litters can range from entirely phenotypically normal to entirely phenotypically abnormal (Padmanabhan et al., 2013). Phenotypes are not inherited in Mendelian ratios (Padmanabhan et al., 2013). This is true of litters derived from the *Mtrr*$^{gt/gt}$homozygous mutant pedigree and the transgenerational maternal grandparental pedigrees (Padmanabhan et al., 2013). This suggests that variants common within the *Mtrr*$^{gt}$colony are unlikely to be responsible for the phenotypes observed.

Overall, the data presented in this chapter supports a conclusion that genetic instability is unlikely to be responsible for the transgenerational effect

observed in the *Mtrr*$^{gt}$mouse line. However, I cannot completely exclude a possible genetic contribution based on the data presented here.

# Chapter 5

# Analysis of DNA methylation in *Mtrr* sperm

Principle component analysis in this chapter was performed in collaboration with Dr Russell Hamilton (Department of Genetics and Centre for Trophoblast Research, Cambridge, UK).

## 5.1 Introduction

DNA methylation has been implicated as a mechanistic candidate in a number of models of intergenerational and transgenerational epigenetic inheritance. In order to be directly inherited, DNA methylation changes induced by an environmental stressor must be present in germ cells and escape epigenetic reprogramming events. Germ cell DNA methylation profiles are unique and highly specialised (Reik and Surani, 2015). Sperm and oocyte epigenetic landscapes are not equivalent. The sperm genome is highly methylated (approximately 80-90% of CpGs are methylated), while oocytes have slightly lower global methylation levels (∼40% of CpGs are methylated in an oocyte) (Kobayashi et al., 2012). Sperm methylation covers most of the genome, except at highly CpG-rich regions, and there is a negative correlation between promoter methylation and gene expression (Kobayashi et al., 2012). DNA methylation in the male germline has roles in the regulation of sperm specific gene expression and transposable element silencing (Bourc'his and Bestor, 2004). Oocyte methylation is mainly in genic rather than intergenic regions (Smallwood et al., 2011) and there is a positive correlation between DNA methylation and transcription (Kobayashi et al., 2012). The role of DNA methylation in oocytes remains elusive as it is not required for early embryo development (Hata et al., 2002). Both the sperm and oocyte epigenomes carry important information influencing both the germ cells themselves and the development of the offspring which they give rise to. The establishment of DNA methylation at imprinted genes, loci that are differentially methylated in the male (paternally imprinted) or female (maternally imprinted) germlines, is vital for successful embryo development (Stewart et al., 2016; McGrath and Solter, 1984). Imprinting is maintained in the zygote and controls parent-of-origin specific expression of imprinted genes. Imprinting is a fantastic example of heritable DNA methylation that intergenerationally regulates gene expression. Whether regions beyond imprints behave in a similar manner has been the focus of much recent research.

DNA methylation patterns in the germline have been assessed in an array of intergenerational or transgenerational inheritance paradigms (Chapter 1, Table 1.2). Paternal inheritance is generally studied for its experimental tractability (Bohacek and Mansuy, 2017). For example *in utero* undernourishment of F1

mice leads to a robust metabolic phenotype in the F2 generation (Jimenez-Chillaron et al., 2009). This was associated with hypomethylated and hypermethylated loci in the F1 male sperm (Radford et al., 2014). Similarly, *in utero* vinclozolin exposure in rats, which is associated with transgenerational inheritance of a number of disease phenotypes including male infertility, led to genome-wide DNA methylation changes in the sperm (Beck et al., 2017). Furthermore, traumatic stress during early post-natal life is associated with reduced DNA methylation levels at the glucocorticoid receptor (GR) promoter in sperm (Gapp et al., 2016). The offspring of these males have a reduced coping behaviour phenotype (Gapp et al., 2016). Intriguingly, the GR promoter DNA methylation change could be rescued if the fathers were housed in an enriched environment, demonstrating that DNA methylation at this locus may be highly labile (Gapp et al., 2016). Overall, DNA methylation patterns in sperm are susceptible to alterations resulting from a range of environmental insults.

Folate metabolism plays a vital role in providing methyl groups for cellular methylation reactions and therefore influences DNA methylation patterns (Jacob et al., 1998; Ghandour et al., 2002). Several models have implicated folate deficiency or abnormal folate metabolism in altered sperm DNA methylation patterns and disease in offspring. Sperm DNA methylation patterns were disrupted in males exposed to folate deficiency from preconception onward (Lambrot et al., 2013). These males had offspring with developmental abnormalities (Lambrot et al., 2013). Furthermore, the differentially methylated loci identified using MeDIP-ChIP in this study coincided with developmentally important genes (Lambrot et al., 2013), though a direct causal link between the sperm methylation changes and the birth defects was not demonstrated. Similarly, BALB/c mice deficient for the folate metabolism enzyme MTHFR displayed a small number of hypomethylated and hypermethylated loci in their sperm (Chan et al., 2010). However, imprinted genes and other loci known to acquire methylation during spermatogenesis did not show altered patterns of DNA methylation in mature sperm of C57Bl/6 *Mthfr*$^{-/-}$ males (Chan et al., 2010). A genome-wide approach will be required to determine if any loci show abnormal DNA methylation in sperm from C57Bl/6 *Mthfr*$^{-/-}$ mice. This data might suggest that genetic background or underlying genetic differences

may influence differential DNA methylation that arises from altered folate metabolism.

DNA methylation has previously be shown to be dysregulated in both adult *Mtrr* deficient tissues and $Mtrr^{+/+}$placentas at E10.5 derived from a $Mtrr^{+/gt}$maternal grandparent (Padmanabhan et al., 2013). However germ cell methylation has never been assessed in the $Mtrr^{gt}$model. I therefore hypothesise that altered DNA methylation patterns may be present in germ cells in the $Mtrr^{gt}$model and could be inherited by subsequent generations resulting in the phenotypes observed.

In this chapter, I will consider DNA methylation in sperm of the $Mtrr^{gt}$model. Since the phenotypes and phenotypic frequencies were similar in the wildtype grandprogeny derived from a $Mtrr^{+/gt}$maternal grandmother or grandfather (Padmanabhan et al., 2013), it is equally valuable to assess DNA methylation in sperm or oocytes of $Mtrr^{gt}$mice. However, obtaining sufficient oocytes to isolate adequate quantities of DNA for methylation analysis would be challenging even with superovulation. Additionally, studying maternal transmission of epigenetic information is complicated by confounding factors, such as the uterine environment and maternal care (Bohacek and Mansuy, 2017). Sperm are experimentally tractable, lack confounding factors and theoretically exclusively pass on genetic and epigenetic information to the zygote (Bohacek and Mansuy, 2017). Once I establish whether differential DNA methylation is present in $Mtrr^{+/+}$, $Mtrr^{+/gt}$and $Mtrr^{gt/gt}$mice, I can investigate whether inheritance of abnormal DNA methylation patterns occurs in the $Mtrr^{gt}$model.

## 5.2 Results

### 5.2.1 Sperm were isolated for DNA methylation analysis

In order to establish if germ line DNA methylation patterns were altered by *Mtrr* deficiency, sperm were isolated from the cauda epididymis of C57Bl/6, $Mtrr^{+/+}$, $Mtrr^{+/gt}$and $Mtrr^{gt/gt}$16-20 week old male mice that had been proven fertile by mating with a female mouse. Eight males were assessed per genotype group, derived from at least four different litters. $Mtrr^{gt/gt}$mice have the most severe metabolic derangement (Padmanabhan et al., 2013) and thus were included to allow me to assess the effect of intrinsic *Mtrr* deficiency on sperm methylation. $Mtrr^{+/gt}$males represent the maternal grandfather in the transgenerational pedigree. Differential methylation in these males has implications for the transgenerational inheritance of phenotypes observed in the $Mtrr^{gt}$model. $Mtrr^{+/+}$males were included to provide insight into the role of parental exposure to the $Mtrr^{gt}$mutation on sperm methylation patterns since they were derived from $Mtrr^{+/gt}$intercrosses. Additionally, they may allow differential methylation resulting from genetic mutations present within the $Mtrr^{gt}$line to be identified. C57Bl/6 males were used as controls throughout as they share the same genetic background as the *Mtrr* mice, but have never been exposed to the $Mtrr^{gt}$mutation.

The purity of the sperm samples collected was confirmed by performing bisulfite pyrosequencing of imprinting control regions. *H19* is paternally imprinted thus should be methylated in sperm (Gebert et al., 2010). *Peg3* is maternally imprinted and therefore should be unmethylated in sperm (Lucifero et al., 2002). I confirmed that the *H19* imprinting control region was fully methylated in all sperm samples, compared to the $\sim$50% methylation seen in somatic C57Bl/6 adult liver (Figure 5.1 A). Likewise, *Peg3* was unmethylated in all sperm samples (Figure 5.1 B). Overall, this confirmed I had obtained pure sperm samples with minimal somatic tissue contamination.

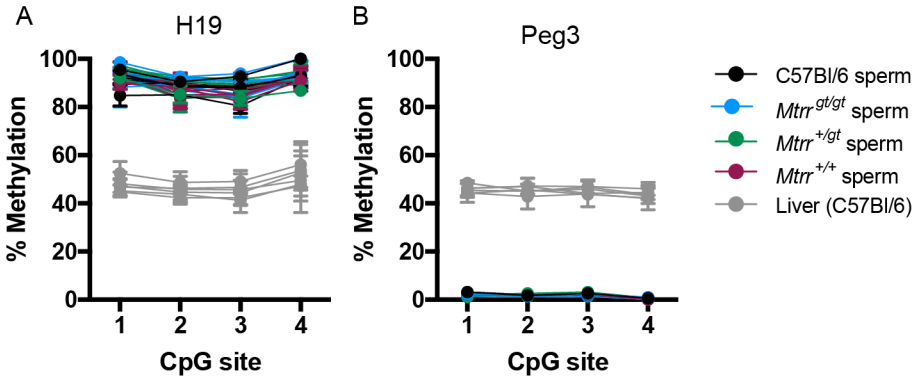



Figure 5.1: **Sperm sample purity was confirmed using bisulfite pyrosequencing of imprinting control regions (ICRs) (A-B)** Graphical data shows the percentage methylation at individual CpG sites assessed by bisulfite pyrosequencing for C57Bl/6 sperm (black), *Mtrr*$^{+/+}$ sperm (purple), *Mtrr*$^{+/gt}$ sperm (green), *Mtrr*$^{gt/gt}$ sperm (blue) and C57Bl/6 adult liver (grey) at **(A)** the *H19* ICR and **(B)** the *Peg3* ICR.

### 5.2.2 Global DNA methylation levels were normal in *Mtrr* sperm

In order to ascertain global DNA methylation levels in sperm I performed mass spectrometry. For each genotype, I analysed three pooled sperm samples whereby each pool comprised sperm from three individuals. In *Mtrr*$^{+/+}$, *Mtrr*$^{+/gt}$and *Mtrr*$^{gt/gt}$sperm, the global 5-methyl cytosine (5mC) level was equivalent to C57Bl/6 controls (Figure 5.2 A). Global hydroxymethylation (5hmC) levels in sperm were also unaffected by the *Mtrr*$^{gt}$mutation (Figure 5.2 B). To determine whether genomic localisation of DNA methylation may be perturbed, I went on to characterise locus-specific DNA methylation patterns.

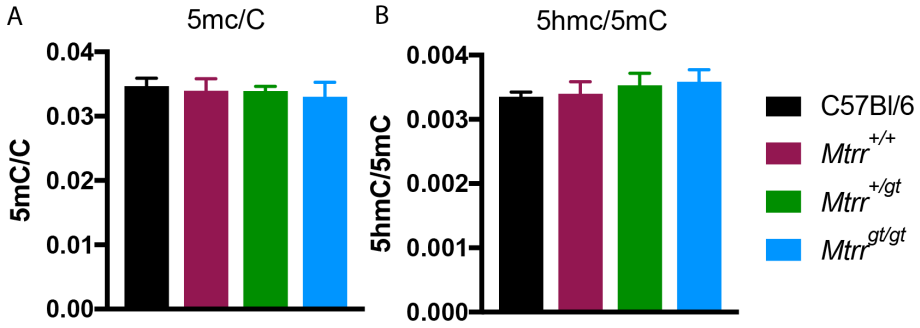



Figure 5.2: **Global DNA methylation levels were unaffected by the *Mtrr*$^{gt}$ mutation. (A-B)** Graphical data shows mass spectrometry quantifications for sperm from C57Bl/6 (black), *Mtrr*$^{+/+}$ (purple), *Mtrr*$^{+/gt}$ (green), and *Mtrr*$^{gt/gt}$ (blue) males for **(A)** 5-methyl cytosine (5mC) **(B)** 5-hydroxymethylcytosine (5hmc).

### 5.2.3 Locus specific DNA methylation was assessed using methylated DNA immunoprecipitation and sequencing (MeDIP-seq)

I analysed the genomic distribution of DNA methyation using methylated DNA immunoprecipitation followed by sequencing (MeDIP-seq) (Weber et al., 2005, 2007). Methylated DNA fragments were enriched using an antibody against 5-methylcytosine and then sequenced (Fouse et al., 2010). A MeDIP-seq approach was adopted as it allowed detection of differential DNA methylation over the whole genome in a nearly unbiased fashion (Taiwo et al., 2012). MeDIP-seq does not have base pair resolution. However, given that CpG methylation at adjacent CpGs tends to be correlated (Eckhardt et al., 2006), I believe the resolution offered by MeDIP (several hundred base pairs) is sufficient to allow functionally relevant DNA methylation changes to be identified. The specificity of the 5-methylcytosine antibody had been previously verified (Radford et al., 2014). No antibody controls were performed to confirm the specificity of the immunoprecipitation reactions. Prior to sequencing, the efficiency of the IP was verified using qPCR to measure enrichment for known methylated regions (*Nanog* and *H19*) in the immunoprecipitatied samples (IP) versus input samples. Regions known to be unmethylated (*H1t* and *TsH2B*) were also assessed (Lambrot et al., 2013). Unmethylated regions were almost undetectable in the IP fraction whereas methylated regions showed enrichment

with respect to the input samples (Figure 5.3 A). Prior to sequencing, library size (300-500bp) was confirmed to be equivalent across all samples by BioAnlyzer analysis (Figure 5.3 B).

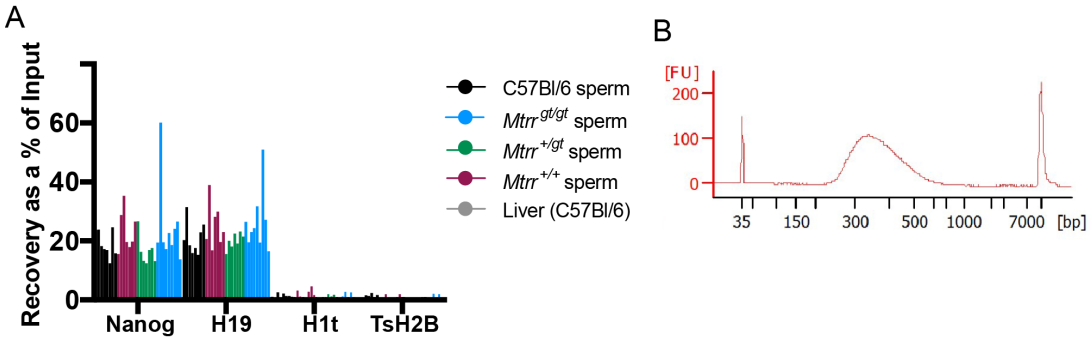



Figure 5.3: **Confirmation of successful MeDIP and verification of MeDIP-seq library size (A)** Graphical data showing recovery as a percentage of input determined using qPCR at regions known to be methylated in sperm, *Nanog* and *H19* ICR, and those known to be unmethylated in sperm, *H1t* and *TsH2B*. Data is shown for C57Bl/6 (black), $Mtrr^{+/+}$ (purple), $Mtrr^{+/gt}$ (green) and $Mtrr^{gt/gt}$ (blue) MeDIP samples. **(B)** A representative BioAnlyzer profile demonstrates a 300-500bp library size.

On average 179 million paired-end mappable reads were obtained per genotype group; 164 million for C57Bl/6 libraries, 172 million for $Mtrr^{+/+}$libraries, 203 million for $Mtrr^{+/gt}$libraries and 179 million for $Mtrr^{gt/gt}$libraries. There was an average of 22 million reads per sample (Table 5.1). This is a lower number of reads (appropriately 30-40% less), both per sample and per comparison groups, than has been obtained in similar studies (Radford et al., 2014; Beck et al., 2017). The number of reads I obtained may only allow interrogation of a subset of around 30% of CpG sites in the genome (Taiwo et al., 2012). Studies have suggested up to 60 millions reads are required to interrogate most CpGs in the genome (Taiwo et al., 2012). Therefore some differentially methylated loci may not have been detected in our analysis.

The quality of MeDIP-seq data was assessed using FastQC. All samples had mean phred quality scores of greater than 30 throughout the read length (Figure 5.4 A). Analysis of the GC content of reads highlighted that MeDIP-Seq libraries did not have normally distributed GC content (Figure 5.4 B). A normal distribution is expected for whole genome sequencing data, however

Table 5.1: **Number of reads obtained for MeDIP-Seq libraries**.

| Genotype | Sample | Total Reads (million bp) |
|---|---|---|
| C57Bl/6 | 1 | 23.8 |
| | 2 | 17.0 |
| | 3 | 19.3 |
| | 4 | 14.9 |
| | 5 | 20.0 |
| | 6 | 21.6 |
| | 7 | 24.4 |
| | 8 | 22.7 |
| $Mtrr^{+/+}$ | 9 | 14.9 |
| | 10 | 22.2 |
| | 11 | 24.4 |
| | 12 | 24.2 |
| | 13 | 19.1 |
| | 14 | 26.8 |
| | 15 | 20.0 |
| | 16 | 20.2 |
| $Mtrr^{+/gt}$ | 17 | 26.0 |
| | 18 | 24.4 |
| | 19 | 22.6 |
| | 20 | 28.6 |
| | 21 | 25.3 |
| | 22 | 23.3 |
| | 23 | 24.5 |
| | 24 | 27.8 |
| $Mtrr^{gt/gt}$ | 25 | 19.9 |
| | 26 | 32.6 |
| | 27 | 23.8 |
| | 28 | 22.6 |
| | 29 | 13.3 |
| | 30 | 21.3 |
| | 31 | 20.5 |
| | 32 | 24.8 |

enrichment of methylated DNA is known to skew the GC content of MeDIP libraries (Staunstrup et al., 2016). Reads were mapped to the reference genome using BowTie2 and alignment rates assessed. Alignment rates were good, with between 95.% and 98.6% of reads mapping in all samples (Figure 5.4 C-D). Approximately 60% of paired-end reads mapped uniquely to the C57Bl/6 reference genome, with some discordant and multimapping (Figure 5.4 C-D). Overall the quality metrics suggested the data was of sufficient quality to proceed with differential methylation analysis.

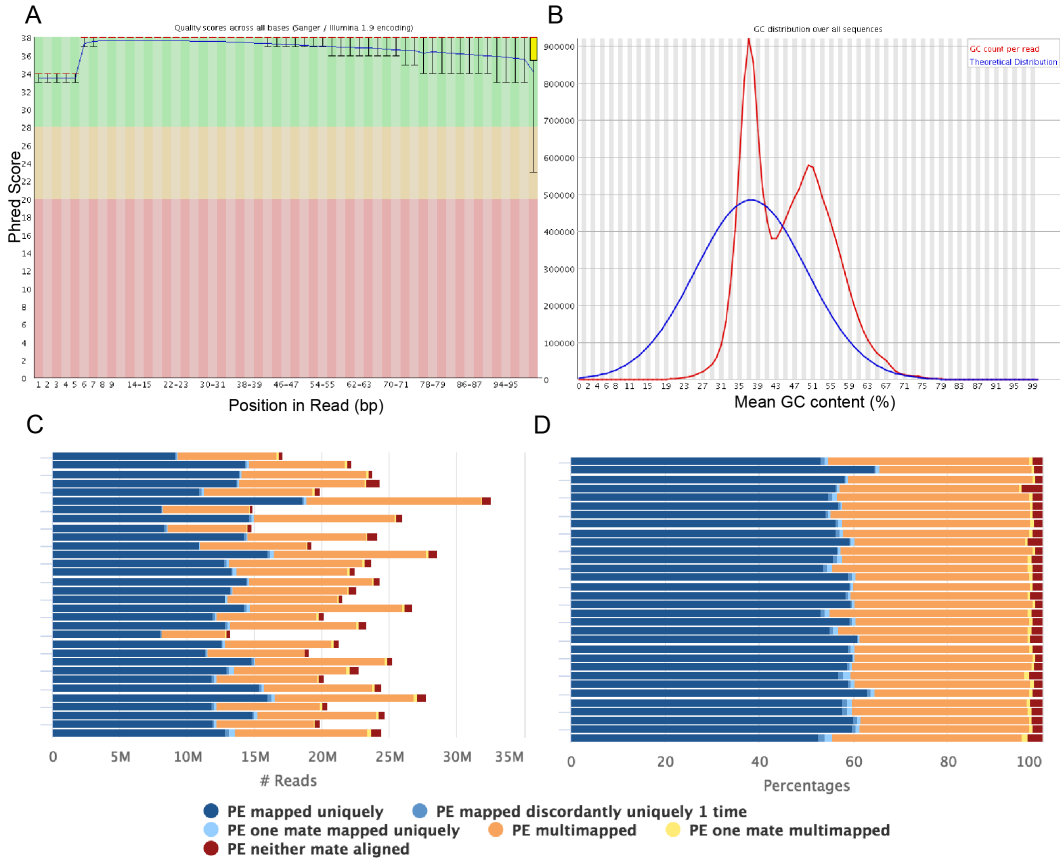



Figure 5.4: **Quality metrics and alignment of MeDIP-seq data. (A)** A representative plot showing Phred quality scores along the read length. **(B)** A representative plot showing the fraction of reads with a given GC content (%). The theoretical distribution is shown in blue and the GC count per read is shown in red. **(C-D)** The **(C)** number and **(D)** percentage of reads that aligned to the reference genome are plotted. Blue: paired-end (PE) reads with at least one mate uniquely aligned, orange: both mates multimapped, yellow: one mate multimapped, red: neither mate aligned. Data is shown for all samples.

Prior to performing differential methylation analysis I used principle component analysis (PCA) to assess sample clustering. The PCA-based clustering was performed on read counts within 5kb genomic windows across the genome. The 500 regions identified as most variable between C57Bl/6 and *Mtrr* samples only were used for PCA analysis. The PCA featuring all samples suggested that the DNA methylation profiles of C57Bl/6, $Mtrr^{+/+}$, $Mtrr^{+/gt}$and $Mtrr^{gt/gt}$samples may not be distinct, demonstrated by the interclustering on the PCA plot (Figure 5.5 A). Performing pairwise PCA however highlighted that while the methylation profiles of C57Bl/6 and $Mtrr^{+/+}$samples appeared shared (Figure 5.5 B), the $Mtrr^{+/gt}$and $Mtrr^{gt/gt}$samples clustered separately from the C57Bl/6 samples suggesting that their methylation profiles differed (Figure 5.5 C,D).

While the $Mtrr^{+/+}$and C57Bl/6 samples generally clustered tightly, $Mtrr^{+/+}$ samples 10 and 12 and C57Bl/6 sample 7 were outlying (Figure 5.5 B). C57Bl/6 sample 7 was also outlying in the $Mtrr^{+/gt}$versus C57Bl/6 and $Mtrr^{gt/gt}$versus C57Bl/6 PCA plots (Figure 5.5 C,D). Therefore C57Bl/6 sample 7 and $Mtrr^{+/+}$ samples 10 and 12 were excluded from differential methylation analysis. I acknowledge that excluding these samples based on the PCA visualisations alone may not have been entirely valid. Their outlying nature may have reflected important heterogeneity within the population. Indeed, phenotypic variability is a key feature of the $Mtrr^{gt}$model (Padmanabhan et al., 2013). As such the exclusion of these samples may have biased the differential methylation calling. A parallel analysis, including samples 10, 12 and 7, would allow me to understand whether excluding this data was influential on differential methylation analysis.

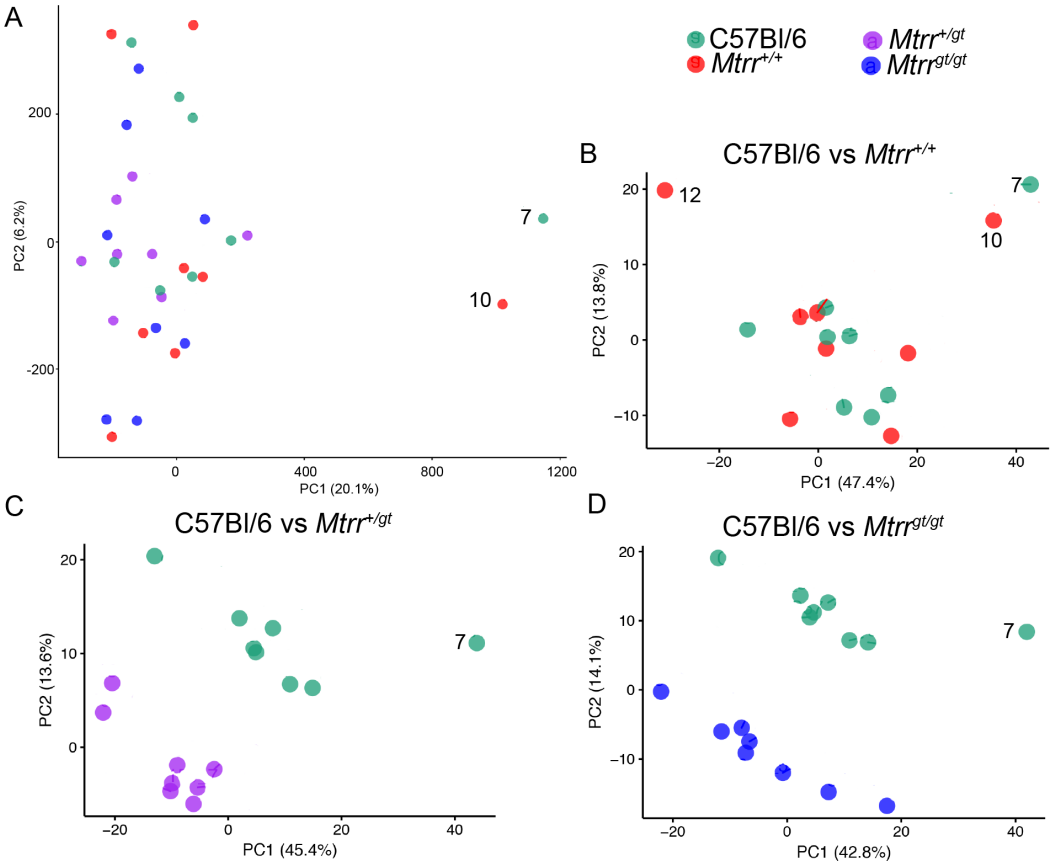



Figure 5.5: **PCA-clustering of C57Bl/6, *Mtrr*$^{+/+}$, *Mtrr*$^{+/gt}$and *Mtrr*$^{gt/gt}$samples.** **(A-D)** PCA based clustering of MeDIP-seq reads over 5kb genomic windows. The 500 most variable windows only were considered in this analysis. PC1 and PC2 were used for clustering and describe the variance between genotypes. Data is shown for **(A)** C57Bl/6 (green) versus *Mtrr*$^{+/+}$(red), *Mtrr*$^{+/gt}$(purple) and *Mtrr*$^{gt/gt}$(blue), **(B)** C57Bl/6 (green) versus *Mtrr*$^{+/+}$only (red), **(C)** C57Bl/6 (green) versus *Mtrr*$^{+/gt}$(purple) only, and **(D)** C57Bl/6 (green) versus *Mtrr*$^{gt/gt}$(blue) only.

Differential methylation analysis was performed using the MEDIPS package in R (Lienhard et al., 2014). Further quality control metrics included as part of this package were performed. Firstly, saturation analysis was used to assess if sufficient sequencing reads were obtained to generate reproducible genome wide methylation profiles. Saturation analysis suggested that the number of sequencing reads I obtained was sufficient to achieve reproducible genome wide coverage for C57Bl/6, *Mtrr*$^{+/+}$, *Mtrr*$^{+/gt}$and *Mtrr*$^{gt/gt}$data sets (Figure 5.6 A-D). Secondly, I assessed sequence pattern coverage in which the number of CpGs covered by short reads is calculated. As expected the majority of short

reads contained very few CpG dinucleotides. The profile of sequence pattern coverage was equivalent between C57Bl/6, *Mtrr*$^{+/+}$, *Mtrr*$^{+/gt}$and *Mtrr*$^{gt/gt}$data sets (Figure 5.6 E-H). Thirdly, I investigated the correlation in short read coverage profiles between samples. High correlation was expected between biological replicates given sufficient sequencing depth. Weaker correlation was expected between divergent groups. Indeed, I observed high correlation within genotype groups (Figure 5.6 I) but slightly weaker correlations when divergent genotypes were compared (Figure 5.6 J). Fourthly, the enrichment of CpGs within the MeDIP-seq datasets versus the whole genome, was calculated. Non-enriched (input) samples would be expected to have an enrichment score of 1. MeDIP-seq should return CpG enriched regions resulting in elevated enrichment scores. I observed an average CpG enrichment score of 2. CpG enrichment was similar across all samples, regardless of genotype (Figure 5.6 K). Finally, I examined the relationship between MeDIP-seq signal intensity and CpG density by plotting a calibration curve. In simplified terms, initially as CpG density increases, MeDIP-Seq signal would be expected to increase as more methylated cytosines would be present. However, as CpG density increases further MeDIP-seq signal no longer increases. This is due to the presence of CpG islands, that are generally unmethylated in mammalian cells (Shen et al., 2007). The calibration curves generated for *Mtrr*$^{+/+}$, *Mtrr*$^{+/gt}$and *Mtrr*$^{gt/gt}$data showed this pattern (Figure 5.6 L-N). Overall, these additional data metrics indicated that the data was suitable for differential methylation analysis.

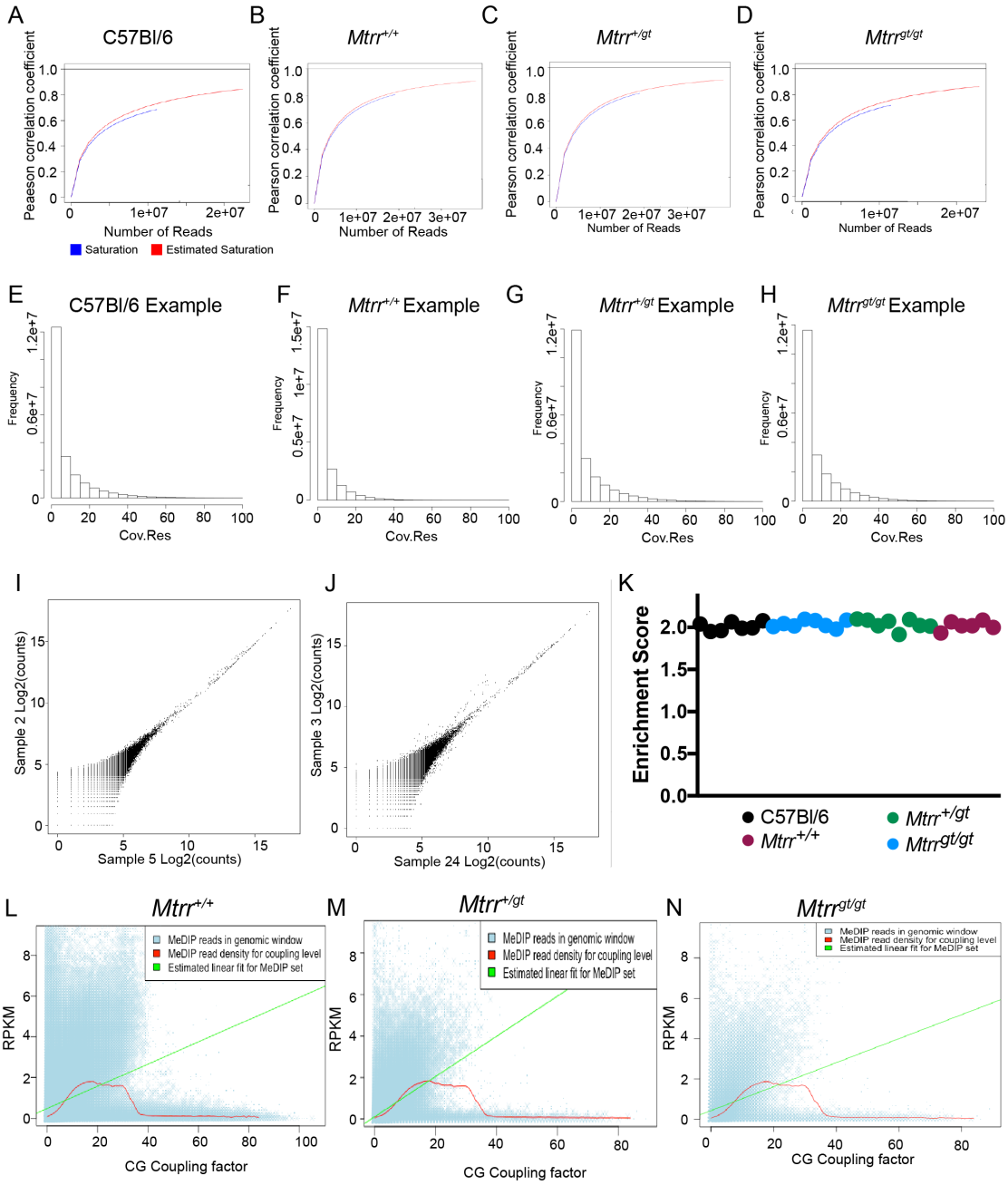

A
C57Bl/6
Peaeson correlation coefficient
Number of Reads
Saturation
Estimated Saturation
B
Mtrr+/+
Pearson correlation coefficient
Number of Reads
C
Mtrr+/gt
Pearson correlation coefficient
Number of Reads
D
Mtrrgt/gt
Pearson correlation coefficient
Number of Reads
E
C57Bl/6 Example
Frequency
Cov.Res
F
Mtrr+/+ Example
Frequency
Cov.Res
G
Mtrr+/gt Example
Frequency
Cov.Res
H
Mtrrgt/gt Example
Frequency
Cov.Res
I
Sample 2 Log2(counts)
Sample 5 Log2(counts)
J
Sample 3 Log2(counts)
Sample 24 Log2(counts)
K
Enrichment Score
C57Bl/6
Mtrr+/+
Mtrr+/gt
Mtrrgt/gt
L
Mtrr+/+
MeDIP reads in genomic window
MeDIP read density for coupling level
Estimated linear fit for MeDIP set
RPKM
CG Coupling factor
M
Mtrr+/gt
MeDIP reads in genomic window
MeDIP read density for coupling level
Estimated linear fit for MeDIP set
RPKM
CG Coupling factor
N
Mtrrgt/gt
MeDIP reads in genomic window
MeDIP read density for coupling level
Estimated linear fit for MeDIP set
RPKM
CG Coupling factor

Figure 5.6: **Quality control metrics included in the MEDIPS package suggest the MeDIP-seq data is good quality (A-D)** Saturation analysis, in which the estimated saturation (red) and calculated saturation (blue) are plotted for **(A)** C57Bl/6, **(B)** $Mtrr^{+/+}$, **(C)** $Mtrr^{+/gt}$and **(D)** $Mtrr^{gt/gt}$MeDIP-seq datasets. **(E-H)** Coverage analysis histograms show the distribution of CpG density in short reads for representative examples of **(E)** C57Bl/6, **(F)** $Mtrr^{+/+}$, **(G)** $Mtrr^{+/gt}$and **(H)** $Mtrr^{gt/gt}$MeDIP-seq libraries. The parameter t, set to 100, specifies the maximal coverage depth to be plotted. **(I-J)** Representative correlation analysis profiles, in which short read coverage is plotted for pairs of MeDIP-seq data, are shown for **(I)** C57Bl/6 versus C57Bl/6 (sample 2 versus sample 5) and **(J)** C57Bl/6 versus $Mtrr^{+/gt}$(sample 3 versus sample 24) comparisons. **(K)** CpG enrichment is plotted for each individual MeDIP-seq dataset for C57Bl/6 (black), $Mtrr^{+/+}$(purple), $Mtrr^{+/gt}$(green) and $Mtrr^{gt/gt}$(blue) samples. **(L-N)** Calibration plots show the dependency between MeDIP-seq signal intensity and CpG density for **(L)** $Mtrr^{+/+}$, **(M)** $Mtrr^{+/gt}$and **(N)** $Mtrr^{gt/gt}$MeDIP-Seq datasets. Green line: estimated linear fit, red line: calibration curve, blue boxes: individual genomic windows.

### 5.2.4 Differentially methylated regions were identified in $Mtrr^{+/+}$, $Mtrr^{+/gt}$and $Mtrr^{gt/gt}$sperm

Differential methylation analysis was performed using pairwise comparisons: C57Bl/6 vs. $Mtrr^{+/+}$, C57Bl/6 vs. $Mtrr^{+/gt}$and C57Bl/6 vs. $Mtrr^{gt/gt}$. I defined differentially methylated loci as windows of DNA (500bp) with greater than 1.5 fold change in reads per kilobase million (as a proxy for methylation) between C57Bl/6 control sperm and *Mtrr* sperm ($p<0.01$). Adjacent differentially methylated loci were merged into differentially methylated regions (DMRs). DMRs were identified in sperm from $Mtrr^{+/+}$, $Mtrr^{+/gt}$and $Mtrr^{gt/gt}$males. The number of sperm DMRs identified increased as the number of male $Mtrr^{gt}$alleles increased. For example, 91 DMRs were identified in sperm from $Mtrr^{+/+}$males, 203 DMRs in sperm from $Mtrr^{+/gt}$males and 599 DMRs in sperm from $Mtrr^{gt/gt}$males. Average DMR size also increased as the number of male $Mtrr^{gt}$alleles increased. For example, DMRs identified in sperm from $Mtrr^{+/+}$males were on average 669bp, DMRs identified in sperm from $Mtrr^{+/gt}$males were 733bp and DMRs identified in sperm from $Mtrr^{gt/gt}$males were 772bp in length. Hypomethylated and hypermethylated DMRs were identified. In sperm from $Mtrr^{+/+}$males 47 (51.6%) DMRs were hypomethylated and 44 (48.4%) DMRs were hypermethylated, in sperm from $Mtrr^{+/gt}$males 71

(35.0%) DMRs were hypomethylated and 132 (65.0%) were hypermethylated and in sperm from *Mtrr*$^{gt/gt}$males 407 (68.0%) DMRs were hypomethylated and 192 (32.0%) hypermethylated (Figure 5.7 A-C). Overall, this data shows that intrinsic or parental exposure to the *Mtrr*$^{gt}$mutation results in changes in sperm DNA methylation patterns.

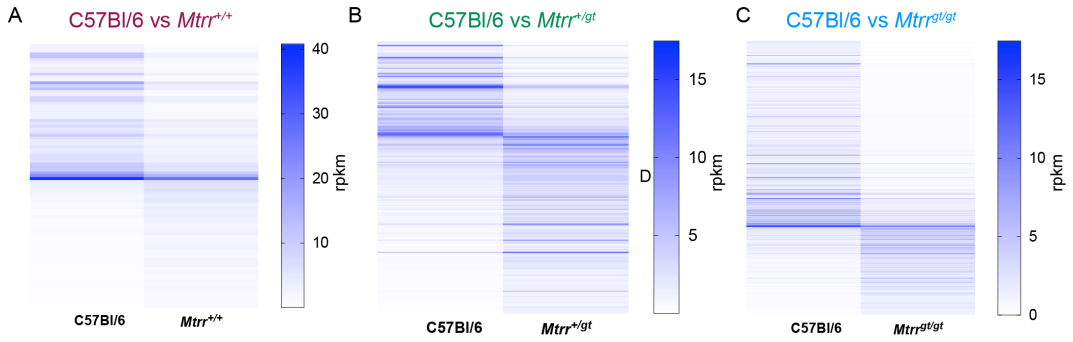



Figure 5.7: **Hypomethylated and hypermethylated DMRs were identified. (A-C)** Heat maps plotting reads per kilobase million (rpkm) show hypomethylated and hypermethylated DMRs identified in sperm from **(A)** *Mtrr*$^{+/+}$, **(B)** *Mtrr*$^{+/gt}$and **(C)** *Mtrr*$^{gt/gt}$males.

### 5.2.5 Validation of DMRs identified using MeDIP-seq via bisulfite pyrosequencing

To ensure the robustness and reliability of MeDIP-seq data I sought to molecularly validate the DMRs identified using bisulfite pyrosequencing. Initially, I performed bisulfite pyrosequencing on a panel of 26 randomly selected DMRs identified in sperm from *Mtrr*$^{gt/gt}$males including 13 hypomethylated and 13 hypermethylated DMRs. I also included two regions not identified as differentially methylated in the MeDIP-seq experiment as controls. Pyrosequencing was performed on sperm samples from eight C57Bl/6 males and eigth *Mtrr*$^{gt/gt}$males. Of these, four sperm samples per group were those sequenced in the MeDIP-seq experiment and four samples per group were independent. Loss DNA of methylation was confirmed in *Mtrr*$^{gt/gt}$sperm at all hypomethylated DMRs assessed (13/13, 100% validation rate) (Figure 5.8 A-M). At these DMRs, differential methylation was generally consistent across multiple CpG sites. Large absolute differences in DNA methylation between C57Bl/6 and *Mtrr*$^{gt/gt}$sperm were observed, ranging from 22.8% to 81.9% (average 50.6%).

On the other hand, gain of methylation was only confirmed at eight hypermethylated DMRs (8/13, 61.5% validation rate) (Figure 5.8 N-Z). This suggests some of the hypermethylated DMRs identified by MeDIP-seq may be false positives. At the hypermethylated DMRs that did validate, substantial differences in methylation were observed ranging from 12% to 59% (average 31.7%). Hypermethylated DMRs that failed to validate tended to have very high levels of methylation ($>$ 90% across all CpGs). The magnitude of LogFC for non-validating DMRs was also lower than that observed for successfully validated DMRs ( the average magnitude of LogFC for non-validating DMRs was 1.3 versus 2.2 for validating DMRs). As expected, the methylation in *Mtrr*$^{gt/gt}$male sperm was equivalent to C57Bl/6 male sperm at the control loci not identified as differentially methylated in MeDIP-seq (Figure 5.8 Aa-Bb).

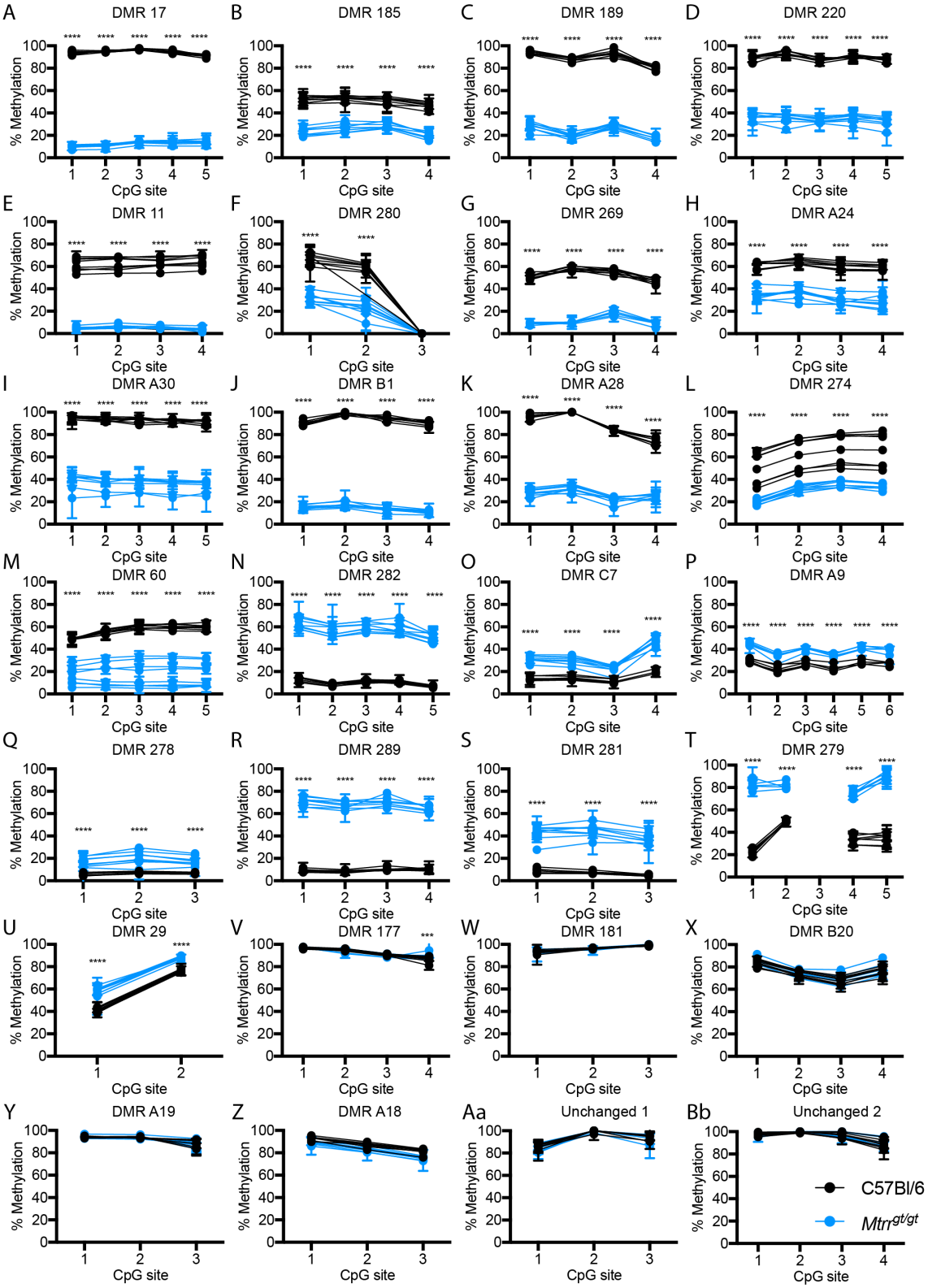

A
DMR 17
B
DMR 185
C
DMR 189
D
DMR 220
E
DMR 11
F
DMR 280
G
DMR 269
H
DMR A24
I
DMR A30
J
DMR B1
K
DMR A28
L
DMR 274
M
DMR 60
N
DMR 282
O
DMR C7
P
DMR A9
Q
DMR 278
R
DMR 289
S
DMR 281
T
DMR 279
U
DMR 29
V
DMR 177
W
DMR 181
X
DMR B20
Y
DMR A19
Z
DMR A18
Aa
Unchanged 1
Bb
Unchanged 2
% Methylation
CpG site
C57Bl/6
Mtrr gt/gt

Figure 5.8: **Bisulfite pyrosequencing validations of DMRs identified in sperm from *Mtrr$^{gt/gt}$* males.** Graphical data is presented for C57Bl/6 sperm (black) and *Mtrr$^{gt/gt}$*sperm (blue) for **(A-M)** hypomethylated DMRs, **(N-Z)** hypermethylated DMRs and **(Aa-Bb)** unchanged regions included as negative controls. N=8 per group. Two-way ANOVA, with Sidak's multiple comparisons test, performed on mean methylation per CpG site per genotype group, **** $p < 0.0001$.

I also sought to validate DMRs identified in *Mtrr$^{+/gt}$*and *Mtrr$^{+/+}$*sperm. I used bisulfite pyrosequencing to assess a panel of 24 DMRs identified in sperm from *Mtrr$^{+/gt}$*males, four of which were hypomethylated and 20 of which were hypermethylated (some of these DMRs were also assessed in *Mtrr$^{gt/gt}$*sperm). Again I used eight sperm samples per genotype group, four of which were used in the MeDIP-seq experiment and four were additional independent samples. Loss of methylation was confirmed in *Mtrr$^{+/gt}$*male sperm at all bar one hypomethylated DMR assessed (3/4, 75% validation rate) (Figure 5.9 A-D). The absolute differences in methylation between sperm from C57Bl/6 and *Mtrr$^{+/gt}$*males were less than seen between sperm from C57Bl/6 and *Mtrr$^{gt/gt}$*males, ranging from 23.4% to 35.9% (average 29.6%). Gain of methylation was confirmed at 10 hypermethlated DMRs assessed (10/20, 50% validation rate) (Figure 5.9 E-X). At one DMR (E66), this represented a significant difference at a single CpG site of the three CpGs assessed (Figure 5.9 L). The absolute differences in methylation between sperm from C57Bl/6 and *Mtrr$^{+/gt}$*males at hypermethylated DMRs were similar to those observed between sperm from C57Bl/6 and *Mtrr$^{gt/gt}$*males, ranging from 5.5% to 54.7% (average 27.3%). Three hypermethylated DMRs were assessed in *Mtrr$^{+/+}$*sperm. Two successfully validated and one showed hypermethylation in *Mtrr$^{+/+}$*sperm at only a single CpG site.

Combining the bisulfite pyrosequencing validation data from *Mtrr$^{+/+}$*, *Mtrr$^{+/gt}$* and *Mtrr$^{gt/gt}$* sperm, differential methylation was confirmed at 16/17 hypomethylated DMRs (94.5%) and 20/36 (55.5%) hypermethylaetd DMRs. While this does suggest some false positive DMR identification within the MeDIP-Seq data, overall there was a high degree of corroboration between the bisulfite pyrosequencing and MeDIP-Seq data. This validation rate is similar to, if not better than, that achieved for other MeDIP-seq experiments (Radford et al.,

2014). I therefore conclude that the majority of DMRs identified by MeDIP-seq represent true methylation changes.

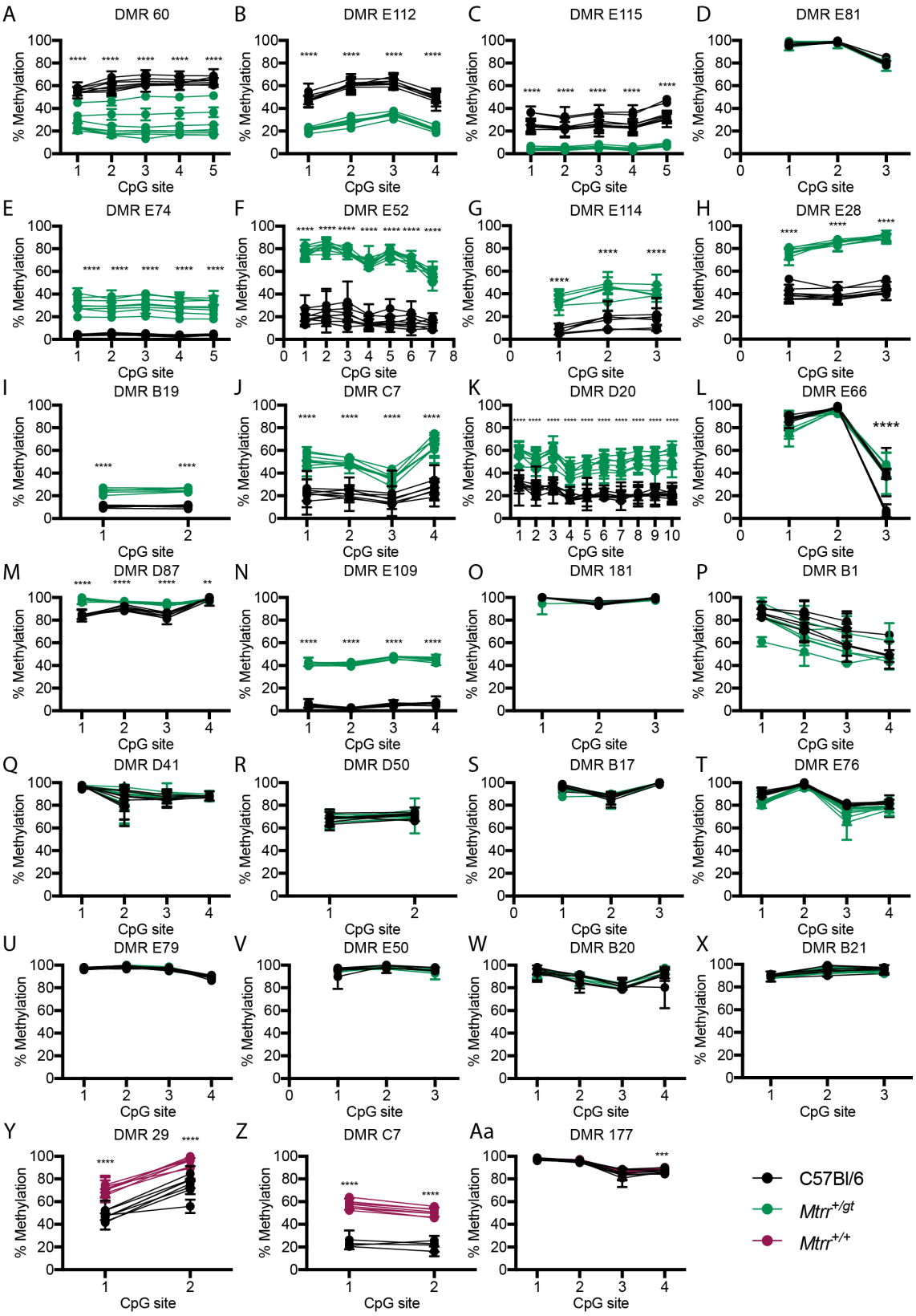

A
DMR 60
B
DMR E112
C
DMR E115
D
DMR E81
E
DMR E74
F
DMR E52
G
DMR E114
H
DMR E28
I
DMR B19
J
DMR C7
K
DMR D20
L
DMR E66
M
DMR D87
N
DMR E109
O
DMR 181
P
DMR B1
Q
DMR D41
R
DMR D50
S
DMR B17
T
DMR E76
U
DMR E79
V
DMR E50
W
DMR B20
X
DMR B21
Y
DMR 29
Z
DMR C7
Aa
DMR 177
% Methylation
CpG site
C57Bl/6
Mtrr+/gt
Mtrr+/+

Figure 5.9: **Bisulfite pyrosequencing validations of DMRs identified in sperm from *Mtrr*$^{+/+}$ and *Mtrr*$^{+/gt}$ males.** Graphical data is presented for C57Bl/6 sperm (black) and *Mtrr*$^{+/gt}$sperm (green) and *Mtrr*$^{+/+}$sperm (purple) for **(A-D)** hypomethylated *Mtrr*$^{+/gt}$DMRs, **(E-X)** hypermethylated *Mtrr*$^{+/gt}$DMRs and **(Y-Aa)** hypermethylated *Mtrr*$^{+/+}$DMRs. N=8 per group. Two-way ANOVA, with Sidak's multiple comparison test, performed on mean methylation per CpG site per genotype group, ** $p< 0.01$, **** $p< 0.0001$.

### 5.2.6 DMRs had clustered chromosomal locations and *Mtrr*$^{gt/gt}$ DMRs were enriched at LTRs

Next I characterised the DMRs to identify if they share common attributes potentially responsible for their differentially methylated status. Firstly, I considered if the methylation level at DMRs was consistent between individuals. The pyrosequencing validations highlighted that the percentage methylation at an individual CpG site was generally highly consistent between animals of the same genotype (Figures 5.8 and 5.9). No differences in DNA methylation were observed between samples used in the MeDIP-Seq experiment and the additional samples used solely for validations. There were two interesting exceptions at which inter-individual variation was observed. At DMR 60, there was variability in methylation levels between *Mtrr*$^{gt/gt}$and *Mtrr*$^{+/gt}$individuals, but not C57Bl/6 individuals (Figure 5.8 M, Figure 5.9 A). At DMR 274, there was C57Bl/6 inter-individual variability, but methylation was consistent between *Mtrr*$^{gt/gt}$individuals (Figure 5.8 L). To investigate inter-individual variability further, I assessed if differential methylation identified at *Mtrr*$^{gt/gt}$male sperm DMRs was present in all *Mtrr*$^{gt/gt}$individuals or just a subset. For each DMR, I calculated the number of *Mtrr*$^{gt/gt}$individuals where the number of reads (in reads per kilobase million, rpkm) in *Mtrr*$^{gt/gt}$mice was at least two standard deviations (SD) different from the mean rpkm of that locus in C57Bl/6 mice. This gave a rough measure of the likelihood of the methylation being statistically different from C57Bl/6 controls at that locus in that individual. In total 85% of *Mtrr*$^{gt/gt}$sperm DMRs (503 of 599) had at least seven individuals with methylation levels at least two SDs different from that in C57Bl/6 mice. Overall, this suggests differential methylation at DMRs was highly consistent across individuals of the same genotype.

Next I considered if DMRs associated with either lowly or highly methylated genomic regions. Bisulfite pyrosequencing allowed me to discern the absolute percentage methylation level of our DMRs. MeDIP-seq only gives a read-out of relative methylation levels at a locus therefore cannot indicate if DMRs are preferentially located in regions with either high or low methylation status. The methylation levels at sperm DMRs spanned from very low, through intermediate, to very high. Sperm DMRs did not seem to associate specifically with lowly methylated regions or highly methylated regions.

I then examined the distribution and genomic localisation of the DMRs to explore if certain regions of the genome were susceptible to or enriched for methylation differences. Firstly, I probed the chromosomal location of the DMRs (Figure 5.10 A-C). Strikingly, DMRs were found in chromosomal clusters (Figure 5.10 A-C). As expected, a cluster of DMRs at the *Mtrr* locus on chromosome 13 was identified in sperm from $Mtrr^{+/gt}$and $Mtrr^{gt/gt}$males (Section 5.2.10). No DMRs were present at the *Mtrr* locus in sperm from $Mtrr^{+/+}$males. Several clusters of DMRs were present in $Mtrr^{+/+}$, $Mtrr^{+/gt}$and $Mtrr^{gt/gt}$sperm on chromosomes 5, 17, 19 and Y. Within these clusters, the DMRs were not always located at exactly the same loci between $Mtrr^{+/+}$, $Mtrr^{+/gt}$and $Mtrr^{gt/gt}$sperm. Clustering of DMRs has been reported in other models of TEI (Haque et al., 2016). The chromosomal locations of the clusters I identified differ from those of previously reported DMR clusters (Haque et al., 2016). This suggests model specific effects on DMR localisation. The clustering of differential methylation could result from increased susceptibility of certain regions of the genome to dysregulation of DNA methylation resulting from the $Mtrr^{gt}$mutation or parental exposure to the $Mtrr^{gt}$mutation. Alternatively, clustering of differential methylation may reflect underlying genetic sequence differences in *Mtrr* mice (explored further in Section 5.2.10).

I considered whether the DMRs were located in coding or non-coding regions. Sperm DMRs identified in $Mtrr^{+/gt}$and $Mtrr^{+/+}$males were most commonly (46-48%) located within introns (Figure 5.11 A-B). However, over 50% of DMRs identified in sperm from $Mtrr^{gt/gt}$males were in intergenic regions (Figure 5.11 C). Methylation in intergenic regions near genes may have a regulatory role controlling gene expression. Indeed, 111 $Mtrr^{gt/gt}$sperm DMRs were

located less than 10 kb upstream of a gene. Furthermore, 21 *Mtrr*$^{gt/gt}$sperm DMRs, 10 *Mtrr*$^{+/gt}$sperm DMRs and 5 *Mtrr*$^{+/+}$sperm DMRs were within potential promoter regions (defined as less than 2kb from a transcription start site). The methylation status of the DMR, hypomethylated or hypermethylated, did not influence their localisation within coding and non-coding regions regardless of whether the DMR was identified in sperm from *Mtrr*$^{+/+}$, *Mtrr*$^{+/gt}$or *Mtrr*$^{gt/gt}$males (Figure 5.11 D-I). The localisation of DMRs with respect to repetitive regions of the genome was then considered. Sperm DMRs identified in *Mtrr*$^{gt/gt}$males appear to be depleted at unique regions and enriched at long terminal repeat elements (LTRs) compared to the proportion of these elements in the genome as a whole ($p<0.01$, two-way ANOVA, Figure 5.11 L). Unique regions and LTRs make up 61.45% and 9.87% of the mouse genome, respectively (Waterston et al., 2002). The enrichment of *Mtrr*$^{gt/gt}$sperm DMRs at LTRs may reflect that LTR elements are more likely to be methylated than other loci. A similar trend toward enrichment at LTRs was observed for *Mtrr*$^{+/gt}$DMRs, although this did not reach statistical significance ($p=0.1662$, two-way ANOVA, Figure 5.11 K). Sperm DMRs identified in *Mtrr*$^{+/+}$males were not found to be enriched or under-represented at any particular repetitive element (Figure 5.11 J). The enrichment of *Mtrr*$^{gt/gt}$DMRs at LTRs, and under-representation at unique regions, was driven by hypomethylated *Mtrr*$^{gt/gt}$DMRs ($p<0.01$, two-way ANOVA, Figure 5.11 Q). For DMRs identified in *Mtrr*$^{+/+}$and *Mtrr*$^{+/gt}$sperm hypomethylated and hypermethylated DMRs had similar localisations with respect to repetitive elements (Figure 5.11 M-R). On the whole, underlying sequence characteristics did not seem to dictate sperm DMR localisation.

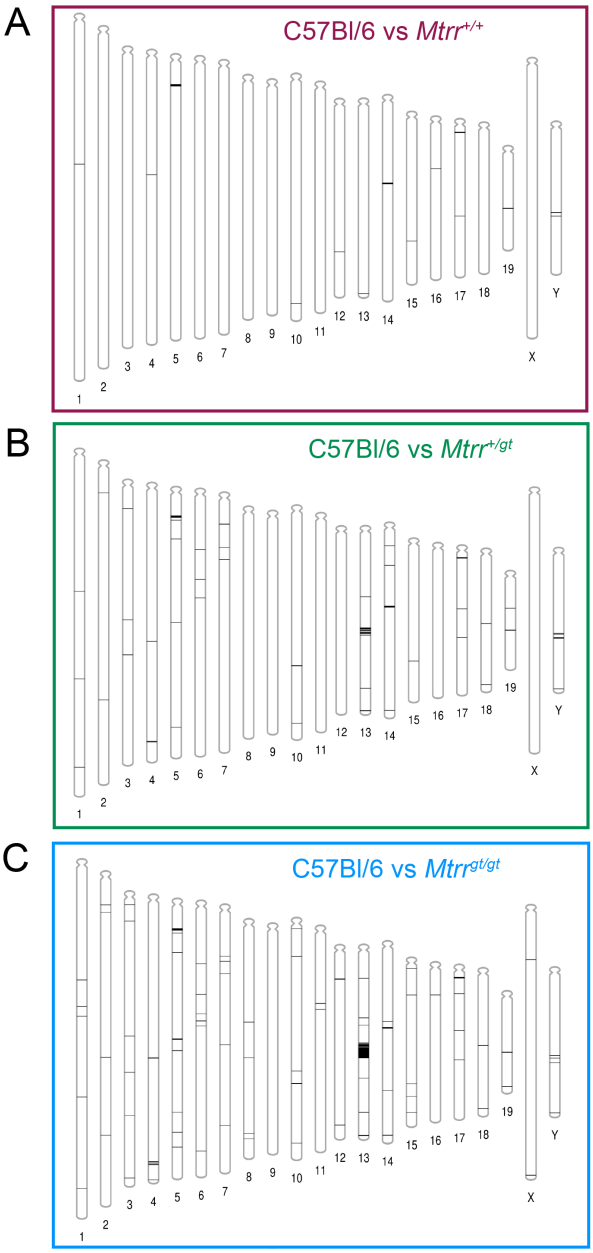



Figure 5.10: **DMRs have clustered chromosomal locations (A-C)** Phenograms show the chromosomal locations of DMRs identified in sperm from **(A)** *Mtrr*$^{+/+}$, **(B)** *Mtrr*$^{+/gt}$and **(C)** *Mtrr*$^{gt/gt}$males.

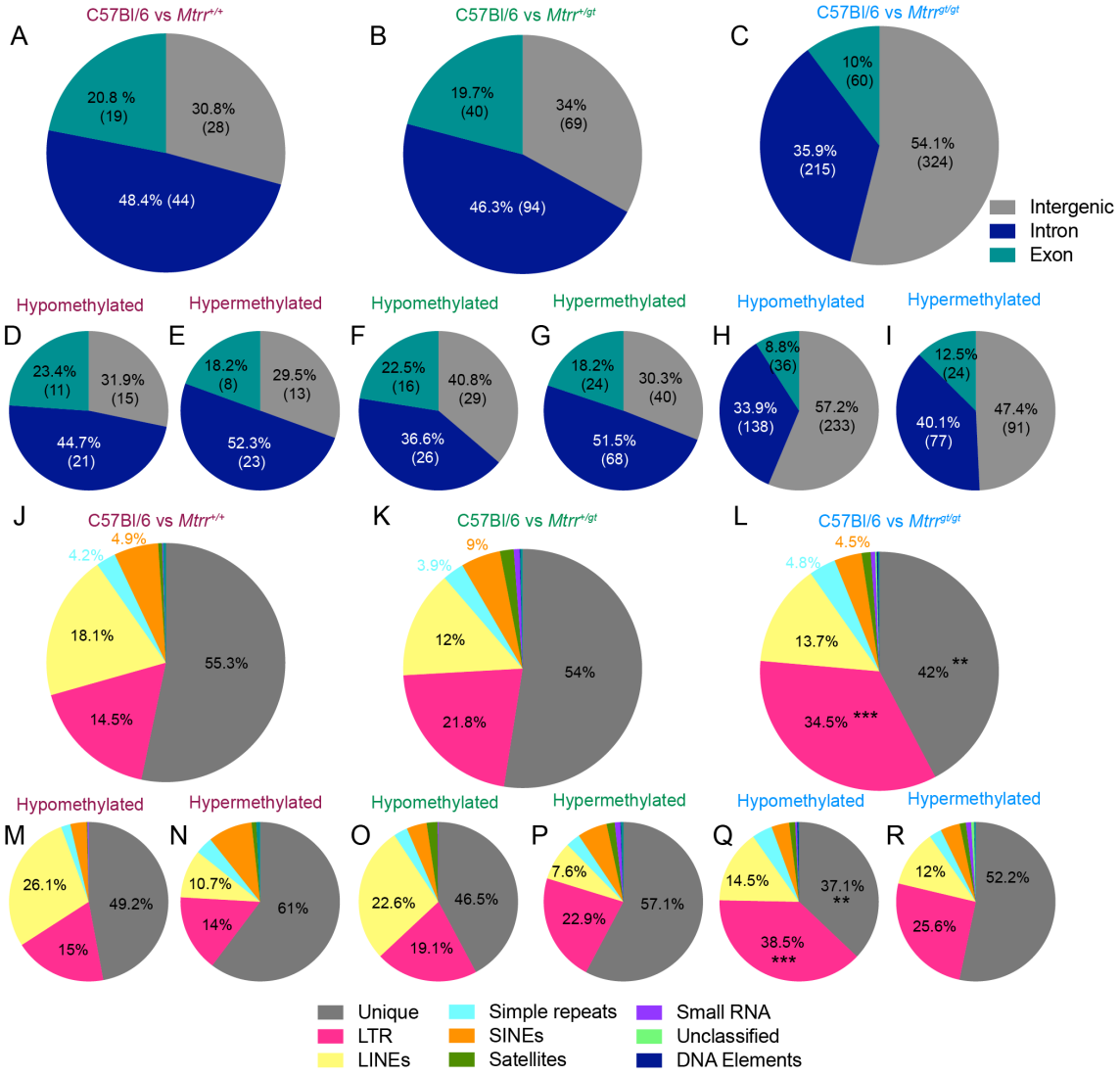



Figure 5.11: ***Mtrr$^{gt/gt}$*DMRs may be enriched at LTRs (A-I)** Relative distribution of DMRs detected by MeDIP-seq among intergenic regions (grey), exons (teal) and introns (dark blue) for **(A)** all *Mtrr*$^{+/+}$DMRs, **(B)** all *Mtrr*$^{+/gt}$DMRs, **(C)** all *Mtrr*$^{gt/gt}$DMRs, **(D)** hypomethylated *Mtrr*$^{+/+}$DMRs, **(E)** hypermethylated *Mtrr*$^{+/+}$DMRs, **(F)** hypomethylated *Mtrr*$^{+/gt}$DMRs, **(G)** hypermethylated *Mtrr*$^{+/gt}$DMRs, **(H)** hypomethylated *Mtrr*$^{gt/gt}$DMRs and **(I)** hypermethylated *Mtrr*$^{gt/gt}$DMRs. **(J-R)** Relative distribution of DMRs detected by MeDIP-seq among unique regions (grey), long terminal repeats (LTRs) (pink), long interspersed repeats (LINEs) (yellow), simple repeats (pale blue), short interspersed repeats (SINEs) (orange), satellites (green), small RNAs (purple), DNA elements (navy blue) and unclassified repeats (lime green) for **(J)** all *Mtrr*$^{+/+}$DMRs, **(L)** all *Mtrr*$^{+/gt}$DMRs, **(M)** all *Mtrr*$^{gt/gt}$DMRs, **(P)** hypomethylated *Mtrr*$^{+/+}$DMRs, **(N)** hypermethylated *Mtrr*$^{+/+}$DMRs, **(O)** hypomethylated *Mtrr*$^{+/gt}$DMRs, **(P)** hypermethylated *Mtrr*$^{+/gt}$DMRs, **(Q)** hypomethylated *Mtrr*$^{gt/gt}$DMRs and **(R)** hypermethylated *Mtrr*$^{gt/gt}$DMRs. Two-way ANOVA with Dunnett's multiple comparisons test performed, ** $p<0.01$, *** $p<0.0001$.

Differential methylation at both CpG Islands (CGIs) and regions of low CpG density may play important roles in the regulation of genes (Stadler et al., 2011). Previous studies have indicated that promoter epimutations identified in sperm are associated with regions with a low CpG density (Skinner and Guerrero-Bosagna, 2014). Therefore, I calculated the CpG density of the DMRs identified in this study. The average CpG densities of *Mtrr*$^{+/+}$, *Mtrr*$^{+/gt}$and *Mtrr*$^{gt/gt}$DMRs were 2.7, 3.6 and 3.6 CpGs per 100 bp, respectively. The CpG density of *Mtrr*$^{+/+}$DMRs was less than that of *Mtrr*$^{+/gt}$and *Mtrr*$^{gt/gt}$DMRs ($p<0.05$, one-way ANOVA, Figure 5.12 A). The biological significance of and underlying reason for this were uncertain. There was a trend towards increasing CpG density with increasing DMR size (Figure 5.12 B). Hypermethylated DMRs had higher CpG densities than hypomethylated DMRs ($p=0.0192$, one-way ANOVA, Figure 5.12 C). CGIs, their shores (0-2 kb either side of the CGI) and shelves (2-4 kb either side of the CGI) are known for their important roles in the regulation of gene expression (Deaton and Bird, 2011; Faulk et al., 2015). A small subset of DMRs identified in sperm from *Mtrr*$^{+/+}$, *Mtrr*$^{+/gt}$and *Mtrr*$^{gt/gt}$males were located within CGIs, CGI shores and CGI shelves (Figure 5.12 D-F). The DMRs located at CGIs, CGI shores and CGI shelves tended to be hypermethylated, although this was not statistically significant ($p>0.05$, two-way ANOVA, Figure 5.12 G-L). Altogether, this data suggests that in the *Mtrr*$^{gt}$model the majority of DMRs are located within CpG deserts, although some DMRs are located in CGI regions.

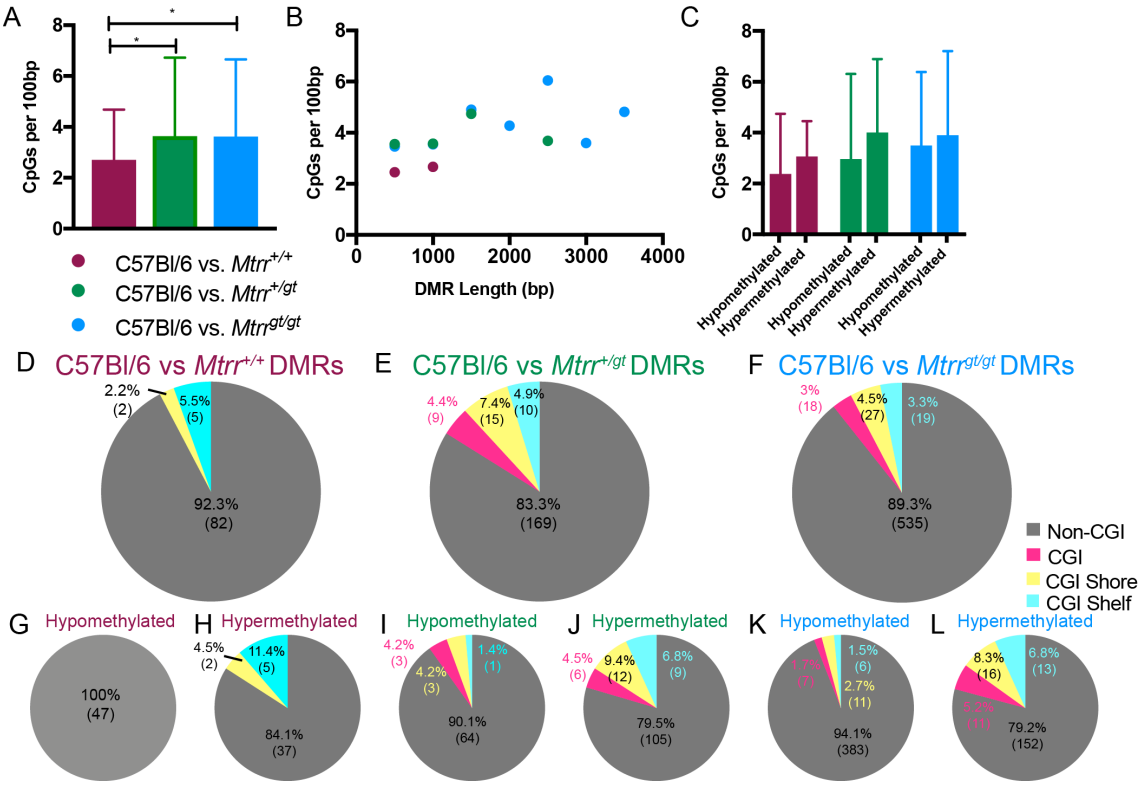



Figure 5.12: **DMRs generally have a low CpG density, but some associate with CpG island regions (A-C)** Graphical data shows CpG density (CpGs per 100bp) for sperm DMRs identified in $Mtrr^{+/+}$(purple), $Mtrr^{+/gt}$(green) and $Mtrr^{gt/gt}$(blue) males. Plots illustrate **(A)** The average CpG density for all DMRs (data is plotted as mean ± sd) **(B)** CpG density as a function of read length and **(C)** Average CpG density for hypomethylated and hypermethylaed DMRs (data is plotted as mean ± sd). **(D-L)** Relative distribution of DMRs detected by MeDIP-seq among CpG islands (CGI) (pink) and CGI shores (yellow) and CGI shelves (turquoise) or non-CGI regions (grey) for **(D)** all $Mtrr^{+/+}$DMRs, **(E)** all $Mtrr^{+/gt}$DMRs, **(F)** all $Mtrr^{gt/gt}$DMRs, **(G)** hypomethylated $Mtrr^{+/+}$DMRs, **(H)** hypermethylated $Mtrr^{+/+}$DMRs, **(I)** hypomethylated $Mtrr^{+/gt}$DMRs, **(J)** hypermethylated $Mtrr^{+/gt}$DMRs, **(K)** hypomethylated $Mtrr^{gt/gt}$DMRs, **(L)** hypermethylated $Mtrr^{gt/gt}$DMRs. (A) one-ANOVA, with Sidak's multiple comparisons test, (C) one-way ANOVA performed, (D-L) Two-way ANOVA with Tukey's multiple comparisons test * $p < 0.05$.

### 5.2.7 DMRs were enriched at nucleosome retaining regions

Although most histones are replaced by protamines in sperm, approximately 1% are retained (Hisano et al., 2013). Nucleosomes are preferentially retained in regions with a high CpG content and low levels of DNA methylation (Erkek et al., 2013). Despite the overall low CpG density of *Mtrr* DMRs, some DMRs with high CpG densities were identified. Therefore, I assessed whether our DMRs coincided with nucleosome retaining regions in sperm (Erkek et al., 2013). Overall, 34.1% of $Mtrr^{+/+}$DMRs (31 of 91), 28.1% of $Mtrr^{+/gt}$DMRs (57 of 203) and 14.5% of $Mtrr^{gt/gt}$(87 of 599) DMRs coincided with regions retaining nucleosomes in sperm. As a crude estimation of genome wide nucleosome retention levels, I calculated the nucleosome retention rate in 10,000 randomly selected 500bp windows. Of these windows only 0.105% coincided with nucleosome retaining regions. With respect to this value, there is a significant enrichment in nucleosome retention at DMRs identified in sperm from $Mtrr^{+/+}$, $Mtrr^{+/gt}$and $Mtrr^{gt/gt}$males ($p<0.0001$, binomial test). This data suggests that some differential methylation in sperm of $Mtrr^{+/+}$, $Mtrr^{+/gt}$and $Mtrr^{gt/gt}$males was present at histone retaining regions therefore exists in a broader chromatin context.

### 5.2.8 DMRs were associated with genes involved in a number of biological processes

Next I considered the potential functional relevance of the sperm DMRs. I identified all genes closest to a DMR (least genomic distance between the DMR and the gene transcription start site). I identified 33, 105 and 222 genes associated with DMRs identified in sperm from $Mtrr^{+/+}$, $Mtrr^{+/gt}$and $Mtrr^{gt/gt}$males, respectively. The majority of these genes (approximately 85%) were within 10kb of a DMR identified in sperm from $Mtrr^{+/+}$, $Mtrr^{+/gt}$or $Mtrr^{gt/gt}$ males. I performed gene ontology analysis using DAVID to investigate if DMRs were preferentially associated with genes involved in any particular biological processes. This analysis showed that differential methylation in $Mtrr^{gt/gt}$sperm was associated with genes involved in the regulation of transcription and a number of immune response pathways (Figure 5.13 C). The highly significant enrichment of genes involved in regulation of transcription near DMRs from $Mtrr^{gt/gt}$sperm likely

reflects the cluster of zinc finger proteins on chromosome 13 at the *Mtrr* locus. Differential methylation in $Mtrr^{+/gt}$and $Mtrr^{+/+}$sperm was associated with genes involved in G-protein coupled receptor (GPCR) signalling and mitotic spindle orientation (Figure 5.13 A-B). Genes involved in ERK signalling and neurogenesis were associated with DMRs identified in $Mtrr^{+/+}$, $Mtrr^{+/gt}$and $Mtrr^{gt/gt}$males (Figure 5.13). Both these terms result from DMRs clustered on chromosome 17 near the *Dynlt1a*, *Dynlt1b* and *Dynlt1c* genes. ERK signalling is involved in the regulation of cell cycle and cell proliferation (Shaul and Seger, 2007). Genes involved in neurogenesis are interesting given the neural tube defects observed in the $Mtrr^{gt}$model (Padmanabhan et al., 2013). Overall, the genes near DMRs could contribute to congenital malformations observed in $Mtrr^{gt}$model, although many genes known to be involved in embryonic development are not associated with sperm DMRs.

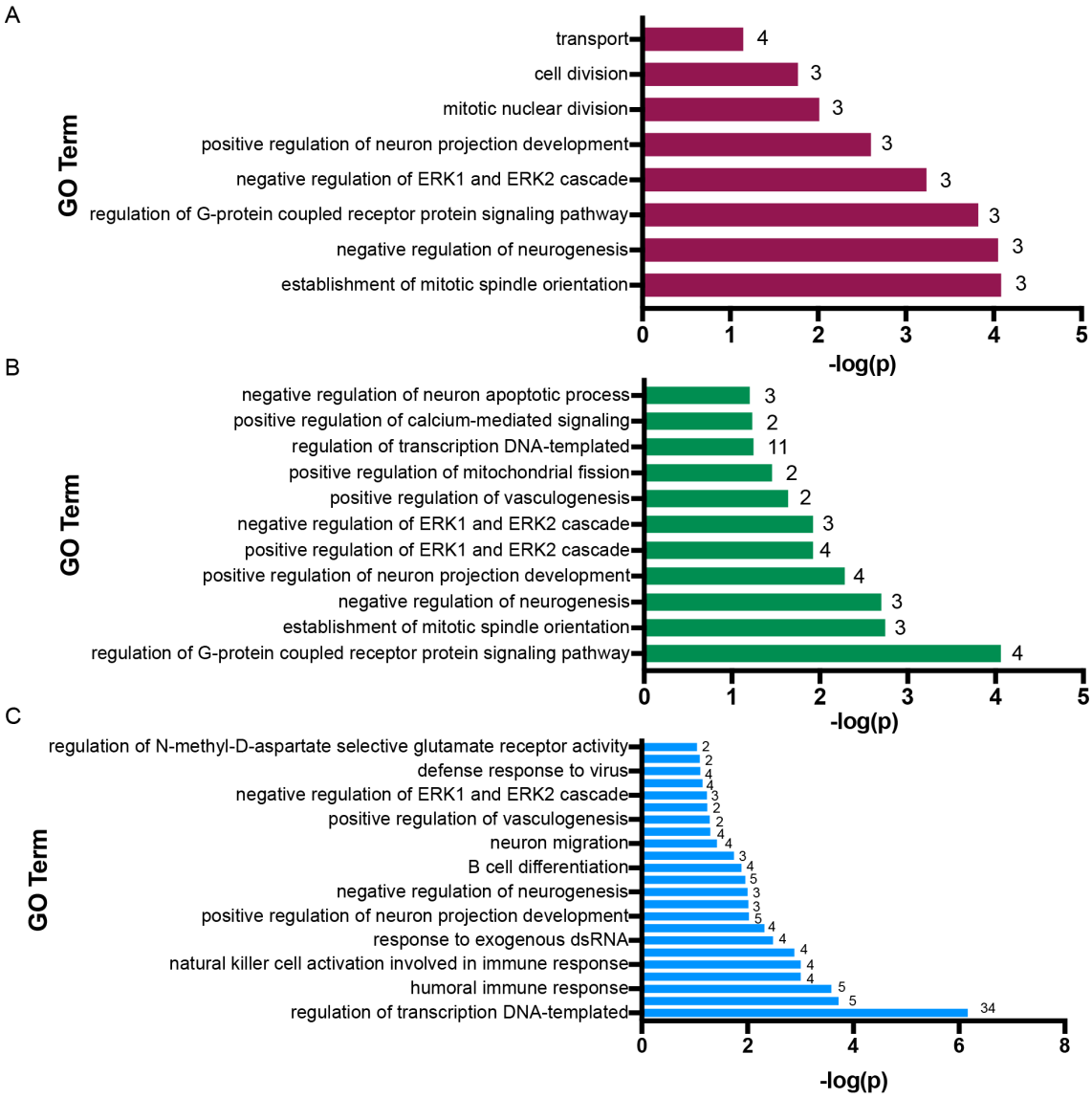



Figure 5.13: **Genes implicated in ERK signalling and neurogenesis are associated with DMRs in *Mtrr*$^{+/+}$, *Mtrr*$^{+/gt}$and *Mtrr*$^{gt/gt}$sperm. (A-C)** Gene ontology analysis performed using DAVID is presented for genes closest to **(A)** *Mtrr*$^{+/+}$sperm DMRs (purple), **(B)** *Mtrr*$^{+/gt}$sperm DMRs (green) and **(C)** *Mtrr*$^{gt/gt}$sperm DMRs (blue). Data is presented as -log(p-value) for each GO term, with the number of genes given at the end of each bar.

### 5.2.9 A subset of DMRs were common across all *Mtrr* genotypes

Gene ontology analysis and chromosomal clustering of DMRs highlighted that some DMRs may be common to all *Mtrr* sperm. Therefore I compared the

DMRs identified in sperm from *Mtrr*$^{+/+}$, *Mtrr*$^{+/gt}$and *Mtrr*$^{gt/gt}$males. This analysis could be used indicate if DNA methylation at certain loci was more susceptible to 1) the intrinsic presence of the *Mtrr*$^{gt}$allele (common to *Mtrr*$^{+/gt}$and *Mtrr*$^{gt/gt}$sperm), 2) parental exposure to abnormal folate metabolism (common to *Mtrr*$^{+/+}$, *Mtrr*$^{+/gt}$and *Mtrr*$^{gt/gt}$sperm) or 3) results from genetic background effects (also common to *Mtrr*$^{+/+}$, *Mtrr*$^{+/gt}$and *Mtrr*$^{gt/gt}$sperm). A total of 147 loci show differential DNA methylation in both *Mtrr*$^{+/gt}$and *Mtrr*$^{gt/gt}$sperm (Figure 5.14 A). Differential methylation at the *Mtrr* locus itself accounts for 37 (25%) of these DMRs. Additionally, 54 of these DMRs are also present in *Mtrr*$^{+/+}$sperm (Figure 5.14 A). A small number of DMRs are common between *Mtrr*$^{+/+}$and *Mtrr*$^{+/gt}$sperm or *Mtrr*$^{+/+}$and *Mtrr*$^{gt/gt}$sperm exclusively (14 and 5 DMRs, respectively). The large number of shared DMRs suggests that overall the *Mtrr*$^{gt}$mutation and parental exposure to the *Mtrr*$^{gt}$mutation affects DNA methylation in a somewhat predictable manner.

I characterised the DMRs common across all genotypes to try and identify any particular features that may have led to their differential methylation. Common DMRs were primarily located in three distinct clusters, one on chromosome 5, one on chromosome 17 and one on chromosome 19 (Figure 5.14 B). The location of common DMRs with respect to coding/non-coding regions (Figure 5.14 C), repetitive elements (Figure 5.14 D) and CGIs regions (Figure 5.14 E) did not identify any distinctive features. The CpG density of common DMRs was on average 2.8 CpGs per 100bp. Overall, the common DMRs showed no striking features or characteristics that could account for their commonly differentially methylated status. Common DMRs, as present in *Mtrr*$^{+/+}$sperm, could represent DMRs resulting from parental exposure to the *Mtrr*$^{gt}$mutation. Sperm methylation is established *in utero*, therefore *Mtrr*$^{+/+}$sperm were exposed to the maternal (*Mtrr*$^{+/gt}$female) metabolic derangement and reduced methyl group availability during PGC formation and epigenetic reprogramming. Alternatively, common DMRs could be associated with genetic sequence changes that exist between the *Mtrr* mice and C57Bl/6 controls. Their tightly clustered localisation and lack of defining characteristics could lend support to this hypothesis. This is further explored in section 5.2.10.

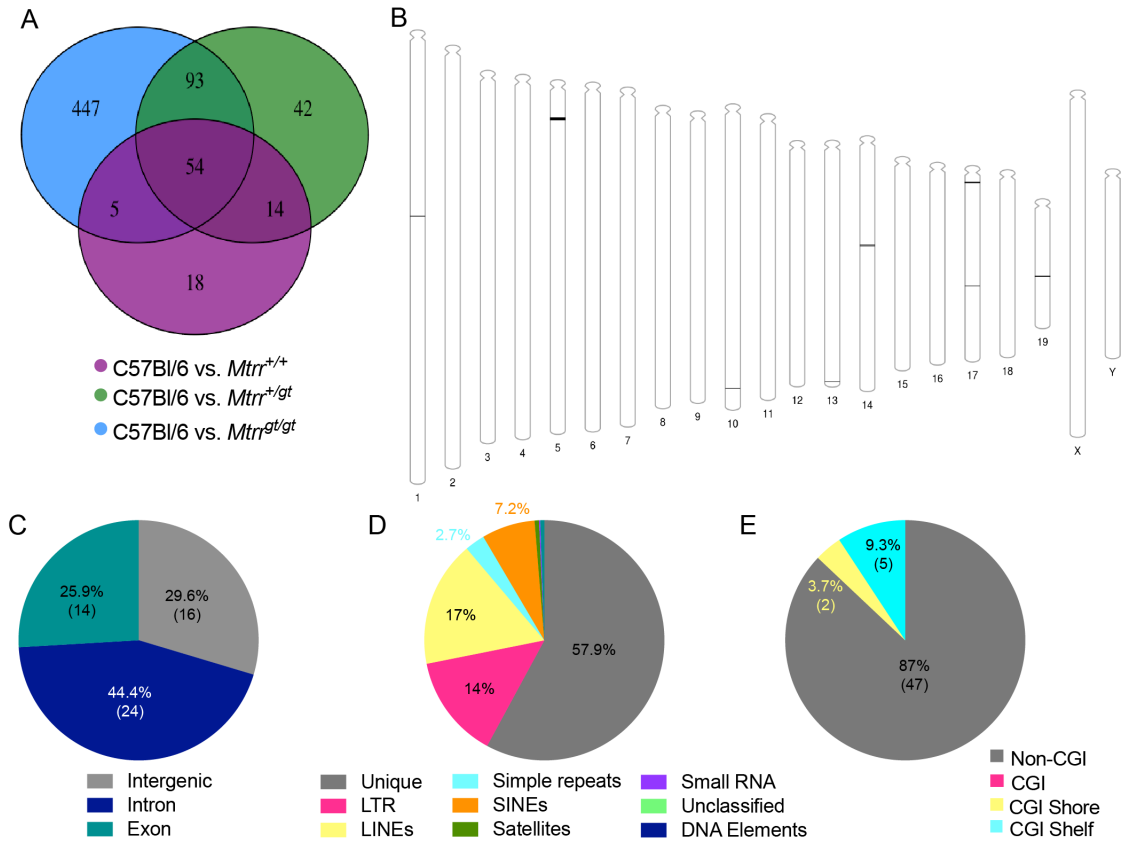



Figure 5.14: **A subset of DMRs were present in sperm from *Mtrr*$^{+/+}$, *Mtrr*$^{+/gt}$ and *Mtrr*$^{gt/gt}$ males.** **(A)** The intersection of DMRs identified in sperm from *Mtrr*$^{+/+}$(purple), *Mtrr*$^{+/gt}$(green) and *Mtrr*$^{gt/gt}$(blue) males. **(B)** A phenogram depicts the chromosomal location of the common DMRs. **(C)** Relative distribution of common DMRs among intergenic regions (grey), exons (teal) and introns (dark blue). **(D** Relative distribution of common DMRs among unique regions (grey), LTRs (pink), LINEs (yellow), simple repeats (pale blue), SINEs (orange), satellites (green), small RNAs (purple), DNA elements (navy blue) and unclassified repeats (lime green). **(E)** Relative distribution of common DMRs among CGIs (pink) and CGI shores (yellow) and CGI shelves (turquoise) or non-CGI regions (grey).

### 5.2.10 Some DMRs associate with underlying genetic changes particulary around the *Mtrr* locus

I wanted to determine if genetic variants could account for the presence of DMRs identified in *Mtrr*$^{+/+}$, *Mtrr*$^{+/gt}$ and *Mtrr*$^{gt/gt}$ males. I compared the sperm DMRs identified using MeDIP-Seq to the SNPs and SVs identified using whole genome sequencing of *Mtrr*$^{gt/gt}$ embryos (Chapter 4). Ideally, whole genome se-

quencing would have been performed on the sperm samples whose methylome was characterised. However, as DMRs were present across multiple individuals within each genotype group, if genetic variants were responsible for the DMRs I identified, they were likely established variants within the *Mtrr*$^{gt}$line rather than *de novo* germline mutations.

Initially, I looked at whether sperm differential DNA methylation coincided directly with SNPs identified in *Mtrr*$^{gt/gt}$embryos. It seems unlikely that a single base pair change, even if directly affecting a cytosine residue, would lead to substantial enough methylation changes for a DMR to be identified. However, the presence of multiple SNPs within a locus could drive differential methylation. I identified that one *Mtrr*$^{+/gt}$DMR (0.5%) and 22 *Mtrr*$^{gt/gt}$DMRs (3.7%) coincided with SNPs (Table 5.2). Some DMRs were associated with numerous SNPs. In total, 65 SNPs fell within *Mtrr*$^{gt/gt}$sperm DMRs (an average 2.9 SNPs per SNP containing DMR) and 4 SNPs were within in the *Mtrr*$^{+/gt}$sperm SNP containing DMR (Table 5.2). Of those 65 SNPs underlying *Mtrr*$^{gt/gt}$DMRs, all were heterozygous, 14 resulted in loss of a cytosine, 19 resulted a gain of cytosine and 32 SNPs did not involve cytosines. I also considered if there were SNPs within the vicinity of DMRs. I identified that 25 sperm DMRs identified in *Mtrr*$^{gt/gt}$males and 2 sperm DMRs identified in *Mtrr*$^{+/gt}$males had SNPs less than 1kb upstream or downstream of the DMR. No DMRs identified in sperm from *Mtrr*$^{+/+}$males were directly associated with SNPs or were within 1kb of a SNP.

I then assessed if DMRs coincided with SVs (Table 5.2). Strikingly, DMRs on chromosome 19 identified in *Mtrr*$^{+/+}$, *Mtrr*$^{+/gt}$and *Mtrr*$^{gt/gt}$sperm, associated with a SV identified in *Mtrr*$^{gt/gt}$embryos (Table 5.2, Figure 5.15 A). This SV on chromosome 19 (36911361-37379467) was a tandem duplication encompassing a region spanning six known genes and was present in all *Mtrr*$^{gt/gt}$embryos sequenced and absent from both C57Bl/6 control embryos (Chapter 4). DMRs identified in *Mtrr*$^{+/+}$, *Mtrr*$^{+/gt}$and *Mtrr*$^{gt/gt}$sperm cluster at the ends of the tandem duplicated region (Figure 5.15 A). Interestingly, differentially methylated regions were located within 10-15 kb of two genes, *Btaf1* and *Ide* which were upregulated ($\sim$ 2-fold) in *Mtrr*$^{gt/gt}$embryos at E10.5 (Chapter 4, Figure 4.9). Other genes at the tandem duplicated region were not misexpressed and inci-

Table 5.2: **Some DMRs coincided with SVs and SNPs**.

| Male Genotype | Total DMRs | DMRs containing SNPs | Total SNPs in DMRs | DMRs associated with SVs |
|---|---|---|---|---|
| *Mtrr*$^{+/+}$ | 91 | 0 | 0 | 14 (15.4%) |
| *Mtrr*$^{+/gt}$ | 203 | 1 (0.5%) | 4 | 23 (11.3%) |
| *Mtrr*$^{gt/gt}$ | 599 | 22 (3.7%) | 65 | 19 (3.2%) |

dentally were located further from the differentially methylated loci. Whether the methylation at these DMRs has a regulatory role in the expression of *Btaf1* and *Ide* should be explored. One might speculate that elevated methylation at this region is a response to suppress transcription resulting from the tandem duplication. Outside of this chromosome 19 locus, no DMRs had an underlying SV or were within 1 kb of a SV identified in *Mtrr*$^{gt/gt}$embryos.

During validation of sperm DMRs identified in *Mtrr*$^{gt/gt}$males by bisulfite pyrosequencing, I identified a genetic variant associated with DMR 60 (chr10:122886000-122886500). When the amplified PCR product was run on an electrophoresis gel, the band size was slightly larger for *Mtrr*$^{gt/gt}$samples than for C57Bl/6 controls (Figure 5.15 B). Sanger sequencing of *Mtrr*$^{gt/gt}$tissue revealed an increase in the number of occurrences a short repetitive sequence (CTCGGCTC) in the *Mtrr*$^{gt/gt}$samples versus C57Bl/6 controls (n=14 repeats in C57Bl/6 liver and n=18 repeats in *Mtrr*$^{gt/gt}$liver, Figure 5.15 C). Curiously, this variant was not identified in whole genome sequencing experiment, perhaps due to poor coverage at this region.

Next, I examined in more detail genetic variation at the *Mtrr* locus with respect to sperm DMRs. I identified that a ~20 Mb region around the *Mtrr* gene had a 129/P2 genetic background and had a high frequency of SNPs and SVs (Chapter 4). I investigated if this may be responsible for the elevated DMR frequency also observed at the *Mtrr* locus. In total, 402 (67.1%) of the sperm DMRs identified in $Mtrr^{gt/gt}$males and 36 (17.7%) of the sperm DMRs identified in $Mtrr^{+/gt}$males were located on chromosome 13 at the *Mtrr* locus. Notably, the distribution of $Mtrr^{gt/gt}$sperm DMRs about the *Mtrr* gene mirrored almost exactly the distribution of SNPs, with a similar asymmetry about the *Mtrr* gene (Figure 5.15 D; Chapter 4, Figure 4.7). Unremarkably, the one $Mtrr^{+/gt}$and 22 $Mtrr^{gt/gt}$sperm DMRs associated with SNPs were on chromosome 13 within the *Mtrr* locus (Table 5.2, Figure 5.15 D). Furthermore, I sought to identify if the 129/P2 background DNA flanking the *Mtrr* gene was responsible for differential DNA methylation. I compared sperm DMRs identified at the *Mtrr* locus to the 129/P2 SNPs identified at the *Mtrr* locus in $Mtrr^{gt/gt}$embryos (Chapter 4). In total, six $Mtrr^{gt/gt}$DMRs coincided with at least one 129/P2 SNP. This suggests that the 129/P2 genetic background may contribute to the differential methylation at a small number of loci near the *Mtrr* gene. However, a large proportion of DMRs at the *Mtrr* locus were not associated with SNPs identified in $Mtrr^{gt/gt}$embryos. These may associate with SVs. I was unable to confirm if DMRs at the *Mtrr* locus associated with SVs as this region was masked during SV analysis. This analysis should be performed to get a full understanding of the role of genetic variants on DNA methylation at the *Mtrr locus*.

Due to the high percentage of sperm DMRs identified in $Mtrr^{gt/gt}$males at the *Mtrr* locus, I speculated that these may skew the characteristics of the $Mtrr^{gt/gt}$sperm DMRs. Therefore I re-characterised the $Mtrr^{gt/gt}$sperm DMRs excluding all DMRs at the *Mtrr* locus. This resulted in the characteristics of the $Mtrr^{gt/gt}$DMRs, in terms of the proportion that were hypomethylated versus hypermethylated and their genomic localisations, becoming more similar to those observed for DMRs identified in $Mtrr^{+/gt}$male sperm (Figure 5.15 E-H, Figure 5.11). Specifically $Mtrr^{gt/gt}$sperm DMRs no longer showed enrichment at LTRs but they did appear to be overrepresented at CGIs (Figure 5.15 E-H). I also characterised the DMRs identified at the *Mtrr* locus in $Mtrr^{gt/gt}$sperm

themselves. The *Mtrr* locus sperm DMRs tended to be hypomethylated (326 of 402 (81.1%)), intergenic and appeared depleted at CGIs.

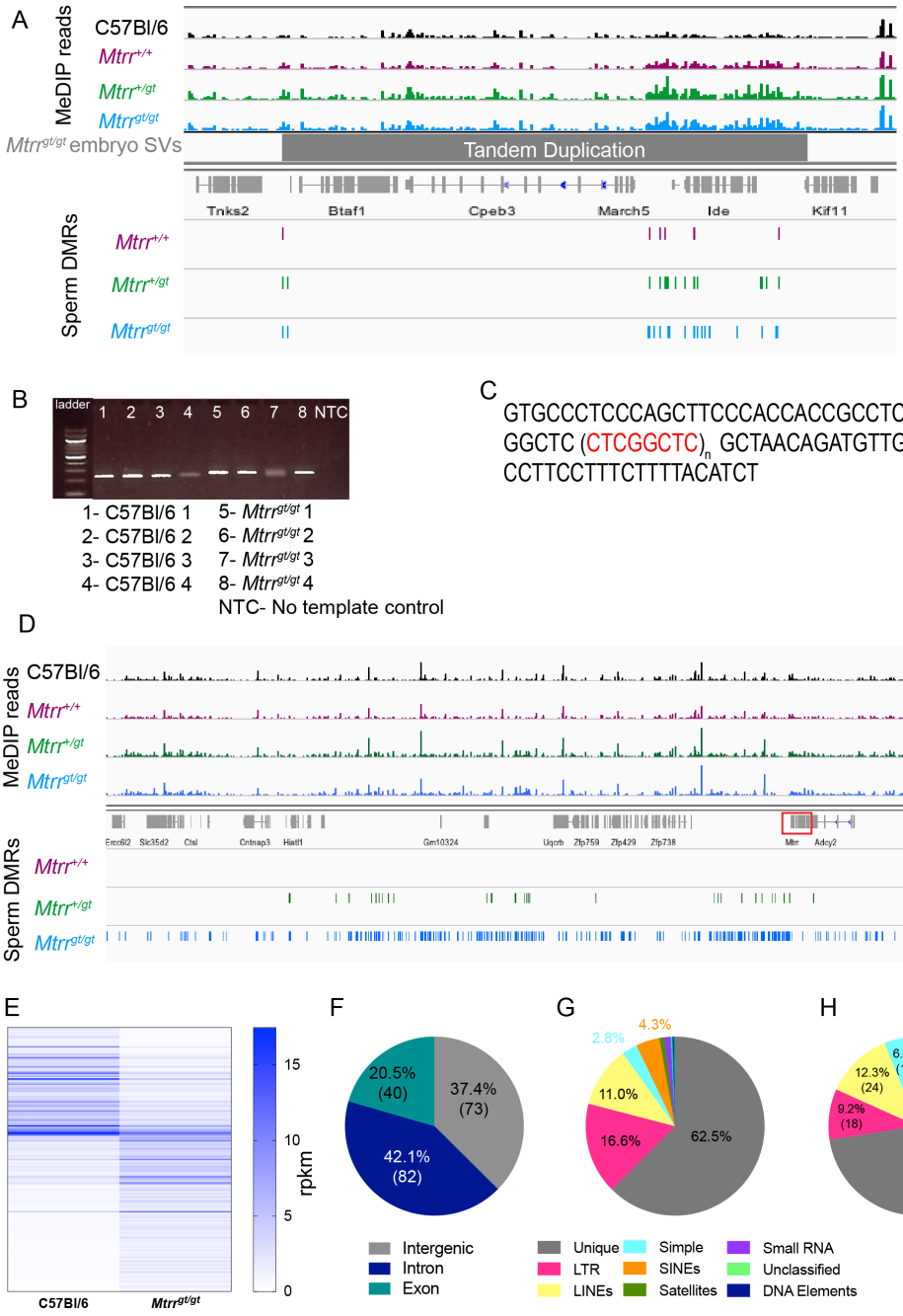


Figure 5.15: **Some DMRs in *Mtrr*$^{+/+}$, *Mtrr*$^{+/gt}$ and *Mtrr*$^{gt/gt}$ sperm are associated with underlying genetic changes.** **(A)** A snapshot from IGV highlights the cluster of DMRs on chromosome 19. MeDIP-seq reads and DMRs are shown for C57Bl/6 (black), *Mtrr*$^{+/+}$(purple), *Mtrr*$^{+/gt}$(green) and *Mtrr*$^{gt/gt}$(blue) sperm. **(B)** Electrophoresis gel of the PCR product of DMR 60, showing the slightly larger band size in *Mtrr*$^{gt/gt}$samples (lanes 5-8) than in C57Bl/6 controls (lanes 1-4). A 1 kb DNA ladder and no template control (NTC) are included. **(C)** The DNA sequence of the PCR product in (B) is shown. "n" denotes the number times the repeat (red) appears in the DNA sequence. **(D)** A snapshot from IGV highlights DMRs on chromosome 13 at the *Mtrr* locus. MeDIP-seq reads and DMRs are shown for C57Bl/6 (black), *Mtrr*$^{+/+}$(purple), *Mtrr*$^{+/gt}$(green) and *Mtrr*$^{gt/gt}$(blue) sperm. The *Mtrr* gene is show in a red box. **(E)** A heat-map plotting reads per kilobase million (rpkm) showing hypomethylated and hypermethylated DMRs, with DMRs at the *Mtrr* locus excluded. **(F)** Relative distribution of *Mtrr*$^{gt/gt}$DMRs, with DMRs at the *Mtrr* locus excluded, among intergenic regions (grey) exons (teal) and introns (dark blue). **(G)** Relative distribution of *Mtrr*$^{gt/gt}$DMRs, with DMRs at the *Mtrr* locus excluded, among unique regions (grey), LTRs (pink), LINEs (yellow), simple repeats (pale blue), SINEs (orange), satellites (green), small RNAs (purple), DNA elements (navy blue) and unclassified repeats (lime green). **(H)** Relative distribution of *Mtrr*$^{gt/gt}$DMRs, with DMRs at the *Mtrr* locus excluded, among CGIs (pink) and CGI shores (yellow) and CGI shelves (turquoise) or non-CGI regions (grey).

I believe the DNA methylation changes identified in sperm on chromosome 13 and 19 are a result of underlying genetic variants. These can be described as secondary epimutations (McCarrey et al., 2016). Beyond chromosomes 19 and 13, only a handful of DMRs including three DMRs identified in sperm from *Mtrr*$^{gt/gt}$males and one DMR identified in sperm from *Mtrr*$^{+/gt}$males associated with SVs. A total of 76 DMRs identified in sperm from *Mtrr*$^{+/+}$males, 142 DMRs identified in sperm from *Mtrr*$^{+/gt}$males, and 174 DMRs identified in sperm from *Mtrr*$^{gt/gt}$males were not directly associated with underlying genetic variants. Particularly intriguingly are the clusters of common DMRs on chromosome 5 and chromosome 17 that do not associate with SVs or SNPs. However, I cannot completely rule out that genetic variants not detected via whole genome sequencing of *Mtrr*$^{gt/gt}$embryos underlie these DMRs. Overall, the data presented here suggests that abnormal folate metabolism, or parental exposure to abnormal folate metabolism, results in DNA methylation changes in sperm in the absence of underlying genetic changes. This is the first evidence to support the hypothesis that epigenetic information, independent of DNA base

sequence, could be inherited and therefore contribute to the transgenerational inheritance of phenotypes observed in the *Mtrr*$^{gt}$model.

## 5.3 Discussion

Here, I have shown that global DNA methylation and hydroxymethylation levels are normal in sperm of *Mtrr*$^{gt}$males. However, I identified a number of differentially methylated regions in sperm from *Mtrr*$^{+/+}$, *Mtrr*$^{+/gt}$and *Mtrr*$^{gt/gt}$males versus C57Bl/6 controls. Both hypomethylated and hypermethylated DMRs were identified. These were commonly found in clusters and generally were associated with regions of low CpG density. At total of 54 DMRs were common to sperm from *Mtrr*$^{+/+}$, *Mtrr*$^{+/gt}$and *Mtrr*$^{gt/gt}$males. Some common DMRs were associated with a tandem duplication identified in *Mtrr*$^{gt/gt}$embryos on chromosome 19. A large number of DMRs identified in *Mtrr*$^{gt/gt}$sperm, and a smaller percentage identified in *Mtrr*$^{+/gt}$sperm, were located at the *Mtrr* locus. Some DMRs at the *Mtrr* locus were associated with SNPs. Apart from a few exceptional loci, all other DMRs identified in sperm from *Mtrr*$^{+/+}$, *Mtrr*$^{+/gt}$and *Mtrr*$^{gt/gt}$males were not associated with any known underlying genetic variants. Overall, the data presented here demonstrate that the *Mtrr*$^{gt}$mutation, or parental exposure to it, results in DNA methylation changes in sperm.

However, the DMR identification and subsequent analysis was not without limitations. Firstly, differential methylation analysis was performed in a pairwise fashion. The methylation profiles of *Mtrr*$^{gt/gt}$, *Mtrr*$^{+/gt}$and *Mtrr*$^{+/+}$sperm were independently compared to a single C57Bl/6 control group. This may have increased the likelihood of errors within the analysis as the comparisons were no longer independent but correlated based on the shared C57Bl/6 control group (Kuan and Chiang, 2012; Howard et al., 2018). No correction for multiple testing was performed, as this is not a feature provided in the MEDIPS analysis package. A post-hoc adjustment of p-values to take into account multiple testing could be performed, but further discussion with a statistician would be required to confirm if this is appropriate. This may alter the number of DMRs identified. Secondly, I was unable to determine if DMRs were statistically over or under represented at certain loci (e.g. introns or CGIs) as a representative background distribution of sperm methylation was not available as a comparator. To estimate the background distribution of sperm methylation the average number of reads per kilobase million (rpkm) should be calculated across all samples for defined genomic windows. An arbitrary (non-zero) rpkm

threshold would then be used to define methylated regions. The genomic location of these methylated regions with respective to coding/noncoding regions, repetitive elements, CGIs and nucleosome retaining regions could then be assessed. Comparison of the genomic localisations of the sperm DMRs to the background methylation distribution could be used to ascertain enrichment or depletion at certain loci. The comparators used in this study: the whole genome DNA sequence covered by repetitive regions and the nucleosome retention in 10000 randomly selected 500bp windows, did not take into account methylation status and therefore did not serve as accurate comparators. Additionally, I acknowledge that MeDIP-seq has limitations. It is low resolution, with the resolution determined by sequence insert size, and regions of low CpG density may be under-represented (Taiwo et al., 2012). However, as the majority of DMRs could be independently validated by bisulfite pyrosequencing, I believe that it captured true methylation differences in sperm from $Mtrr^{+/+}$, $Mtrr^{+/gt}$and $Mtrr^{gt/gt}$males.

I report that global methylation is not reduced in sperm in the $Mtrr^{gt}$model. This is somewhat surprising given the importance of the folate cycle in cellular methylation reactions and the previously reported global hypomethylation in the $Mtrr^{gt}$model (Padmanabhan et al., 2013). However, despite global hypomethylation being reported in *Mtrr* liver and non-pregnant uterine tissues, in the brains of $Mtrr^{+/gt}$and $Mtrr^{gt/gt}$mice DNA methylation was equivalent to C57Bl/6 controls (Padmanabhan et al., 2013). It was suggested that a sparing mechanism may protect the brain from reduced methyl group availability. A similar scenario could be envisaged for sperm. SAM levels in sperm should be assessed to test this hypothesis. Methodological differences also need to be considered. Global methylation in the $Mtrr^{gt}$model was previous measured using methylated DNA quantification kits (Padmanabhan et al., 2013) rather than mass spectrometry which may be considered a more robust approach. Furthermore, I did not identify differential methylation near any imprinted genes in sperm. This is despite imprinted genes showing locus specific dysregulation of DNA methylation in $Mtrr^{+/+}$placentas at E10.5 derived from a $Mtrr^{+/gt}$maternal grandparent (Padmanabhan et al., 2013). Perhaps in sperm, imprinted loci are prioritised to maintain their methylation due to their crucial

role in development. Indeed, in *Mthfr*$^{-/-}$ sperm, imprinted genes also showed no methylation differences as determined by quantitative analysis of methylation PCR (Chan et al., 2010).

Some of the differentially methylated loci I identified were hypermethylated. The presence of hypermethylated loci seems counter-intuitive in a model in which methyl group availability is expected to be reduced. However, hypermethylation has been reported at other loci in *Mtrr*$^{gt/gt}$tissues (Padmanabhan et al., 2013). Additionally, both hypomethylated and hypermethylated loci were reported in sperm of *Mthfr*$^{-/-}$ mice (Chan et al., 2010). The mechanism for the locus-specific dysregulation of DNA methylation is unknown. At many of the DMRs I identified there are large differences in methylation across multiple CpG sites. The differences in methylation between C57Bl/6 and *Mtrr* sperm are greater than those previously reported between groups at DMRs resulting from dietary perturbations (Radford et al., 2014; Wei et al., 2014). I believe this likely reflects the genetic basis of the *Mtrr*$^{gt}$model and direct link between folate metabolism and DNA methylation. However, the number of DMRs identified in sperm from *Mtrr*$^{+/gt}$and *Mtrr*$^{gt/gt}$males, once the *Mtrr* locus is excluded, is approximately equivalent to that observed in sperm in other models of intergenerational or transgenerational inheritance (Radford et al., 2014; Wei et al., 2014; Beck et al., 2017). However, I identified many more DMRs than previously reported in *Mthfr*$^{-/-}$ or folate-deficient sperm. Additionally, no overlap was observed between the DMRs identified in *Mtrr* sperm and those identified in folate-deficient or *Mthfr*$^{-/-}$ sperm (Lambrot et al., 2013; Chan et al., 2010). This may reflect the differing approaches used for DMR identification, MeDIP-Seq (this study), MeDIP-Chip (Lambrot et al., 2013) and restriction landmark genomic scanning (RLGS) (Chan et al., 2010). Alternatively, it may reflect the underlying differences in models of folate deficiency versus disrupted folate metabolism.

It has been reported that there is a high level of epigenetic variation between individuals including within inbred mouse strains. For example, in mice fed a protein restricted diet from weaning until sexual maturity, whether control versus protein-restricted mice were littermates or not had a greater effect on methylation differences than diet (Carone et al., 2010). However, the DNA

methylation differences that I identified were highly conserved across individuals, regardless of littermate status. The low inter-individual variability was clearly observed in the bisulfite pyrosequencing data, with the methylation of individuals of the same genotype tightly clustered (Figures 5.8 and 5.9). There were two exceptions to this: at DMR 60 in sperm from *Mtrr*$^{+/gt}$and *Mtrr*$^{gt/gt}$mice and at DMR 274 in sperm of C57Bl/6 mice (Figure 5.8 L,M, Figure 5.9 A). These loci have methylation patterns similar to metastable epialleles, defined as ERV-associated elements with variable methylation between individuals (Kazachenka et al., 2018). However neither DMR 60 nor DMR 274 are associated with ERV elements. I did identify that DMR 60 is associated with a variant in *Mtrr*$^{gt/gt}$mice which causes an increase in the number of a short repeat sequence versus C57Bl/6 controls (Figure 5.15 B,C). I postulate that the varying methylation between *Mtrr*$^{gt/gt}$individuals at DMR 60 may reflect variability in the number of repeats present in the *Mtrr*$^{gt/gt}$mice. DMR 274 was not associated with a SNP or SV identified in *Mtrr*$^{gt/gt}$embryos or any repetitive elements therefore I am unsure of the cause of the variable methylation in C57Bl/6 individuals. Sanger sequencing of the DMR 274 region may help identify an underlying genetic cause. Studies have also shown that tandem repeat regions and rDNA copy number variations are associated with variable methylation between inbred individuals (Shea et al., 2015). Indeed, in sperm from *Mtrr*$^{+/gt}$and *Mtrr*$^{gt/gt}$males a DMR is present at the start of the *Rn45s* rDNA locus. However, rDNA copy number variation was not detected by whole genome sequencing. Overall, genetic variation between individuals may drive variation in the epigenome, although this seems to occur at only isolated loci in *Mtrr* sperm.

Interpreting the functional relevance of sperm DMRs is challenging. The DMRs identified did not overtly associate with genes involved in biological processes implicated in the phenotypes observed in the *Mtrr*$^{gt}$model. While DNA methylation is known to regulate gene expression, sperm are generally considered transcriptionally silent (Ward and Zalensky, 1996). Therefore, DMRs may not be functionally relevant until post-fertilisation. Some *Mtrr* sperm DMRs I identified fall in introns, which may be associated intronic enhancers (Hoivik et al., 2011). About 20% of *Mtrr*$^{+/+}$and *Mtrr*$^{+/gt}$DMRs fall

in exons. The role of exon methylation remains elusive, although it may play a role in splicing (Lev Maor et al., 2015). DMRs may have additional functions beyond direct regulation of gene expression either within the sperm or the zygote, provided they escape reprogramming. DNA methylation at DMRs may influence other chromatin factors, act as binding sites for regulatory molecules or influence nucleosome or protamine positioning (Tirado-Magallanes et al., 2017). Indeed, *Mtrr* DMRs seem to be preferentially associated with nucleosome retaining regions (Erkek et al., 2013). Nucleosomes are known to be retained at developmentally important genes (Erkek et al., 2013). It would be interesting to explore histone marks and histone variant composition at nucleosome retaining loci associated with DMRs. Modifications, such as H3K27me3, in retained nucleosomes are associated with gene repression in early embryos (Erkek et al., 2013). This analysis could help elucidate a role for DMRs, and their chromatin context, in embryo development.

Overall, I have identified DNA methylation changes in sperm of $Mtrr^{+/+}$, $Mtrr^{+/gt}$and $Mtrr^{gt/gt}$males. The $Mtrr^{+/gt}$males represent the maternal grandfather in the transgenerational pedigree. Therefore differential methylation present in these males may be a factor in the transgenerational inheritance of congenital abnormalities observed in the $Mtrr^{gt}$model.

# Chapter 6

# Determining if differential DNA methylation is inherited over multiple generations

The introduction of this chapter contains elements published in the following textbook chapter and review: Blake et al. (2018) and Blake and Watson (2016), Permission was granted by the publisher for their inclusion in this thesis.

Western blots were performed by Billy Yung (Centre for Trophoblast Research, Cambridge, UK).

## 6.1 Introduction

DNA methylation has been investigated as a mechanism responsible for the inheritance of phenotypes in many transgenerational and intergenerational inheritance paradigms. Indeed, germ line DNA methylation patterns are susceptible to various environmental insults (Radford et al., 2014; Beck et al., 2017; Gapp et al., 2016; Wei et al., 2014). However, epigenetic reprogramming stands as an obstacle to direct inheritance of DNA methylation patterns. Two waves of reprogramming, the first in the post-fertilisation embryo and the second in the developing primordial germ cells, involve widespread DNA demethylation (Reik and Surani, 2015; Lee et al., 2014; Marcho et al., 2015).

However, loci have been identified that escape epigenetic reprogramming in the germline and the blastocyst. Initial studies demonstrated that IAP elements are resistant to demethylation in the preimplantation zygote (Lane et al., 2003). Subsequent genome-wide approaches revealed further loci that are resistant to the post-fertilisation reprogramming wave including CGIs with intermediate methylation levels (Smallwood et al., 2011), non-imprinted promoters primarily of genes expressed in the male germline (Borgel et al., 2010) and regions identified to be differentially methylated between sperm and oocytes (Kobayashi et al., 2012). Furthermore, a large number of loci that are resistant to reprogramming in PGCs have been identified (Hackett et al., 2013). The vast majority of these resistant loci occur at repeat elements, predominately IAPs (Hackett et al., 2013). However, over 200 single copy loci, often near IAP elements or telomeric regions, were PGC reprogramming resistant (Hackett et al., 2013). These loci offer exciting possibilities for direct inheritance of epigenetic information between generations.

Indeed, in some studies differential methylation patterns that were identified in sperm have been successfully identified in offspring tissues. For example, a pre-diabetic mouse model, induced by high fat diet and streptozotocin exposure, is characterised by inheritance of impaired glucose tolerance and insulin resistance up to the F2 generation (Wei et al., 2014). A substantial proportion of regions identified as differentially methylated in the pre-diabetic male sperm were also differentially methylated in pancreatic islets of the F1 generation offspring (Wei et al., 2014). Moreover, traumatic stress in early

post-natal life led to altered methylation of the glucocorticoid receptor (GR) promoter in sperm of stressed males, although the magnitude of change was small (less than 10%) (Gapp et al., 2016). Their offspring had behavioural phenotypes and importantly also showed aberrant methylation of the GR receptor in the hippocampus (Gapp et al., 2016). This might suggest inheritance of DNA methylation via the germline is possible in certain circumstances. However, these studies did not assess if DNA methylation changes identified in the father's sperm were present in grandprogeny to test a true transgenerational effect.

In many other models in which differential DNA methylation has been identified in sperm, tracking the inheritance of DNA methylation in subsequent generations has proved challenging. *In utero* undernourishment of mice (F1) led to intergenerational inheritance of metabolic phenotypes and locus-specific alterations in the F1 sperm methylome (Radford et al., 2014; Jimenez-Chillaron et al., 2009). However, differential DNA methylation was not maintained in somatic tissues (brain and liver) of the F2 generation (Radford et al., 2014). Intriguingly in these mice, despite the loss of differential DNA methylation, the expression of genes near some DMRs was perturbed in the F2 offspring (Radford et al., 2014). Ancestral exposure to vinclozolin in rats was associated with widespread alterations in DNA methylation patterns in mature sperm of F3 generation males (Beck et al., 2017). However, these were shown to be entirely distinct from the DNA methylation changes identified in F1 sperm (Beck et al., 2017). This argues against direct inheritance and suggests rather differential methylation is re-established in F3 tissues. Indeed, while it was demonstrated that DNA methylation differences are induced in prospermatogonia exposed to endocrine disruptors *in utero* (F1), the changes seen in the F1 generation are not present in F2 generation prospermatogonia (Iqbal et al., 2015). Even at the $A^{vy}$ epiallele it appears the DNA methylation pattern is not directly inherited. The methylation at the $A^{vy}$ LTR, present in oocytes, is erased in the blastocyst following maternal transmission and then reestablished (Blewitt et al., 2006). Altogether, these studies demonstrate that while DNA methylation differences may be present in sperm following various environmental insults, direct inheritance of these DNA methylation patterns may not occur.

Despite the balance of evidence being against direct inheritance of DNA methylation between generations, there is still the possibility it plays a role under certain circumstances. The direct link between the folate cycle and DNA methylation could make the inheritance of DNA methylation feasible in the *Mtrr*$^{gt}$model. Therefore, I hypothesised that differential DNA methylation present in sperm may be inherited over multiple generations resulting in the transgenerational inheritance of congenital malformations observed in the *Mtrr*$^{gt}$model. I identified 203 DMRs in *Mtrr*$^{+/gt}$male sperm (Chapter 5). *Mtrr*$^{+/gt}$males represent the maternal grandfather in the *Mtrr*$^{gt}$transgenerational pedigree. The F2, F3 and F4 generations derived from a *Mtrr*$^{+/gt}$maternal grandfather have an increased risk of having embryonic congenital malformations (Padmanabhan et al., 2013). In this chapter I aim to establish if the differential DNA methylation present in sperm from *Mtrr*$^{+/gt}$males is inherited and therefore potentially contributes to the inheritance of phenotypes observed in the *Mtrr*$^{gt}$model.

## 6.2 Results

### 6.2.1 Some DMRs were associated with regions that show resistance to reprogramming in the germline or blastocyst

Non-imprinted loci that escape post-fertilisation epigenetic reprogramming have been identified (Smallwood et al., 2011; Kobayashi et al., 2012). Therefore I assessed if DMRs identified in $Mtrr^{+/+}$, $Mtrr^{+/gt}$and $Mtrr^{gt/gt}$male sperm coincided with regions known to show resistance to post-fertilisation DNA demethylation (Kobayashi et al., 2012). Indeed, 40.7% of $Mtrr^{+/+}$sperm DMRs, 47.3% of $Mtrr^{+/gt}$sperm DMRs and 54.3% of $Mtrr^{gt/gt}$sperm DMRs coincided with regions previously identified as resistant to zygotic reprogramming (Table 6.1). This suggests DMRs in the sperm may persist in the developing embryo.

However, for DMRs to be inherited over multiple generations, differential methylation would also have to be resistant to reprogramming in the developing germline. Hence, I determined if DMRs identified in sperm from $Mtrr^{+/+}$, $Mtrr^{+/gt}$and $Mtrr^{gt/gt}$males coincided with genomic regions known to show resistance to germline DNA demethylation (Hackett et al., 2013). Only a small number of DMRs correlated with regions of known resistance to germline reprogramming, including two sperm DMRs identified in $Mtrr^{+/+}$males, five sperm DMRs identified in $Mtrr^{+/gt}$males and 23 sperm DMRs identified in $Mtrr^{gt/gt}$males (Table 6.1). Interestingly, a few sperm DMRs coincided with regions that were resistant to both zygotic and germline reprogramming, including two sperm DMRs identified in $Mtrr^{+/+}$males, four sperm DMRs identified in $Mtrr^{+/gt}$males and 16 sperm DMRs identified in $Mtrr^{gt/gt}$males (Table 6.1). The $Mtrr^{+/gt}$sperm DMRs that are both zygotic and germline reprogramming resistant were located in intergenic regions, at least 10kb from the nearest gene. These DMRs may have the potential to be inherited over multiple generations.

Table 6.1: **DMRs resistant to zygotic and germline reprogramming**.

| Male Genotype | Blastocyst Resistant | Germline Resistant | Blastocyst and Germline Resistant |
|---|---|---|---|
| *Mtrr*$^{+/+}$ | 37 (40.7%) | 2 (2.2%) | 2 (2.2%) |
| *Mtrr*$^{+/gt}$ | 96 (47.3%) | 5 (2.5%) | 4 (2.0%) |
| *Mtrr*$^{gt/gt}$ | 325 (54.3%) | 23 (3.8%) | 16 (2.7%) |

DMRs that are resistant to DNA demethylation in the blastocyst (> 20% methylation in blastocyst, Kobayashi et al., 2012), germline (> 20% methylation in PGCs, Hackett et al., 2013) and both germline and blastocyst (> 20% methylation in PGCs and blastocysts) are shown.

### 6.2.2 Differential methylation at *Mtrr*$^{+/gt}$ male sperm DMRs does not persist in the F1 and F2 generations

In order to determine if differential DNA methylation identified in sperm could be inherited, I assessed DNA methylation at a panel of twelve randomly selected, validated *Mtrr*$^{+/gt}$ sperm DMRs in a range of *Mtrr*$^{+/+}$ F1 and F2 generation offspring tissues of the maternal grandfather pedigree (Table 6.2). Some of the DMRs I assessed in the offspring coincided with regions resistant to DNA demethylation in the blastocyst (Table 6.2).

I first assessed DNA methylation at DMRs identified in *Mtrr*$^{+/gt}$ male sperm in the *Mtrr*$^{+/+}$ F1 generation derived from *Mtrr*$^{+/gt}$ males crossed to C57Bl/6 females. I analysed *Mtrr*$^{+/+}$ embryos and trophoblast at E10.5 and *Mtrr*$^{+/+}$ female adult liver tissue. Three of the F1 adult liver samples were from the *Mtrr*$^{+/+}$ daughters of three *Mtrr*$^{+/gt}$ males in which the sperm DMRs were initially identified. I observed that differential DNA methylation was lost at all DMRs assessed in *Mtrr*$^{+/+}$ F1 embryos and trophoblast at E10.5 and *Mtrr*$^{+/+}$ F1 adult liver (Figure 6.1). Importantly, loss of differential methylation was observed even at loci identified as being resistant to blastocyst reprogramming (Table 6.2). Blastocyst reprogramming resistant loci were identified informationally (Kobayashi et al., 2012) and therefore may not be an accurate representation of the reprogramming resistant loci *in vivo*. Alternatively, methylation may have been retained in the blastocyst at these loci but subsequently lost during post-implantation development.

Table 6.2: **Panel of sperm DMRs identified in *Mtrr*$^{+/gt}$males assessed in F1 and F2 generation tissues.**

| DMR | Chr | Start | Stop | DMR status | C57Bl/6 sperm methylation | *Mtrr*$^{+/gt}$sperm methylation | PGC methylation | Blastocyst methylation |
|---|---|---|---|---|---|---|---|---|
| E28 | 14 | 13512001 | 13512500 | Hypermethylated | 41.0% | 84.4% | <20% | 16.04% |
| E66 | 17 | 6562501 | 6563000 | Hypermethylated | 67.9% | 73.4% | <20% | - |
| E109 | 5 | 28168501 | 28169000 | Hypermethylated | 4.7% | 43.9% | <20% | 7.1% |
| B19 | 19 | 37251001 | 37252500 | Hypermethylated | 10.6% | 24.6% | <20% | 9.2% |
| E115 | 7 | 36770501 | 36772000 | Hypomethylated | 28.6% | 5.2% | <20% | 9.8% |
| E74 | 18 | 47505501 | 47506000 | Hypermethylated | 4.1% | 29.9% | <20% | 6.4% |
| E52 | 17 | 6324501 | 6325500 | Hypermethylated | 17.7% | 72.4% | <20% | 67.6% |
| E112 | 6 | 33270501 | 33271000 | Hypomethylated | 55.7% | 26.4% | <20% | 52.2% |
| 60 | 10 | 122886001 | 122886500 | Hypermethylated | 61.5% | 25.6% | <20% | 37.5% |
| C7 | 19 | 37238001 | 37239000 | Hypermethylated | 21.6% | 50.6% | <20% | 12.2% |
| D87 | 5 | 15671001 | 15671500 | Hypermethylated | 89.6% | 96.3% | <20% | 48.4% |
| A10 | 16 | 18975501 | 18977000 | Hypermethylated | 15.9% | 41.5% | 23.4% | 17.5% |
| B16 | 17 | 39842501 | 39843000 | Hypermethylated | | | <20% | 46.7% |

Blastocyst methylation: Kobayashi et al., 2012. PGC methylation: Hackett et al., 2013.

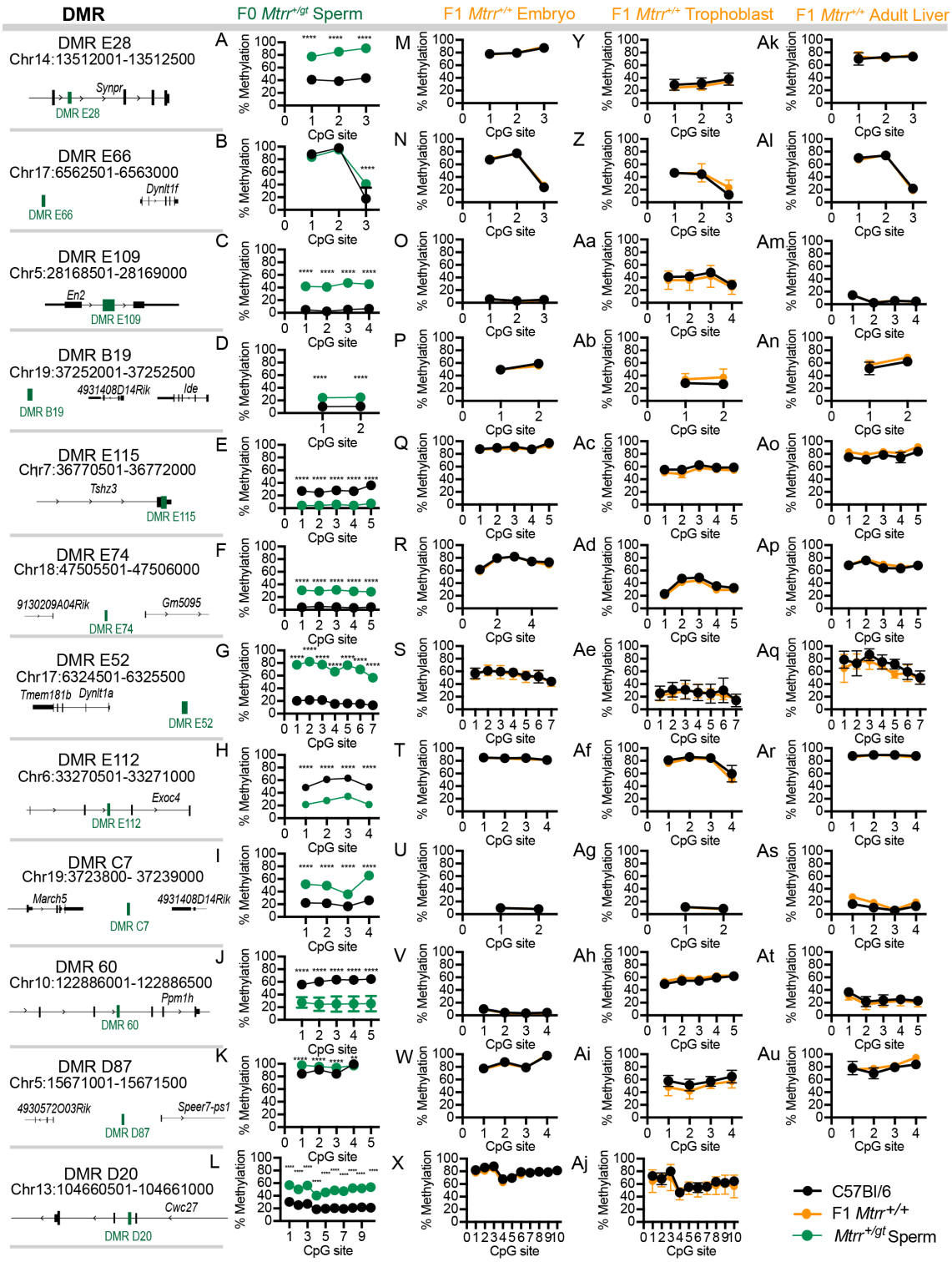

DMR
F0 Mtrr+/gt Sperm
F1 Mtrr+/+ Embryo
F1 Mtrr+/+ Trophoblast
F1 Mtrr+/+ Adult Liver
DMR E28
Chr14:13512001-13512500
Synpr
DMR E28
DMR E66
Chr17:6562501-6563000
Dynlt1f
DMR E66
DMR E109
Chr5:28168501-28169000
En2
DMR E109
DMR B19
Chr19:37252001-37252500
4931408D14Rik
Ide
DMR B19
DMR E115
Chr7:36770501-36772000
Tshz3
DMR E115
DMR E74
Chr18:47505501-47506000
9130209A04Rik
Gm5095
DMR E74
DMR E52
Chr17:6324501-6325500
Tmem181b
Dynlt1a
DMR E52
DMR E112
Chr6:33270501-33271000
Exoc4
DMR E112
DMR C7
Chr19:3723800- 37239000
March5
4931408D14Rik
DMR C7
DMR 60
Chr10:122886001-122886500
Ppm1h
DMR 60
DMR D87
Chr5:15671001-15671500
4930572O03Rik
Speer7-ps1
DMR D87
DMR D20
Chr13:104660501-104661000
Cwc27
DMR D20
% Methylation
CpG site
C57Bl/6
F1 Mtrr+/+
Mtrr+/gt Sperm

Figure 6.1: **Differential methylation at *Mtrr*$^{+/gt}$male sperm DMRs was lost in F1 generation tissues. (A-Au)** Graphical data shows average percentage methylation at individual CpGs for the DMRs illustrated schematically, as determined by bisulfite pyrosequencing. C57Bl/6 tissues (black) are compared to **(A-L)** *Mtrr*$^{+/gt}$sperm (also shown in Chapter 5), **(M-Au)** *Mtrr*$^{+/+}$F1 generation (orange) **(M-X)** embryos at E10.5, **(Y-Aj)** trophoblast at E10.5 and **(Ak-Au)** female adult liver. N=4-8 per group. Two-way ANOVA, with Sidak's multiple comparisons test, performed on mean methylation per CpG site per genotype group. ** $p<0.01$, **** $p<0.0001$.

However, embryonic phenotypes are not observed in the F1 generation derived from a *Mtrr*$^{+/gt}$maternal grandfather (Padmanabhan et al., 2013). I hypothesised that differential DNA methylation may only be apparent in embryos with congenital malformations or their placentas in the F2 generation. Therefore, I assessed DNA methylation at the panel of DMRs (Table 6.2) in *Mtrr*$^{+/+}$F2 generation tissues. The F2 generation was derived from *Mtrr*$^{+/+}$F1 females crossed to C57Bl/6 males. I quantified DNA methylation in embryos, both phenotypically normal and those with congenital malformations, and trophoblast at E10.5. The F2 generation embryos were the grandprogeny of three of the *Mtrr*$^{+/gt}$males in which the sperm DMRs were identified. Overall, differential DNA methylation was not maintained at *Mtrr*$^{+/gt}$male sperm DMRs in *Mtrr*$^{+/+}$F2 generation embryos or trophoblast at E10.5 (Figure 6.2). Isolated CpG sites at a few DMRs did show differential methylation in F2 trophoblast at E10.5 (D87: CpGs 2 and 4, E74: CpGs 3,4 and 5) and in embryos at E10.5 with congenital malformations (E115: CpG 4) (Figure 6.2 E,R,W). At these isolated CpG sites, the DNA methylation differences observed between C57Bl/6 and F2 generation tissues were far smaller than observed at the same CpG sites between C57Bl/6 and *Mtrr*$^{+/gt}$sperm. I believe that small methylation changes at isolated CpG sites are unlikely to have functional relevance. Altogether, this data suggests that differential DNA methylation in sperm is not inherited by *Mtrr*$^{+/+}$offspring generations. Similar normalisation of DNA methylation in offspring tissues at DMRs identified in sperm has been reported in other epigenetic inheritance models (Radford et al., 2014).

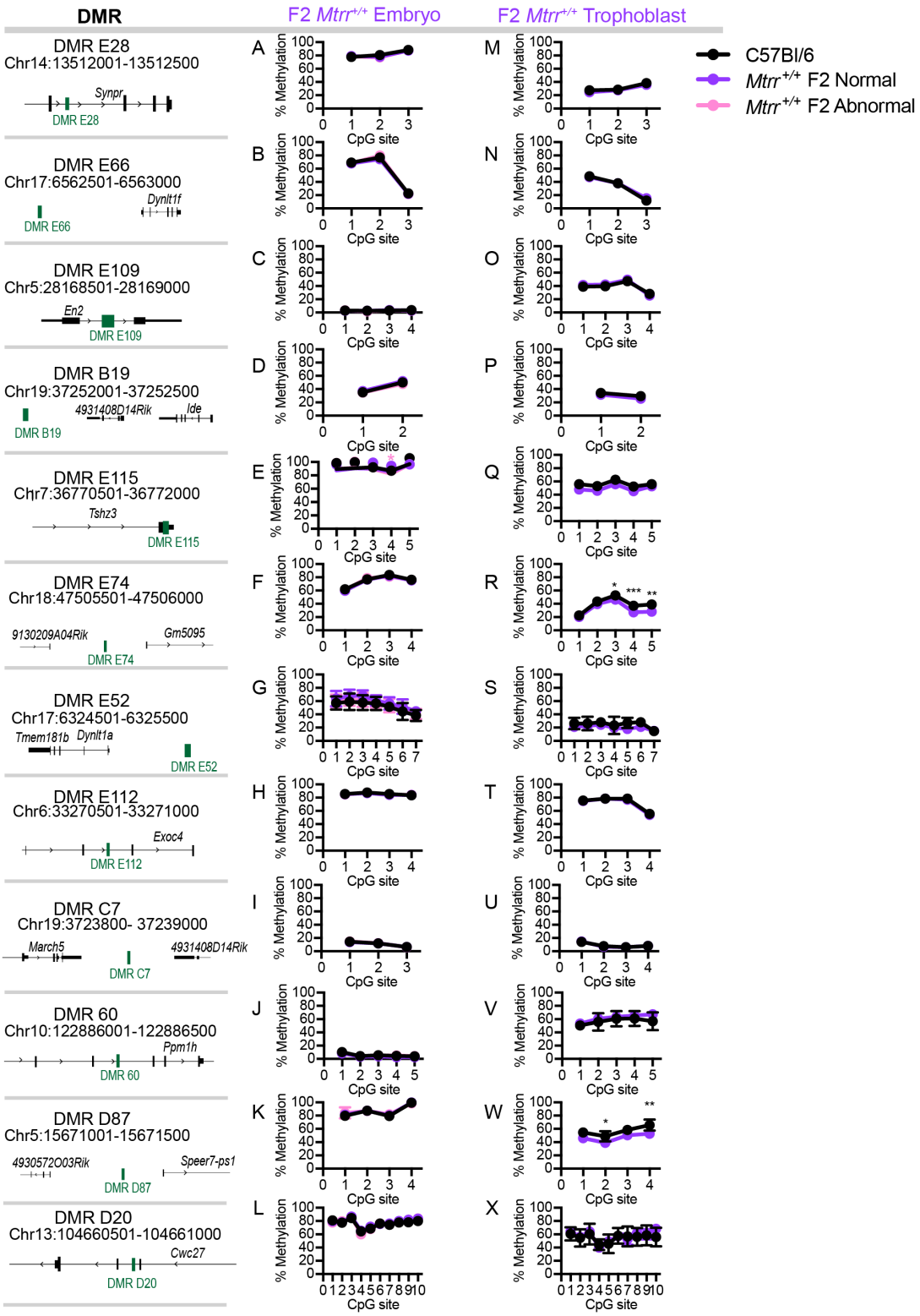

DMR
F2 Mtrr+/+ Embryo
F2 Mtrr+/+ Trophoblast
C57Bl/6
Mtrr+/+ F2 Normal
Mtrr+/+ F2 Abnormal
DMR E28
Chr14:13512001-13512500
Synpr
DMR E28
DMR E66
Chr17:6562501-6563000
Dynlt1f
DMR E66
DMR E109
Chr5:28168501-28169000
En2
DMR E109
DMR B19
Chr19:37252001-37252500
4931408D14Rik
Ide
DMR B19
DMR E115
Chr7:36770501-36772000
Tshz3
DMR E115
DMR E74
Chr18:47505501-47506000
9130209A04Rik
Gm5095
DMR E74
DMR E52
Chr17:6324501-6325500
Tmem181b
Dynlt1a
DMR E52
DMR E112
Chr6:33270501-33271000
Exoc4
DMR E112
DMR C7
Chr19:3723800- 37239000
March5
4931408D14Rik
DMR C7
DMR 60
Chr10:122886001-122886500
Ppm1h
DMR 60
DMR D87
Chr5:15671001-15671500
4930572O03Rik
Speer7-ps1
DMR D87
DMR D20
Chr13:104660501-104661000
Cwc27
DMR D20
% Methylation
CpG site

Figure 6.2: **Differential methylation at *Mtrr*$^{+/gt}$male sperm DMRs was lost in F2 generation tissues. (A-X)** Graphical data shows average percentage methylation at individual CpGs within the DMRs illustrated schematically as determined by bisulfite pyrosequencing. C57Bl/6 tissues (black) are compared to **(A-L)** *Mtrr*$^{+/+}$F2 generation phenotypically normal embryos at E10.5 (normal, purple) or embryos at E10.5 with congenital malformations (abnormal, pink), **(M-X)** *Mtrr*$^{+/+}$F2 generation trophoblast at E10.5 from phenotypically normal conceptuses (purple). N=4-8 per group. Two-way ANOVA, with Sidak's multiple comparisons test, performed on mean methylation per CpG site per genotype group. * $p < 0.05$, ** $p<0.01$, *** $p<0.001$.

### 6.2.3 Differential DNA methylation at *Mtrr*$^{gt/gt}$sperm DMRs is generally lost in *Mtrr*$^{gt/gt}$embryos

The loss of differential methylation observed in F1 generation *Mtrr*$^{+/+}$tissues indicates that aberrant sperm DNA methylation patterns in *Mtrr*$^{+/gt}$males are reprogrammed in the blastocyst. Normal DNA methylation patterns are then established in the *Mtrr*$^{+/+}$F1 and F2 generation tissues, despite general epigenetic instability in some of these tissues (Padmanabhan et al., 2013). I speculated whether establishment of normal DNA methylation patterns in the *Mtrr*$^{+/+}$offspring was concomitant with the loss of the *Mtrr*$^{gt}$allele. If the normalisation of methylation resulted only from the absence of the *Mtrr*$^{gt}$allele, one would have expected DMRs to be present in *Mtrr*$^{+/gt}$or *Mtrr*$^{gt/gt}$tissues. Therefore I assessed the DNA methylation at ten *Mtrr*$^{gt/gt}$ sperm DMRs (that were also present in sperm from *Mtrr*$^{+/gt}$males) in *Mtrr*$^{gt/gt}$phenotypically normal embryos at E10.5. Only two DMRs continued to show differential methylation in *Mtrr*$^{gt/gt}$embryos versus C57Bl/6 controls (DMRs E109 and E52, Figure 6.3 C,F). Methylation at all other *Mtrr*$^{gt/gt}$DMRs was equivalent in *Mtrr*$^{gt/gt}$and C57Bl/6 control embryos (Figure 6.3). I considered what may be behind the persistence of abnormal DNA methylation at DMRs E109 and E52. Firstly I considered their genomic location. DMRs E109 and E52 fall on chromosome 5 and 17, respectively, in DMR clusters identified in *Mtrr*$^{+/+}$, *Mtrr*$^{+/gt}$and *Mtrr*$^{gt/gt}$sperm. However, other DMRs in the chromosome 5 and 17 clusters (D87 and E66, respectively) did not show differential methylation in *Mtrr*$^{gt/gt}$ embryos (Figure 6.3 B,D). Next I considered resistance to reprogramming. DMR E52 is located at a region known to be resistant to reprogramming

in the blastocyst. However, other loci thought to be blastocyst reprogramming resistant (e.g. DMRs E112 and D87) are not differentially methylated in *Mtrr*$^{gt/gt}$embryos (Table 6.2). Interestingly, DMR E109 is not blastocyst reprogramming resistant suggesting differential methylation may be re-established in *Mtrr*$^{gt/gt}$embryos. What is exceptional about the E109 and E52 loci resulting in their differential methylation in *Mtrr*$^{gt/gt}$embryos remains unclear and should be investigated further.

The normalisation of DNA methylation at the majority of DMRs identified in sperm from *Mtrr*$^{gt/gt}$males in *Mtrr*$^{gt/gt}$embryos suggests that the differential methylation present in sperm is not due simply to the presence of the *Mtrr*$^{gt}$allele. Tissue specific factors must influence whether the loci are differentially methylated or not. DMR loci may have differing susceptibilities to low-methyl group availability resulting from the *Mtrr*$^{gt}$allele in sperm and embryos. Overall, this data suggests that at the majority of DMRs, normalisation of DNA methylation in *Mtrr*$^{+/+}$offspring tissues is not purely a result of the loss of the *Mtrr*$^{gt}$allele.

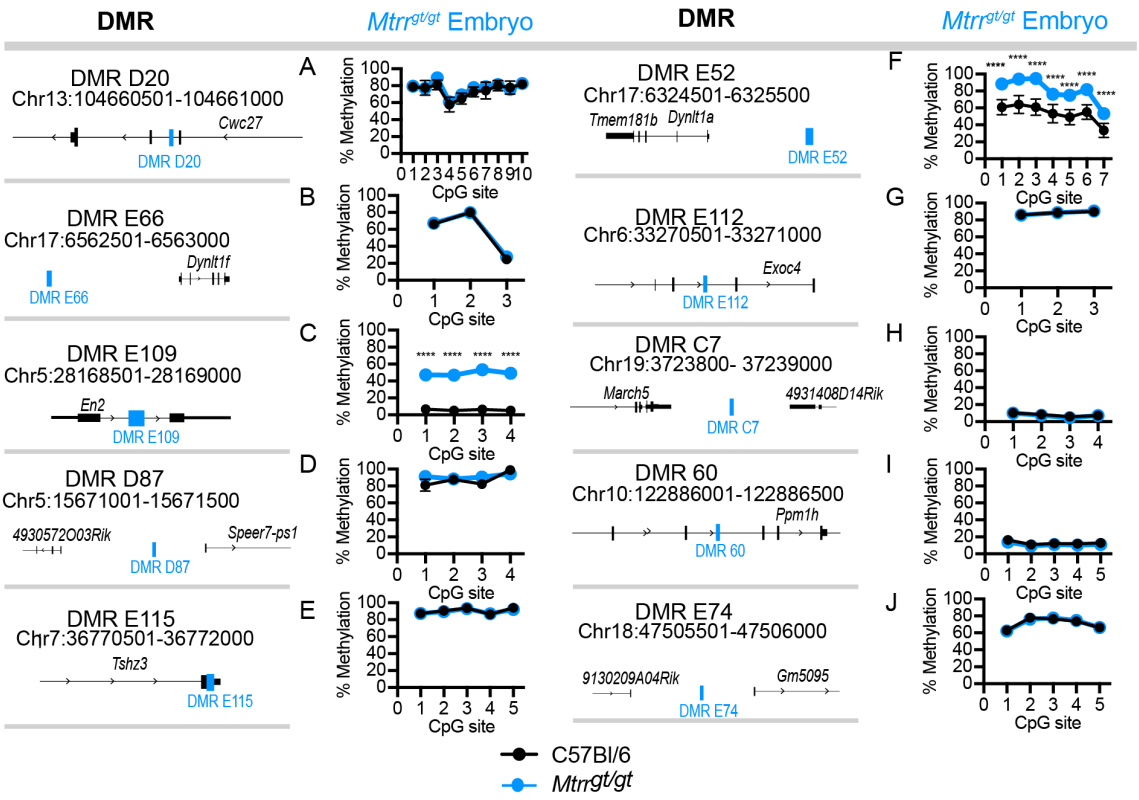



Figure 6.3: **Differential methylation at *Mtrr*$^{gt/gt}$male sperm DMRs was generally not maintained in *Mtrr*$^{gt/gt}$embryos.** **(A-J)** Graphical data shows average percentage methylation at individual CpGs within DMRs identified in *Mtrr*$^{gt/gt}$male sperm as determined by bisulfite pyrosequencing in C57Bl/6 embryos at E10.5 (black) and *Mtrr*$^{gt/gt}$embryos at E10.5 (blue). N=6-8 per genotype. Two-way ANOVA, with Sidak's multiple comparisons test, performed on mean methylation per CpG site per genotype group, **** $p<0.0001$.

### 6.2.4 Expression of genes near *Mtrr*$^{+/gt}$ male sperm DMRs is largely unaffected in F1 and F2 generations

Previous studies have shown that despite loss of differential methylation in offspring generations, expression of genes near sperm DMRs can be altered (Radford et al., 2014). Therefore I used RT-qPCR to assess the expression of a subset of seven genes located near DMRs identified in sperm from *Mtrr*$^{+/gt}$males in offspring tissues. I assessed the expression of *Ide*, *Tshz3*, *Dynlt1a*, *Exoc4*, *March5*, *Ppm1h* and *Cwc27* in *Mtrr*$^{+/+}$F1 and F2 generation embryos and trophoblast at E10.5 and F1 generation adult liver.

I also examined expression of these genes near DMRs in *Mtrr*$^{gt/gt}$phenotypically normal embryos at E10.5 (Figure 6.4 A-G). *Ide* was up-regulated in *Mtrr*$^{gt/gt}$embryos whereas *Exoc4* was down-regulated in *Mtrr*$^{gt/gt}$embryos versus C57Bl/6 controls (Figure 6.4 A,D). Differential methylation was lost in *Mtrr*$^{gt/gt}$embryos at the DMRs associated with these genes (Figure 6.3). This suggests altered expression of these genes was not directly mediated by differential methylation at DMRs identified in *Mtrr*$^{gt/gt}$sperm. The *Ide* gene is on chromosome 19, and was associated with a tandem duplication in *Mtrr*$^{gt/gt}$mice (Chapter 4). This likely explains the *Ide* misexpression. The drivers of *Exoc4* misexpression in *Mtrr*$^{gt/gt}$embryos are unknown but may include embryo specific DNA methylation or histone modification differences associated with the *Mtrr*$^{gt}$allele. The expression of the *Dynlt1a* gene was comparable in *Mtrr*$^{gt/gt}$and C57Bl/6 control embryos (Figure 6.4 C). This was despite persistence of differential methylation at the DMR approximately 6 kb upstream of this gene in *Mtrr*$^{gt/gt}$embryos (Figure 6.3 F). This suggests methylation at this locus does not play a role in regulation of *Dynlt1a* expression in embryos. Unfortunately, RT-qPCR primers could not be optimised to examine the expression *En2*, which was also associated with a DMR showing differential methylation in both *Mtrr*$^{gt/gt}$sperm and embryos at E10.5 (Figure 6.3 C). Altogether, this data suggests *Mtrr*$^{gt/gt}$male sperm DMRs do not generally associate with altered gene expression in *Mtrr*$^{gt/gt}$phenotypically normal embryos at E10.5. However, misexpression of *Exoc4* and *Ide* in *Mtrr*$^{gt/gt}$embryos occurs despite normalisation of DNA methylation at the associated sperm DMRs.

Gene expression was then quantified by RT-qPCR in *Mtrr*$^{+/+}$offspring tissues. Expression of *Ide*, *Tshz3*, *Dynlt1a*, *Exoc4*, *March5*, *Ppm1h* and *Cwc27* was equivalent to C57Bl/6 controls in *Mtrr*$^{+/+}$F1 generation embryos and trophoblast at E10.5 and *Mtrr*$^{+/+}$F1 adult liver (Figure 6.4 H-O). Expression was also equivalent to C57Bl/6 controls in *Mtrr*$^{+/+}$F2 embryos and trophoblast at E10.5 at all bar one gene (Figure 6.4 P-V). *Ppm1h* had decreased expression specifically in *Mtrr*$^{+/+}$F2 generation trophoblast compared to C57Bl/6 controls, while *Ppm1h* expression in F2 generation embryos was equivalent to controls (Figure 6.4 U). The DMR associated with *Ppm1h* (DMR 60) was associated with a change in repeat copy number in *Mtrr*$^{gt/gt}$tissues (Chapter 5). If this genetic

variant is also present in *Mtrr*$^{+/+}$F2 generation tissues, and therefore may affect gene expression, is uncertain and should be explored. Additionally, I examined whether any genes near sperm DMRs identified in *Mtrr*$^{+/gt}$males were present in RNA-seq data from *Mtrr*$^{+/+}$F2 placentas from phenotypically normal conceptuses derived from maternal grandfathers (K. Menelaou and E. Watson, unpublished data). No genes with greater than 1.5 fold change in expression between C57Bl/6 control and *Mtrr*$^{+/+}$F2 placentas at E10.5 were identified and no genes showing any degree of misexpression were associated with DMRs identified in *Mtrr*$^{+/gt}$sperm.

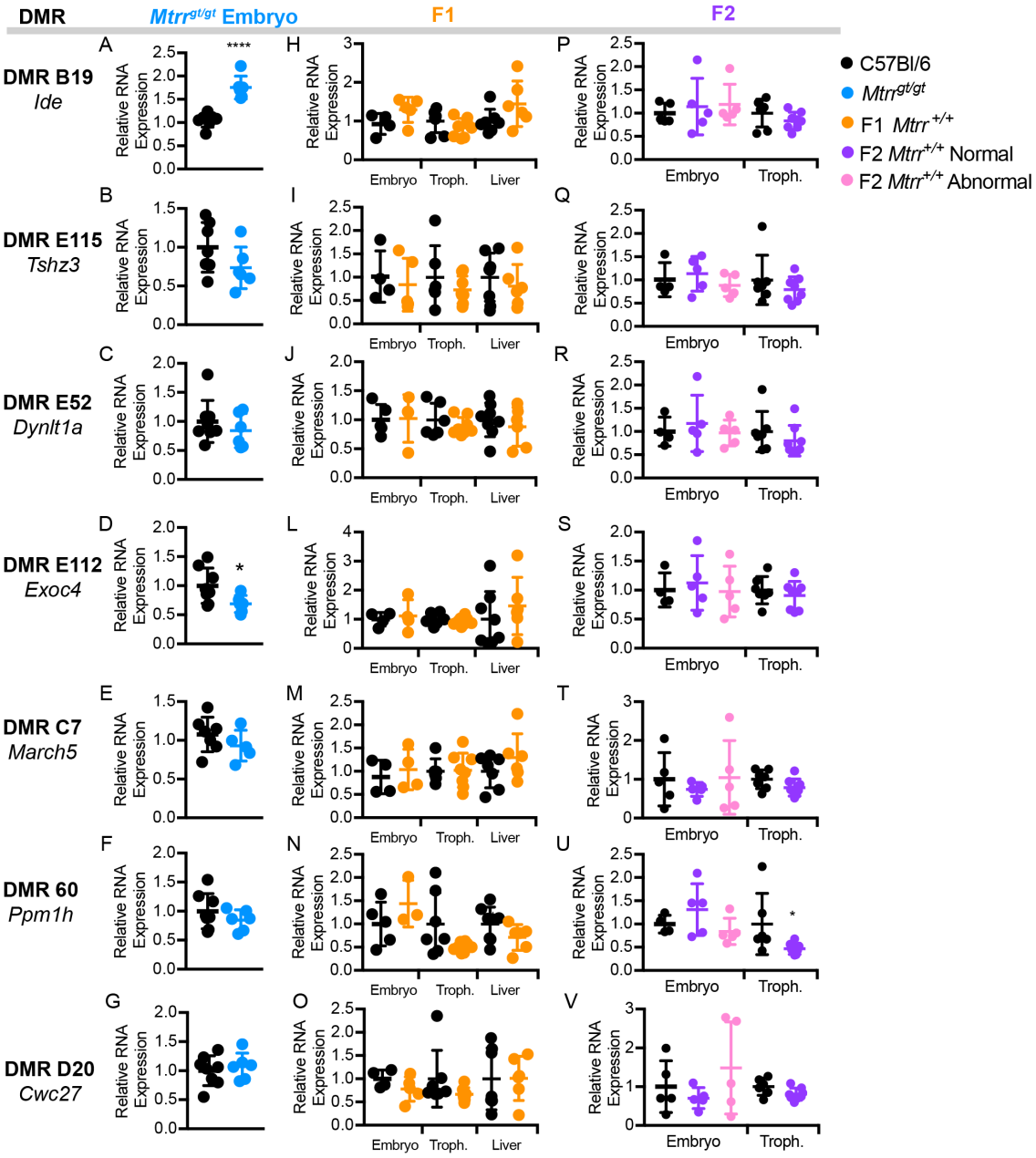



Figure 6.4: **Expression of genes near *Mtrr*$^{+/gt}$male sperm DMRs is largely normal in *Mtrr*$^{gt/gt}$embryos and F1 and F2 generation tissues. (A-V)** Graphs showing relative mRNA expression as determined by RT-qPCR analysis of genes close to DMRs identified in sperm from *Mtrr*$^{+/gt}$males in C57Bl/6 control tissues (black) and **(A-G)** *Mtrr*$^{gt/gt}$phenotypically normal embryos at E10.5 (blue), **(H-O)** *Mtrr*$^{+/+}$F1 generation embryos at E10.5, trophoblast at E10.5 and female adult liver (orange), **(P-V)** *Mtrr*$^{+/+}$F2 generation phenotypically normal embryos and trophoblast at E10.5 (normal, purple) and embryos at E10.5 with congenital malformations (abnormal, pink). Data is plotted as mean ± sd. Tissue from 4-7 individuals per group were assessed. Data is represented as fold change compared to C57Bl/6 controls, which were normalised to 1. Un-paired independent t-tests or one-way ANOVA, with Dunnett's multiple comparisons tests were performed on each data set. * $p<0.05$, **** $p<0.0001$.

However, two DMRs identified in *Mtrr*$^{+/gt}$male sperm, A10 and B16, were associated with gene expression changes in offspring tissues. DMR B16 lies at the start of the *Rn45s* rDNA locus on chromosome 17 (Figure 6.5 A). *Rn45s* is the pre-ribosomal RNA transcript consisting of the 18S, 5.8S, and 28S rRNA genes and additionally a 5′ external transcribed spacer (ETS), an internal transcribed spacer (ITS), and a 3′ ETS (Yoshikawa and Fujii, 2016). This DMR has yet to be validated by bisulfite pyrosequencing in *Mtrr*$^{+/gt}$or *Mtrr*$^{gt/gt}$sperm and this should be performed. Differential DNA methylation at the B16 DMR was not maintained in *Mtrr*$^{+/+}$F1, F2 or F3 generation embryos at E10.5, *Mtrr*$^{+/+}$F1 or F2 generation trophoblast at E10.5 or *Mtrr*$^{+/+}$F1 adult liver (Figure 6.5 B-G). However, expression of the *Rn45s* pre-ribosomal RNA was reduced in F1, F2 and F3 generation embryos at E10.5, regardless of phenotype, compared to C57Bl/6 controls (Figure 6.5 I-K). *Rn45s* expression in *Mtrr*$^{gt/gt}$embryos at E10.5, *Mtrr*$^{+/+}$F1 adult liver and *Mtrr*$^{+/+}$F1 and F2 generation trophoblast at E10.5 was equivalent to C57Bl/6 controls (Figure 6.5 H-J). The *Rn45s* qPCR primers lay within the 5′ ETS sequence thus should have recognised only the unprocessed pre-ribosomal RNA transcript. Overall, *Rn45s* pre-ribosomal RNA was mi-sexpressed in a tissue specific manner despite the loss of differential DNA methylation at the sperm DMR associated with it.

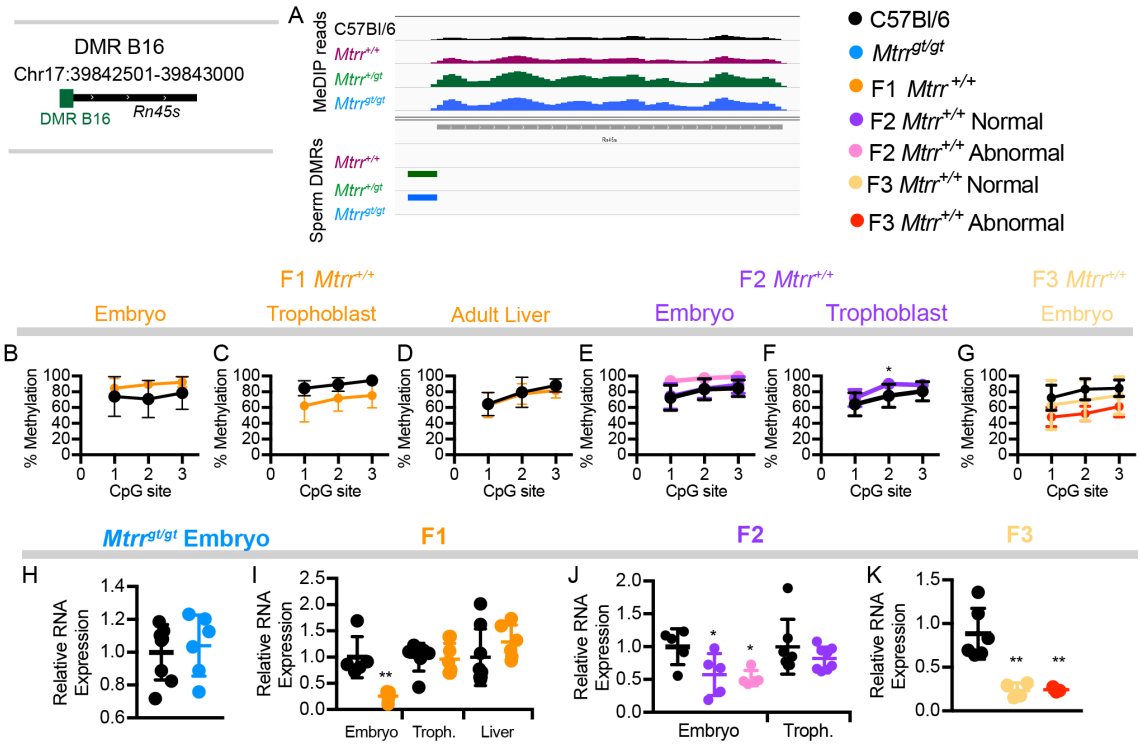



Figure 6.5: **Expression of *Rn45s* is reduced in *Mtrr*$^{+/+}$F1, F2 and F3 embryos.** **(A)** A snapshot from IGV highlights the B16 DMR at the *Rn45s* locus on chromosome 17. MeDIP-seq reads and DMRs are shown for C57Bl/6 (black), *Mtrr*$^{+/+}$(purple), *Mtrr*$^{+/gt}$(green) and *Mtrr*$^{gt/gt}$(blue) sperm. **(B-G)** Graphical data shows average percentage methylation at individual CpGs within the B16 DMR as determined by bisulfite pyrosequencing. C57Bl/6 tissues (black) are compared to **(B-D)** *Mtrr*$^{+/+}$F1 generation (orange) **(B)** phenotypically normal embryos at E10.5 **(C)** trophoblast at E10.5 and **(D)** adult liver, **(E-F)** *Mtrr*$^{+/+}$F2 generation **(E)** embryos at E10.5 that were phenotypically normal (purple) or had congenital malformations (pink) or **(F)** trophoblast at E10.5 from phenotypically normal conceptuses (purple) and **(G)** *Mtrr*$^{+/+}$F3 generation embryos at E10.5 that were phenotypically normal (pale orange) or had congenital malformations (red). N=4-8 per group. **(H-K)** Graphs showing relative expression as determined by RT-qPCR analysis of *Rn45s* in C57Bl/6 control tissues (black) and **(H)** *Mtrr*$^{gt/gt}$phenotypically normal embryos at E10.5 (blue), **(I)** *Mtrr*$^{+/+}$F1 generation embryos at E10.5, trophoblast at E10.5 and female adult liver (orange), **(J)** *Mtrr*$^{+/+}$F2 generation phenotypically normal embryos and trophoblast at E10.5 (purple) and embryos at E10.5 with congenital malformations (pink) and **(K)** *Mtrr*$^{+/+}$F3 generation phenotypically normal embryos at E10.5 (pale orange) and embryos at E10.5 with congenital malformations (red). Data is plotted as mean ± sd. Tissues from 4-7 individuals per group were assessed. Data is represented as fold change compared to C57Bl/6 controls, which were normalised to 1. Un-paired independent t-tests or one-way ANOVA, with Dunnett's multiple comparisons test, were performed. * $p<0.05$, ** $p<0.01$.

DMR A10 was initially identified only in *Mtrr*$^{gt/gt}$sperm via MeDIP-Seq, however a trend towards hypermethylation was also observed in *Mtrr*$^{+/gt}$male sperm in the MeDIP-seq data and via bisulfite pyrosequencing (Figure 6.6 A,B). The high variability in DNA methylation apparent at the A10 DMR, particularly in *Mtrr*$^{+/gt}$male sperm (Figure 6.6 B), might be a technical artefact due to degradation of the sperm DNA but it may reflect true biological variability. Sperm from additional individuals must be collected to verify this finding. Differential DNA methylation at the A10 DMR was not maintained in *Mtrr*$^{+/+}$F1, F2 or F3 generation embryos at E10.5, *Mtrr*$^{+/+}$F1 or F2 generation trophoblast at E10.5, *Mtrr*$^{+/+}$F1 adult liver or *Mtrr*$^{gt/gt}$embryos at E10.5 (Figure 6.6 C-I). DMR A10 is located downstream of the gene *Hira*. *Hira* mRNA expression was upregulated in *Mtrr*$^{gt/gt}$phenotypically normal embryos at E10.5 compared to C57Bl/6 controls (Figure 6.6 J). Conversely, *Hira* showed reduced gene expression compared to C57Bl/6 controls in F2 and F3 generation embryos at 10.5 that were phenotypically normal and that had congenital malformations (Figure 6.6 L,M). The degree of *Hira* down-regulation correlated with the severity of the embryonic phenotype, with a greater reduction in mRNA expression in embryos with congenital malformations (Figure 6.6 L,M). *Hira* mRNA expression in *Mtrr*$^{+/+}$F1 tissues and F2 trophoblast was equivalent to C57Bl/6 controls (Figure 6.6 K,L). To determine if the decreased *Hira* mRNA levels were functionally relevant I assessed HIRA protein levels in *Mtrr*$^{+/+}$F1 and F2 phenotypically normal embryos at E10.5 and *Mtrr*$^{+/+}$F2 generation embryos at E10.5 with congenital malformations compared to C57Bl/6 control embryos. Remarkably, HIRA protein levels were elevated in *Mtrr*$^{+/+}$F2 generation embryos at E10.5 with congenital malformations compared to C57Bl/6 controls (Figure 6.6 N,O). This was despite the pronounced reduction in *Hira* mRNA expression in this tissue.

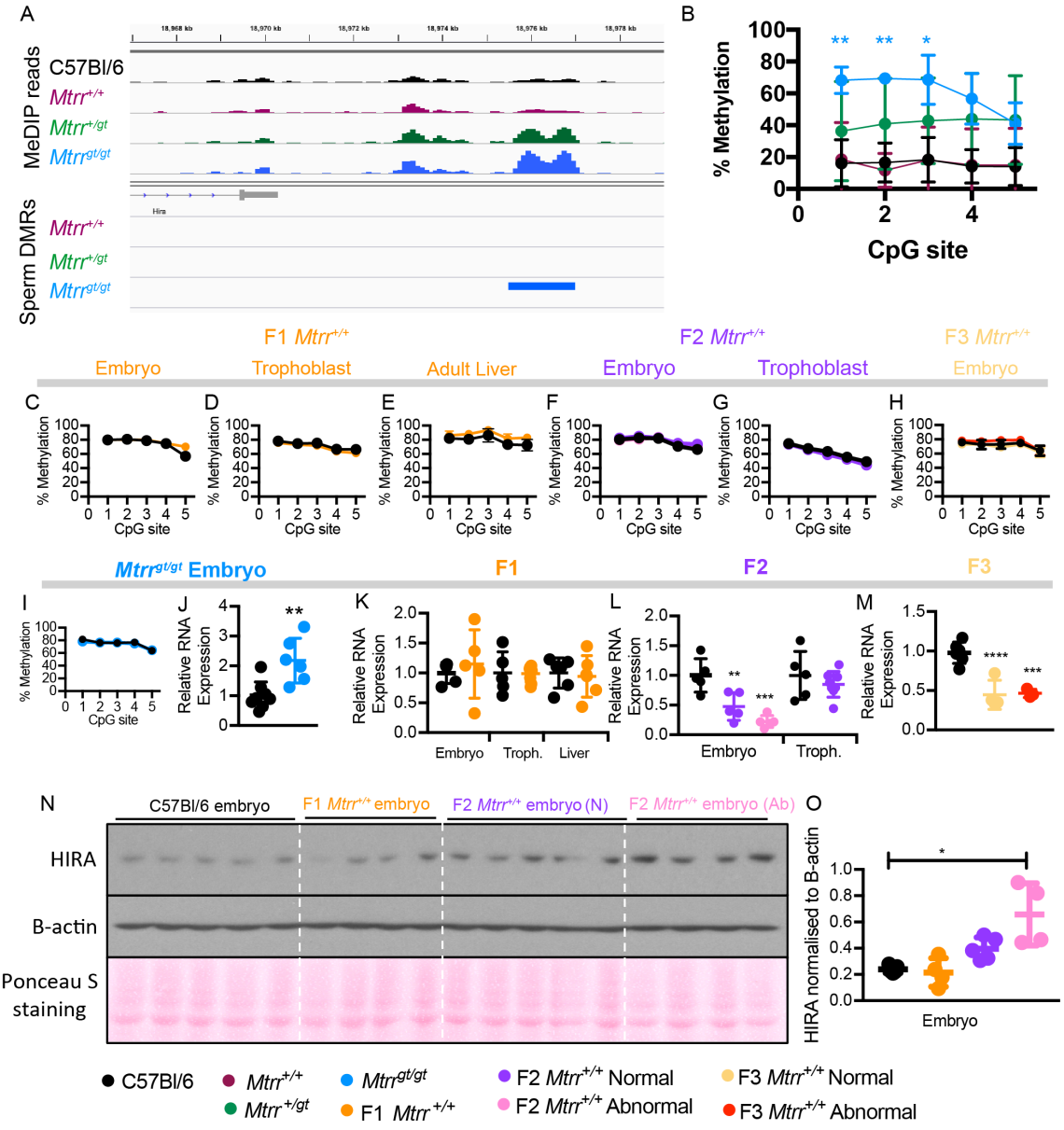

A
MeDIP reads
C57Bl/6
Mtrr+/+
Mtrr+/gt
Mtrrgt/gt
Sperm DMRs
Mtrr+/+
Mtrr+/gt
Mtrrgt/gt
B
% Methylation
CpG site
F1 Mtrr+/+
F2 Mtrr+/+
F3 Mtrr+/+
Embryo
Trophoblast
Adult Liver
Embryo
Trophoblast
Embryo
C
D
E
F
G
H
Mtrrgt/gt Embryo
F1
F2
F3
I
J
K
L
M
Relative RNA Expression
Embryo
Troph.
Liver
Troph.
N
C57Bl/6 embryo
F1 Mtrr+/+ embryo
F2 Mtrr+/+ embryo (N)
F2 Mtrr+/+ embryo (Ab)
HIRA
B-actin
Ponceau S staining
O
HIRA normalised to B-actin
Embryo
C57Bl/6
Mtrr+/+
Mtrr+/gt
Mtrrgt/gt
F1 Mtrr+/+
F2 Mtrr+/+ Normal
F2 Mtrr+/+ Abnormal
F3 Mtrr+/+ Normal
F3 Mtrr+/+ Abnormal

Figure 6.6: **Expression of *Hira* mRNA is reduced in *Mtrr*$^{+/+}$F2 and F3 generation embryos. (A)** A snapshot from IGV highlights the A10 DMR downstream of the *Hira* gene on chromosome 16. MeDIP-seq reads and DMRs are shown for C57Bl/6 (black), *Mtrr*$^{+/+}$(purple), *Mtrr*$^{+/gt}$(green) and *Mtrr*$^{gt/gt}$(blue) sperm. **(B-H)** Graphical data shows average percentage methylation at individual CpGs for the A10 DMR as determined by bisulfite pyrosequencing for C57Bl/6 tissues (black), **(B)** *Mtrr*$^{+/+}$(purple), *Mtrr*$^{+/gt}$(green) and *Mtrr*$^{gt/gt}$(blue) sperm, **(C-E)** *Mtrr*$^{+/+}$F1 generation (orange) **(C)** phenotypically normal embryos at E10.5 **(D)** trophoblast at E10.5 and **(E)** adult liver, **(F-G)** *Mtrr*$^{+/+}$F2 generation **(F)** embryos at E10.5 that were phenotypically normal (purple) or had congenital malformations (pink) or **(G)** trophoblast at E10.5 from phenotypically normal conceptuses (purple), **(H)** *Mtrr*$^{+/+}$F3 generation embryos at E10.5 that were phenotypically normal (pale orange) or had congenital malformations (red) and **(I)** *Mtrr*$^{gt/gt}$phenotypically normal embryos at E10.5 **(J-M)** Graphs showing relative mRNA expression as determined by RT-qPCR analysis of *Hira* in C57Bl/6 control tissues (black) and **(J)** *Mtrr*$^{gt/gt}$phenotypically normal embryos at E10.5 (blue), **(K)** *Mtrr*$^{+/+}$F1 generation embryos at E10.5, trophoblast at E10.5 and female adult liver (orange), **(L)** *Mtrr*$^{+/+}$F2 generation phenotypically normal embryos and trophoblast at E10.5 (purple) and embryos at E10.5 with congenital malformations (pink) and **(M)** *Mtrr*$^{+/+}$F3 generation phenotypically normal embryos at E10.5 (pale orange) and embryos at E10.5 with congenital malformations (red). Data is plotted as mean $\pm$ sd. Data is represented as fold change compared to C57Bl/6 controls, which were normalised to 1. **(N)** Western blot showing HIRA, $\beta$-actin and Ponceau S staining in C57Bl/6 and *Mtrr*$^{+/+}$F1 phenotypically normal embryos at E10.5 and *Mtrr*$^{+/+}$F2 generation embryos at E10.5 that were phenotypically normal (n) or had congenital malformations (Ab). **(O)** Quantification of HIRA staining shown in (N) normalised to $\beta$-actin (arbitrary units) for C57Bl/6 (black) and *Mtrr*$^{+/+}$F1 phenotypically normal embryos at E10.5 (orange) and *Mtrr*$^{+/+}$F2 generation embryos at E10.5 that were phenotypically normal (purple) or had congenital malformations (pink). (J) An un-paired independent t-test, (K-M) one-way ANOVA, with Dunnett's multiple comparisons test and (N) Kruskal-Wallis test with Dunn's multiple comparisons test were performed. * $p<0.05$, ** $p<0.01$, *** $p<0.001$, **** $p<0.0001$. N=4-8 per group.

Overall, these data demonstrate that generally the expression of genes near DMRs identified in sperm from *Mtrr*$^{+/gt}$ males was not disrupted in F1 and F2 generation tissues. This correlated with the normalisation of DNA methylation at the DMRs associated with these genes in these tissues. However, *Hira* and *Rn45s* were misexpressed in a tissue specific manner despite the loss of differential DNA methylation at the DMRs associated with them. This suggests that

germline DNA methylation changes may act as an initiator or demarcator of gene misexpression at certain loci, before being reprogrammed.

## 6.3 Discussion

Here I show that differential methylation identified in sperm from *Mtrr*$^{+/gt}$males was not maintained in subsequent wildtype generations. DMRs identified in sperm from *Mtrr*$^{+/gt}$males did not persist in embryonic, placental and adult tissue of the *Mtrr*$^{+/+}$F1 generation or embryonic and placental tissues of the F2 generation. This indicates that DMRs were likely reprogrammed in the F1 blastocyst. Expression of the majority of genes near regions of differential methylation identified in sperm of *Mtrr*$^{+/gt}$males was also unaffected in *Mtrr*$^{+/+}$F1 and F2 generation tissues. Nonetheless, *Hira* and *Rn45s* were downregulated in *Mtrr*$^{+/+}$F2 and F3 generation embryos at mid-gestation. This gene misepxression occurred despite DNA methylation normalisation at the associated DMRs. Furthermore, HIRA is upregulated at the protein level in *Mtrr*$^{+/+}$F2 generation embryos at E10.5 with congenital malformations. Overall, this data suggests that DNA methylation alone is not a long-term heritable memory of abnormal folate metabolism, but may act as an initiator or demarcator of gene misexpression at some loci.

I used a candidate based approach to examine DNA methylation and transcriptional changes in offspring tissues. Only a small subset of DMRs identified in sperm of *Mtrr*$^{+/gt}$males were investigated in selected offspring tissues. This may have resulted in differential methylation or gene expression changes in the offspring tissues being missed. Indeed, dysregulation of DNA methylation has previously been reported in *Mtrr*$^{+/+}$adult liver and *Mtrr*$^{+/+}$F2 generation placentas at E10.5 (Padmanabhan et al., 2013). Furthermore, the four *Mtrr*$^{+/gt}$sperm DMRs that show resistance to both zygotic and germline reprogramming (Table 6.1) were not examined in the offspring due to difficulty designing pyrosequencing primers. The DNA methylation status of these loci should be assessed in F1 and F2 generation tissues. Additionally, although I did not observe differential methylation at *Mtrr*$^{+/gt}$sperm DMRs or associated gene expression changes in the F1 and F2 embryos and placentas at E10.5, this does not preclude that differences may be observed in other tissues. There is a suggestion that DNA methylation changes may persist in offspring tissues in a highly tissue specific manner. For example, differential DNA methylation at the glucocorticoid receptor (GR) promoter was observed in traumatised F1

generation male sperm. GR promoter methylation changes were also observed in the offspring hippocampus but not pre-frontal cortex (Gapp et al., 2016). This suggests that DNA methylation may not be inherited by all cell types. Embryos at E10.5 are highly heterogeneous and made up of numerous tissues and cell types. Perhaps examining DNA methylation within specific tissues within E10.5 embryos, for example hearts or brains, would reveal methylation changes inherited in a tissue specific manner. Furthermore, genes associated with DMRs were chosen by their genomic proximity to the DMR. Differential methylation may act as a long-range regulator of gene expression (Skinner et al., 2012). Long-range impacts of DMRs on gene expression would be missed with the approach used in this study. Only by adopting combined genome-wide methylome and transcriptome profiling across a range of offspring tissues could these issues be negated.

I observe tissue specific down-regulation of *Hira* and *Rn45s*. If differential DNA methylation at sperm DMRs is to contribute to the gene misexpression it must act very early in development prior to reprogramming as DNA methylation was normal in offspring tissues. It would be interesting to assess the methylation at $Mtrr^{+/gt}$male sperm DMRs in $Mtrr^{+/+}$F1 zygotes, 2 or 4 cell embryos and blastocysts to determine if sperm differential methylation is indeed present and for how long it persists. This would be technically challenging due to the need to genotype F1 zygotes/blastocysts, derived from $Mtrr^{+/gt}$males crossed to C57Bl/6 females, prior to pooling for bisulfite conversion. Exploring if other epigenetic changes are present in sperm will be important to understand how offspring gene expression may be altered despite DNA methylation being reprogrammed. Indeed, the DNA methylation changes I identified in $Mtrr^{+/gt}$sperm were enriched in nucleosome retaining regions suggesting they are inherited in a chromatin context. This might indicate that histone modifications may play a role in mediating the inheritance of information from F0 $Mtrr^{+/gt}$males. Additionally, studies have shown that histone modification profiles in sperm can be altered by folate deficiency (Lambrot et al., 2013). Histone modifications in sperm should be explored further in the $Mtrr^{gt}$model.

Interestingly, HIRA was upregulated at the protein level in $Mtrr^{+/+}$F2 generation embryos at E10.5 with congenital malformations despite downregula-

tion of *Hira* mRNA expression. This may result from *Hira* mRNA transcripts being rapidly turned over due to elevated HIRA translation. Alternatively, HIRA protein may not be being degraded and this may negatively feedback to suppress mRNA expression. Analysis of HIRA ubiquitination levels could help elucidate what is responsible for elevated HIRA protein in the $Mtrr^{+/+}$F2 generation embryos at E10.5 with congenital malformations. HIRA protein levels should also be assessed in $Mtrr^{gt/gt}$and $Mtrr^{+/+}$F3 generation embryos at E10.5 to see if a similar apparent discrepancy between mRNA and protein levels is present.

*Rn45s* encodes the pre-ribosomal RNA. The *Rn45s* locus consists of the 18S, 5.8S, and 28S rRNA genes and additionally a non-transcribed spacer (NTS), a 5′ external transcribed spacer (ETS), an internal transcribed spacer (ITS), and a 3′ ETS (Yoshikawa and Fujii, 2016). *Rn45s* is transcribed by RNA Polymerase I to give rise to a pre-ribosomal RNA transcript (Moss et al., 2019). It is then processed in a tightly regulated manner to give rise to the 28s, 18s and 5.8s mature rRNAs (Grozdanov and Karagyozov, 2002; Yoshikawa and Fujii, 2016). The sequences coding for ribosomal RNAs are present as rDNA repeats on chromosomes 12, 15, 16, 18 and 19. The *Rn45s* locus is annotated on chromosome 17 in the mouse reference genome. Ribosomal DNA (rDNA) loci constitute the nucleolar organiser regions (NORs) responsible for the formation of nucleoli and thus ribosome synthesis within the cell (Moss et al., 2019). The DMR associated with *Rn45s* (B16, chr17:39842500-39843000) is present in $Mtrr^{+/gt}$and $Mtrr^{gt/gt}$sperm and is located over the transcription start site and the promoter (McStay and Grummt, 2008) (Figure 6.5 A). It is not associated with any SNPs or SVs identified in $Mtrr^{gt/gt}$embryos. Copy number variation is associated with this locus across mouse strains, including in C57Bl/6 and 129/P2 mice, and the number of rDNA repeats is also known to be variable between inbred individuals (Shea et al., 2015). CNVs at rDNA loci are associated with epigenetic variation between individuals (Shea et al., 2015). Environmental stress can alter rDNA copy number in yeast (Kwan et al., 2016). Indeed, a high degree of variability in methylation levels between individuals was detected by pyrosequencing at the B16 DMR (Figure 6.5). However, the variability in DNA methylation did not correspond with variability in the expression

levels between individuals. Reduced *Rn45s* transcription might be expected to lead to reduced ribosomal RNA. This could negatively impact ribosomal biogenesis, ribosomal function and consequently translation of proteins (McStay and Grummt, 2008). This could have far reaching consequences for the cell. Ribosomal function assays should be performed to test whether this may be a contributory factor to the phenotypes observed in the *Mtrr*$^{gt}$model (Brar and Weissman, 2015). Remarkably, a DMR associated with *Rn45s* was also identified in sperm of males exposed to *in utero* undernutrition (Radford et al., 2014), which may suggest this locus is sensitive to epigenetic perturbations.

The DMR associated with *Hira*, A10 (chr16: 18975500-1897700), lies 5.2kb downstream of the *Hira* gene, and contains a CTCF binding site and promoter flank. Differential methylation at this DMR was lost in F2 generation embryos suggesting differential methylation at this promoter flank region is not directly driving differential expression in these embryos. No SNPs or SVs identified in *Mtrr*$^{gt/gt}$embryos are found within 10kb of the *Hira* gene. No variants are reported to occur between the 129/P2 and C57Bl/6 mouse strains at this region (Keane et al., 2011). HIRA is a histone chaperone for the histone variant H3.3 (Tang et al., 2006). H3.3 deposition is associated with gene activation (Ray-Gallet et al., 2018; Shindo et al., 2018). In mice, maternally supplied HIRA is required for H3.3 is incorporation into the paternal pronucleus post-fertilisation (Torres-Padilla et al., 2006; Lin et al., 2014). HIRA also has roles in developing oocytes, with depletion leading to reduced fertility and chromatin structure abnormalities (Nashun et al., 2015). The role of HIRA in sperm has not been investigated. HIRA has also been shown to have specific roles in neural development (Li and Jiao, 2017) and cardiac development (Saleh et al., 2018). Importantly, *Hira*$^{-/-}$ mice display congenital malformations similar to those reported in *Mtrr*$^{gt}$mice (Padmanabhan et al., 2013; Roberts et al., 2002). These include abnormal heart development, pericardial oedema, abnormal allantois morphology, failure of chorioallantoic attachment, abnormal neural tube morphology, failed neural tube closure and reduced embryo size (Roberts et al., 2002). Therefore, I speculate that HIRA may play an important role in the inheritance of congenital malformations in the *Mtrr*$^{gt}$model.

Overall, differential DNA methylation identified in sperm from *Mtrr*$^{+/gt}$males is not maintained in offspring somatic tissues. This occurs despite epigenetic instability reported in *Mtrr*$^{+/+}$adult liver and *Mtrr*$^{+/+}$F2 generation placentas at E10.5 (Padmanabhan et al., 2013). This suggests DNA methylation does not serve as a long-term heritable memory of abnormal folate metabolism. However, germline DNA methylation changes may be an initiating or contributory factor in transgenerational inheritance in the *Mtrr*$^{gt}$model though the dysregulation of *Hira* and *Rn45s*.

# Chapter 7

# Analysis of sperm small non-coding RNA profiles in *Mtrr*$^{gt}$ mice

Small non-coding RNA library preparation and sequencing was performed in collaboration with Katharina Gapp (Wellcome Sanger Institute, Hinxton, UK) and Eric Miska (Department of Genetics and Gurdon institute, Cambridge, UK). Bioinformatics analysis in this chapter was performed in collaboration with Dr Russell Hamilton (Department of Genetics & Centre for Trophoblast Research, Cambridge, UK).

The introduction of this chapter contains elements published in the following textbook chapter and review: Blake et al. (2018); Blake and Watson (2016). Permission was granted by the publisher for their inclusion in this thesis.

## 7.1 Introduction

In recent years non-coding (nc) RNAs have become the focus of those studying epigenetic inheritance. There is a huge diversity of ncRNA species within mammalian germ cells (Liebers et al., 2014). Furthermore, the RNA profiles of sperm and oocytes differ (Liebers et al., 2014). Oocytes contain the mRNA transcripts required for embryonic development before zygotic genome activation in addition to a range of ncRNAs (Veselovska et al., 2015). Sperm contain some mRNAs (mostly of genes involved in spermatogenesis), fragmented ribosomal-RNA (rsRNA), fragmented longer transcripts, miRNAs, piRNAs, tRNA and tRNA fragments (tsRNAs), and other small RNAs (Johnson et al., 2011; Casas and Vavouri, 2014; Ostermeier et al., 2002). RNAs in sperm can be delivered to the oocyte upon fertilisation (Ostermeier et al., 2004). Importantly, recent studies have shown that sperm derived RNAs can influence phenotypes in offspring (Grandjean et al., 2015; Gapp et al., 2014; Sharma et al., 2016).

Indeed many recent studies have implicated ncRNAs (particularly sncRNAs) in models of epigenetic inheritance. ncRNAs have long been known to be inherited in *C. elegans* (Rechavi and Lev, 2017). For example, starvation-induced developmental arrest in *C. elegans* causes the transgenerational inheritance of small RNAs and an increased lifespan over multiple generations (Rechavi et al., 2014). In the classic $A^{vy}$ mouse model of epigenetic inheritance, sperm ncRNAs have recently been implicated in the obesity phenotype (Chapter 1). Sperm ncRNA profiles, including tRNA fragments, are abnormal in the sperm of F1 males derived from mating an $A^{vy}$ obese male to wildtype lean female (Cropley et al., 2016). The F2 generation derived from these males had defects in glucose and lipid metabolism (Cropley et al., 2016). There is also evidence that the profile of ncRNAs in sperm can be altered by a range of environmental exposures and these changes are associated with phenotypes in the offspring (Table 7.1).

Many studies have suggested a direct causative link between the dysregulated sperm ncRNAs and offspring phenotypes using microinjection studies (Table 7.1). sncRNAs purified from sperm exposed to the environmental stressor are injected into a naive, unexposed fertilised oocyte. The embryo is transferred to a recipient mother and offspring phenotypes are assessed. For

Table 7.1: **Epigenetic inheritance paradigms in which sperm ncRNA profiles are disrupted following exposure to an environmental insult**.

| Exposure | Offspring Phenotype | Sperm ncRNAs Misexpressed | Microinjection? | Phenotype Recapitulated? | Other Details | Reference |
|---|---|---|---|---|---|---|
| Western-diet (high-fat, high-sugar) | Obesity, GI and IR | miRNAs and piRNAs | Yes (miR19b) | ↑ body weight, variable GI and IR | | Grandjean et al. (2015) |
| High-fat diet | Obesity, GI and IR | miRNAs and tsRNAs | Yes 1) 30-40nt tsRNAs, 2) synthetic tRNAs | 1) GI, no ↑ body weight or IR, 2) No metabolic phenotype | RNA-seq identified transcriptional changes in HFD embryos | Chen et al. (2016a) |
| High-fat diet | Female GI and resistance to HFD-induced weight gain | miRNA (let7c) | No | | Altered let7c target expression in offspring. Rat study | de Castro Barbosa et al. (2016) |
| High-fat diet | Sub-fertility | miRNAs | No | | miRNAs not dysregulated in sperm of offspring | Fullston et al. (2013) |
| Low-protein diet | Altered hepatic cholesterol biosynthesis | tsRNAs, miRNA, piRNA and let-7 | Yes | ↓ expression of MERVL targets in 2-cell embryos | Showed tsRNAs gained on transit through epididymis | Sharma et al. (2016) |

| Exposure | Offspring Phenotype | Sperm ncRNAs Misexpressed | Microin-jection? | Phenotype Recapitulated? | Other Details | Reference |
|---|---|---|---|---|---|---|
| Maternal separation and unpredictable maternal stress (MSUS) | Behavioural phenotypes: ↑ risk taking, fear, despair. Metabolic: ↓ body weight, insulin hyersensitivity, GI | miRNA, lncRNA, sncRNA | Yes 1) miRNA, 2) lncRNA, 3) sncRNA | 1) behavioural phenotypes, ↓ body weight, 2) insulin hyper-sensitivity, GI and ↓ fear, 3) ↑ body weight, GI and behavioural despair | Injection of both sncRNA and lncRNA required for some phenotypes | Gapp et al. (2014, 2018) |
| *in utero* vinclozolin exposure | Reduced male fertility, prostate and kidney disease | tsRNAs, miRNAs, rRNA-derived sncRNAs, piRNAs | No | | Analysis performed on F3 generation sperm. Rat study | Schuster et al. (2016). |
| Environmental enrichment | Enhanced synaptic plasticity, cognition and memory | miRNAs (miRs 212/132) | Yes | Enhanced synaptic plasticity, improved memory | | Benito et al. (2018) |
| Exercise | Reduced anxiety and fear memory in males | miRNAs, tsRNAs | No | | | Short et al. (2017) |

A non-exhaustive table of key studies showing disruption of sperm ncRNA profiles and associated offspring phenotypes. All studies were carried out in mice unless otherwise stated. GI: glucose intolerance, IR insulin resistance, HFD: high fat diet, ↓: decreased, ↑: increased.

instance, traumatic stress in early postnatal life of the F1 generation leads to altered miRNA profiles in the F1 sperm and severe behavioural phenotypes in the F2 generation (Gapp et al., 2014). Offspring derived from microinjection of miRNAs from stressed F1 male sperm into fertilised oocytes had similar behavioural phenotypes to conventionally derived offspring (Gapp et al., 2014). Poor diet in male mice can alter metabolic gene expression in early development of the F1 generation (Sharma et al., 2016) and lead to adult metabolic disease phenotypes (Chen et al., 2016a). These effects have been linked to altered tRNA fragment profiles in the sperm (Chen et al., 2016a; Sharma et al., 2016). When sperm tRNA fragments isolated from high-fat diet fed mice were microinjected into fertilised oocytes, it partially recapitulated metabolic disease in adulthood (Chen et al., 2016a). Interestingly, microinjection of synthetic tRNA fragments was unable to recapitulate the phenotype (Chen et al., 2016a) suggesting that RNA modifications (e.g., methylation) may be important in epigenetic inheritance of disease. Additionally, microinjection of sperm tRNA fragments from mice fed a low-protein diet into oocytes specifically repressed early embryonic genes associated with MERVL retroelements (Sharma et al., 2016). These genes are hypothesised to influence feto-placental development, the dysregulation of which might lead to adult-onset metabolic disease. However, in many microinjection studies only a partial recapitulation of phenotype is achieved (Table 7.1) (Chen et al., 2016a; Gapp et al., 2018)). This indicates that sncRNAs may work in concert with other epigenetic mechanisms.

The effect of folate deficiency or disrupted folate metabolism on ncRNAs has not been widely studied. One report observed that obese mice fed a high-fat diet, when also supplemented with folic acid, have alterations in the expression of lncRNAs in the heart (Ma et al., 2017). There is some circumstantial evidence in the *Mtrr*$^{gt}$model that might suggest ncRNA modifications may be relevant to investigate further. I observed that mRNA expression of the RNA methyltransferase *Dnmt2* was down-regulated specifically in the epididymis of *Mtrr*$^{+/+}$(derived from *Mtrr*$^{+/gt}$intercrosses), *Mtrr*$^{+/gt}$and *Mtrr*$^{gt/gt}$males compared to C57Bl/6 controls. It is intriguing that decreased *Dnmt2* mRNA expression was observed in *Mtrr*$^{+/+}$epididymis, suggesting exposure to the parental *Mtrr*$^{gt}$allele is sufficient for its misexpression. This has important implica-

tions in the context of transgenerational inheritance. *Dnmt2* is required for the methylation of tsRNAs and it has been implicated in other models of TEI (Tuorto et al., 2012; Zhang et al., 2018; Kiani et al., 2013). Therefore, I aimed to characterise the sncRNA profiles of sperm from $Mtrr^{+/+}$, $Mtrr^{+/gt}$and $Mtrr^{gt/gt}$males to further understand the mechanism of TEI in the $Mtrr^{gt}$model.

## 7.2 Results

### 7.2.1 The sncRNA-seq data from *Mtrr* male sperm was high quality

In order to explore a potential role for sncRNAs in transgenerational inheritance of congenital malformations in the $Mtrr^{gt}$model, I examined the sncRNA profiles of sperm. Sperm were isolated from C57Bl/6, $Mtrr^{+/+}$, $Mtrr^{+/gt}$and $Mtrr^{gt/gt}$16-20 week old male mice. Six males were assessed per genotype group, derived from at least three different litters. I confirmed the purity of RNA extracted from sperm using BioAnalyzer analysis (Figure 7.1 A-D). Importantly, no 28s rRNA peak and no/minimal 18s rRNA peak was observed (at around 3600nt and 1500nt, respectively) (Figure 7.1 A-D). rRNA is cleaved in sperm to cause translational arrest (Johnson et al., 2011). Therefore, the presence of 28s and 18s rRNA would indicate somatic cell contamination of sperm samples. From total sperm RNA I generated small-RNA sequencing libraries to assess the sncRNA portion of the sperm transcriptome. Library size was quantified using TapeStation analysis (Figure 7.1 E). This confirmed all libraries had the expected size of 160-180bp, which corresponds to adapter ligated small RNAs. Libraries were sequenced using 50bp single-end reads.

On average 8.7 million reads were sequenced per library (Table 7.2). In total, an average 52.5 million mappable reads were obtained per genotype group; 34.5 million reads for C57Bl/6 libraries, 57.8 million reads for $Mtrr^{+/+}$libraries, 56.3 million reads for $Mtrr^{+/gt}$libraries and 61.3 million reads for $Mtrr^{gt/gt}$libraries. This was a comparable number of reads to that obtained in similar studies (Chen et al., 2016a; Sharma et al., 2016). Three of the C57Bl/6 libraries had a very low number of reads (Table 7.2). There was no obvious cause for this but it may have arisen from poor library preparation or errors during multiplexing. It was noted and kept in consideration during subsequent analyses.

Table 7.2: **Number of reads obtained for sncRNA-seq libraries**.

| Genotype | Sample | Total Reads (million bp) |
|---|---|---|
| C57Bl/6 | 1 | 10.6 |
| | 2 | 10.1 |
| | 3 | 7.1 |
| | 4 | 2.2 |
| | 5 | 2.6 |
| | 6 | 1.8 |
| *Mtrr*$^{+/+}$ | 1 | 9.6 |
| | 2 | 10.1 |
| | 3 | 11.7 |
| | 4 | 9.7 |
| | 5 | 8.0 |
| | 6 | 7.4 |
| *Mtrr*$^{+/gt}$ | 1 | 9.2 |
| | 2 | 12.4 |
| | 3 | 10.0 |
| | 4 | 9.6 |
| | 5 | 8.6 |
| | 6 | 6.5 |
| *Mtrr*$^{gt/gt}$ | 1 | 11.9 |
| | 2 | 9.1 |
| | 3 | 11.9 |
| | 4 | 11.2 |
| | 5 | 8.7 |
| | 6 | 8.5 |

The quality of sncRNA-sequencing data was assessed using FastQC. All samples had mean phred quality scores of greater than 30 throughout the read length (Figure 7.1 F). Analysis of the GC content of reads highlighted that while sncRNA-Seq libraries had an average GC content of 52.5%, this was bimodally distributed (Figure 7.1 G). The higher of the two peaks can likely be attributed to a high percentage of rRNA in some libraries. rRNA is known to have a high GC content (Johnson et al., 2011). Analysis of overrepresented sequences supports this as one the most overrepresented sequences corresponds to a fragment of 28s rRNA (Figure 7.1 H). While library contamination could not be entirely excluded, overrepresentation of certain sequences is not unexpected in small RNA libraries not subjected to fragmentation. Some sequences can be present in significant proportions in such libraries. The tRNA-valine sequence

was the most commonly overrepresented sequence I identified. I confirmed that overrepresented sequences did not correspond to adaptor contamination. Average read length was 35bp. Reads had a size distribution corresponding to three fragment size peaks, 28-30 bp, 32-34 bp and 40-42 bp (Figure 7.1 I). Overall the quality metrics indicated the data was of sufficient quality to proceed with sncRNA mapping and differential expression analysis.

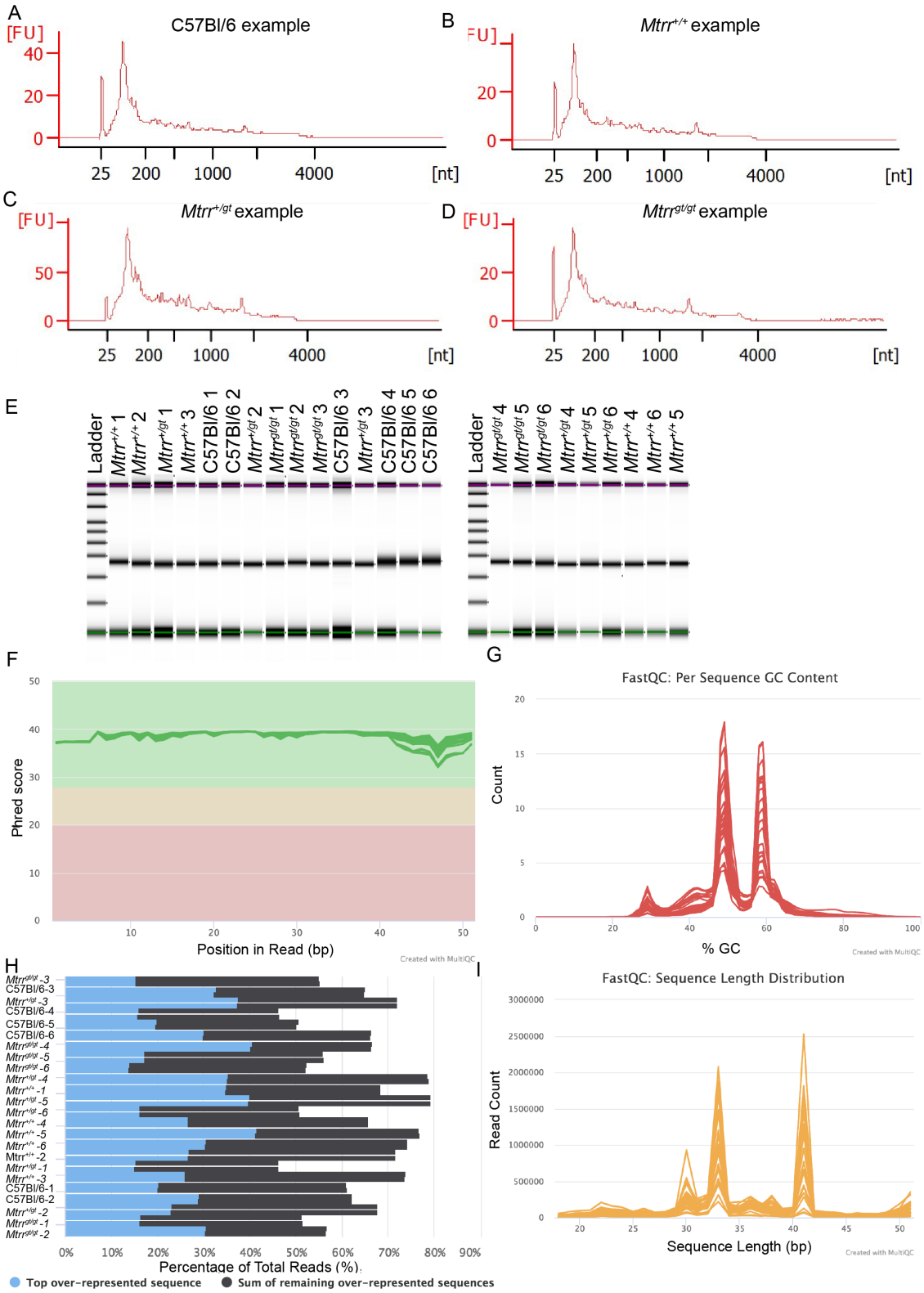

A
C57Bl/6 example
[FU]
40
20
0
25
200
1000
4000
[nt]
B
Mtrr+/+ example
C
Mtrr+/gt example
D
Mtrrgt/gt example
E
Ladder
Mtrr+/+ 1
Mtrr+/+ 2
Mtrr+/gt 1
Mtrr+/+ 3
C57Bl/6 1
C57Bl/6 2
Mtrr+/gt 2
Mtrrgt/gt 1
Mtrrgt/gt 2
Mtrrgt/gt 3
C57Bl/6 3
Mtrr+/gt 3
C57Bl/6 4
C57Bl/6 5
C57Bl/6 6
Ladder
Mtrrgt/gt 4
Mtrrgt/gt 5
Mtrrgt/gt 6
Mtrr+/gt 4
Mtrr+/gt 5
Mtrr+/gt 6
Mtrr+/+ 4
Mtrr+/+ 6
Mtrr+/+ 5
F
Phred score
Position in Read (bp)
Created with MultiQC
G
FastQC: Per Sequence GC Content
Count
% GC
Created with MultiQC
H
Mtrrgt/gt -3
C57Bl/6-3
Mtrr+/gt -3
C57Bl/6-4
C57Bl/6-5
C57Bl/6-6
Mtrrgt/gt -4
Mtrrgt/gt -5
Mtrrgt/gt -6
Mtrr+/gt -4
Mtrr+/+ -1
Mtrr+/gt -5
Mtrr+/gt -6
Mtrr+/+ -4
Mtrr+/+ -5
Mtrr+/+ -6
Mtrr+/+ -2
Mtrr+/gt -1
Mtrr+/+ -3
C57Bl/6-1
C57Bl/6-2
Mtrr+/gt -2
Mtrrgt/gt -1
Mtrrgt/gt -2
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
Percentage of Total Reads (%)
Top over-represented sequence
Sum of remaining over-represented sequences
I
FastQC: Sequence Length Distribution
Read Count
Sequence Length (bp)
Created with MultiQC

Figure 7.1: **Sperm RNA purity, library size confirmation and quality metrics of sncRNA-seq data. (A-D)** BioAnalyzer profiles demonstrate sperm RNA purity for representative sperm samples from **(A)** C57Bl/6, **(B)** *Mtrr*$^{+/+}$, **(C)** *Mtrr*$^{+/gt}$, and **(D)** *Mtrr*$^{gt/gt}$males. **(E)** TapeStation analysis was used to assess small RNA-seq library size and quantity. Libraries are the central band, the upper and lower bands are makers. **(F)** Phred quality scores along the read length are shown for all libraries. **(G)** The fraction of reads with given GC content (%) is plotted for all libraries. **(H)** The percentage of all reads made up of overrepresented reads, with the top overrepresented reads (blue) and sum of the other overrepresented reads (grey) plotted for each library. **(I)** The distribution of fragment sizes (read) lengths is plotted for each sample.

### 7.2.2 Profiling the sncRNA content of sperm from C57Bl/6, *Mtrr*$^{+/+}$, *Mtrr*$^{+/gt}$and *Mtrr*$^{gt/gt}$ males

Alignment was first performed to the reference genome (mm10) using STAR (Dobin et al., 2013). This quality control step allowed the mapping consistency between libraries to be assessed. Alignment rates were poor, with only an average of 18.4% of reads mapping uniquely to the reference genome. Alignment rates improved to approximately 50% of the total reads when reads that mapped to multiple loci were included (Figure 7.2). Up to 40% of reads in some libraries failed to map, primarily because they were too short (Figure 7.2). However, the poor alignment rates likely reflect the limitations of the STAR aligner when faced with reads length that would be expected in sncRNA libraries. This highlighted the need to map sncRNA-seq reads to specialised sncRNA databases e.g. miRBase (Kozomara and Griffiths-Jones, 2014). I used two parallel sncRNA mapping approaches. Firstly, I used SPORTS1.0, a small RNA annotation pipeline that performs mapping of tsRNAs, rRNAs, miRNA, piRNA and other ncRNAs as part of an integrated package (Shi et al., 2018). In parallel, reads were mapped to individual sncRNA databases using Salmon (Patro et al., 2017) as performed in Gapp et al. (2018) (see chapter 2 for details of databases used).

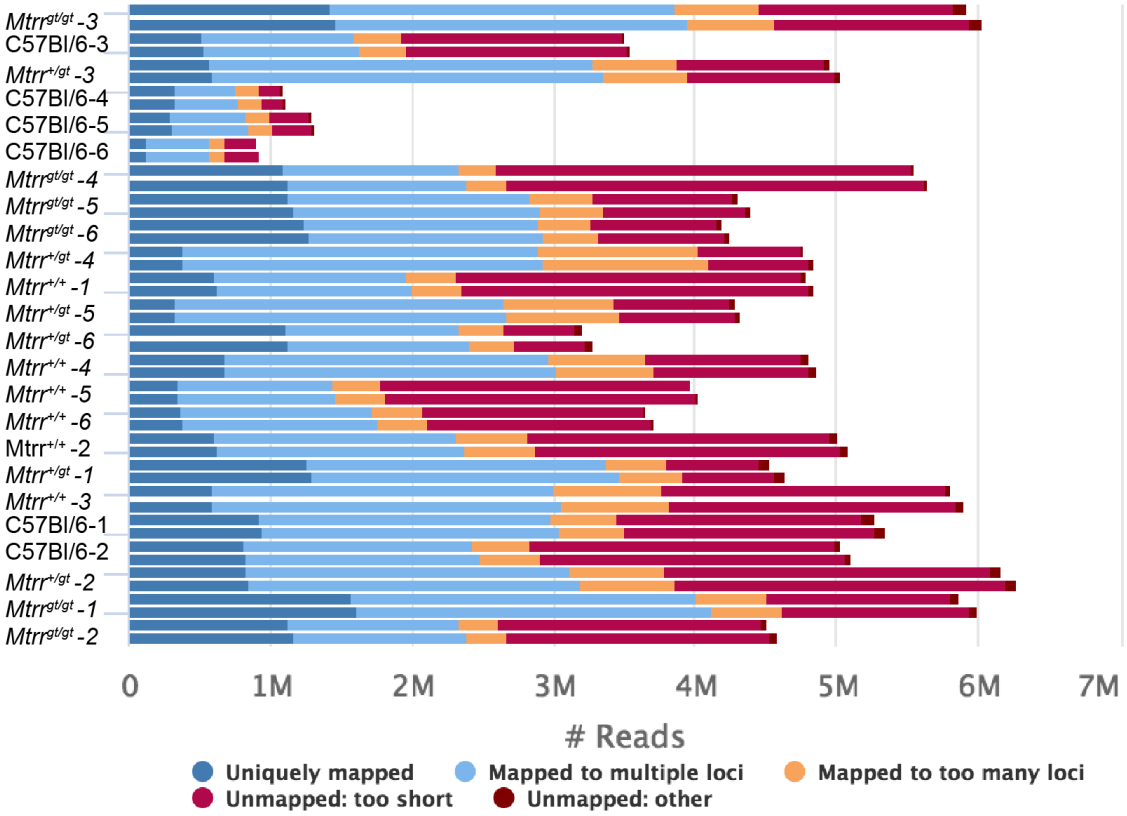



Figure 7.2: **Alignment of sncRNA-seq reads to the reference genome using STAR**. The number of reads that mapped to the mm10 reference genome uniquely (dark blue), at multiple loci (pale blue), at too many loci (orange) or that failed to map to the reference genome due to being too short (red) or another reason (maroon) are shown for all samples. Each sample is represented by two bars, one per sequencing lane.

I compared the annotation of sncRNAs performed using Salmon and SPORTS1.0. While it is difficult to perform direct comparisons between the two methods, overall there was good consensus in the proportion of transcripts assigned to each ncRNA species between the two approaches. For example, an average of 35.3% of reads mapped to rRNA using Salmon and 35.0% using SPORTS1.0, similarly 34.8% of reads aligned to the tRNA database using Salmon and 31.4% using SPORTS1.0. The raw number of reads mapped was substantially lower using Salmon than SPORTS 1.0. This likely reflects differences between the aligner used as part of SPORTS1.0 (Bowtie) and Salmon and also slight differences in the sncRNA databases used for mapping.

To compare the overall sncRNA profiles of sperm from C57Bl/6, *Mtrr*$^{+/+}$, *Mtrr*$^{+/gt}$and *Mtrr*$^{gt/gt}$males I used the categorisation of sncRNAs generated using SPORTS1.0 (Shi et al., 2018). For each sperm sample a sncRNA profile was produced (Figure 7.3). The sncRNA profiles were variable between individuals across all genotypes (Figure 7.3). In most sperm samples, the most abundant sncRNA species present were fragmented tRNAs (tsRNAs) between 28-32 bp in length ($\sim$ 31% of all sncRNAs on average). However, in some samples fragmented rRNA (rsRNA) accounted for the majority of sncRNA present (e.g. Figure 7.3 B, G, K, T, V). The abundance of rsRNAs present was highly variable across all libraries, with rsRNA accounting for between 17% and 57% of total reads (discussed further in Section 7.2.3). In addition to tsRNAs and rsRNAs, sperm also contained piRNAs ($\sim$ 9.7% of all sncRNAs on average), miRNAs ($\sim$ 1.8% of all sncRNAs on average), and other ncRNAs ($\sim$ 0.3% of all sncRNAs on average). Additionally, there were unannotated reads some which could be mapped to the genome (match) and some which could not (unmatch). This included mitochondrial DNA, snoRNA, pseudogenes and mRNAs. The mapping of sncRNAs in sperm from C57Bl/6, *Mtrr*$^{+/+}$, *Mtrr*$^{+/gt}$and *Mtrr*$^{gt/gt}$males was in line with that reported in other studies using SPORTS1.0 (Zhang et al., 2018).

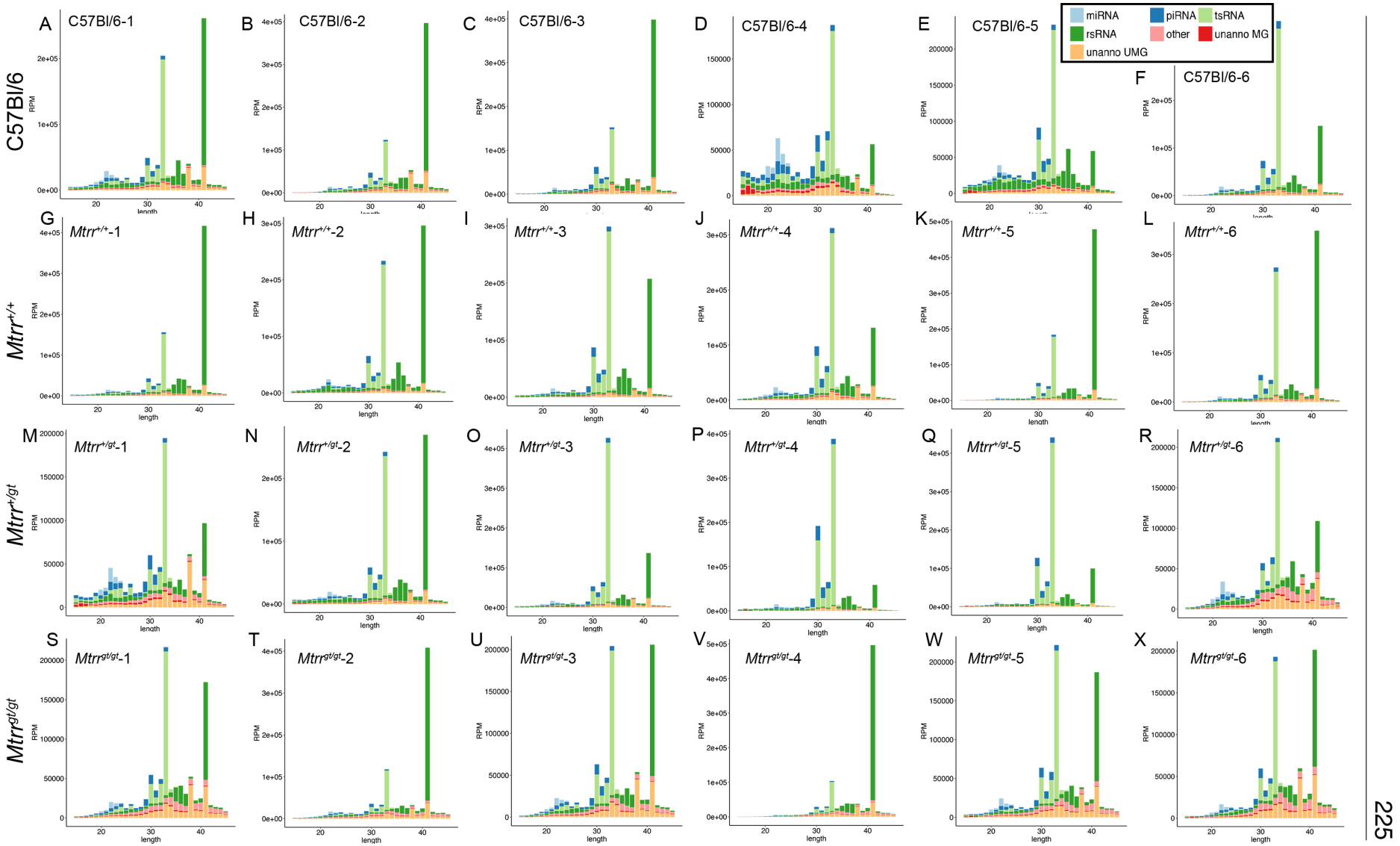

C57Bl/6
Mtrr+/+
Mtrr+/gt
Mtrrgt/gt
A C57Bl/6-1
B C57Bl/6-2
C C57Bl/6-3
D C57Bl/6-4
E C57Bl/6-5
F C57Bl/6-6
G Mtrr+/+-1
H Mtrr+/+-2
I Mtrr+/+-3
J Mtrr+/+-4
K Mtrr+/+-5
L Mtrr+/+-6
M Mtrr+/gt-1
N Mtrr+/gt-2
O Mtrr+/gt-3
P Mtrr+/gt-4
Q Mtrr+/gt-5
R Mtrr+/gt-6
S Mtrrgt/gt-1
T Mtrrgt/gt-2
U Mtrrgt/gt-3
V Mtrrgt/gt-4
W Mtrrgt/gt-5
X Mtrrgt/gt-6
miRNA
piRNA
tsRNA
rsRNA
other
unanno MG
unanno UMG
RPM
length

Figure 7.3: **Categorisation and length distribution analysis of sperm sncRNAs.** The sncRNA length distribution profiles are shown for sperm from **(A-F)** C57Bl/6, **(G-L)** *Mtrr*$^{+/+}$, **(M-R)** *Mtrr*$^{+/gt}$and **(S-X)** *Mtrr*$^{gt/gt}$males. Reads per million versus read length is plotted for sncRNAs categorised as tsRNA (pale green), rsRNAs (dark green), miRNAs (pale blue), piRNAs (dark blue), other ncRNAs (pink), unannotated but matching genome (MG) (red) and unannotated and not matching genome (UMG) (orange). Analysis performed using SPORTS1.0.

To perform comparisons between sncRNA profiles of sperm from C57Bl/6, *Mtrr*$^{+/+}$, *Mtrr*$^{+/gt}$and *Mtrr*$^{gt/gt}$males, for each genotype group I calculated the number of reads annotated to each sncRNA species using SPORTS1.0 and then normalised this figure by the total number of sequencing reads per genotype group (Table 7.3). There was no significant difference in the overall normalised sncRNA proportions in sperm from *Mtrr*$^{+/+}$, *Mtrr*$^{+/gt}$and *Mtrr*$^{gt/gt}$males compared to C57Bl/6 controls ($p > 0.05$, two-way ANOVA with Dunnett's multiple comparison's test, Figure 7.4).

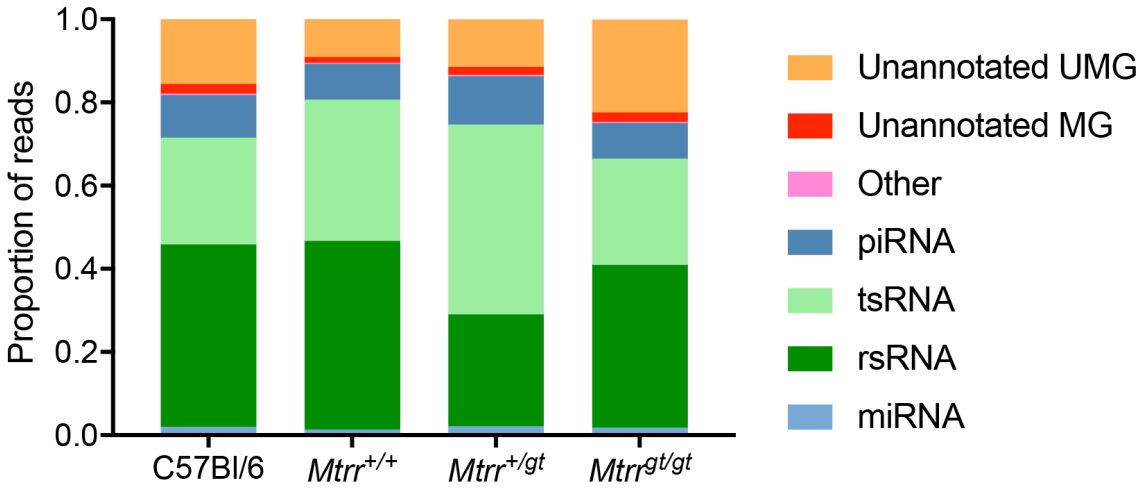



Figure 7.4: **sncRNA profiles are consistent in C57Bl/6, *Mtrr*$^{+/+}$, *Mtrr*$^{+/gt}$and *Mtrr*$^{gt/gt}$male sperm.** The number of reads annotated to each sncRNA species was normalised by the total number of sequencing reads for C57Bl/6, *Mtrr*$^{+/+}$, *Mtrr*$^{+/gt}$and *Mtrr*$^{gt/gt}$male sperm. sncRNAs were categorised as tsRNA (pale green), rsRNAs (dark green), miRNAs (pale blue), piRNAs (dark blue), other ncRNAs (pink), unannotated but matching genome (MG) (red) and unannotated and not matching genome (UMG) (orange). Analysis was performed with SPORTS1.0. Statistical test performed: two-way ANOVA with Dunnett's multiple comparisons test.

Table 7.3: **Summary of sncRNA read categorisation using SPORTS1.0**.

| Annotation | C57Bl/6 sperm | *Mtrr*$^{+/+}$sperm | *Mtrr*$^{+/gt}$sperm | *Mtrr*$^{gt/gt}$sperm |
|---|---|---|---|---|
| Clean reads | 31628736 | 53822646 | 55005120 | 55814955 |
| Match genome reads | 13484088 | 23992598 | 32670092 | 24838127 |
| tsRNA reads | 7725362 | 17744027 | 23726845 | 12412566 |
| rsRNA reads | 13168474 | 23605404 | 13955560 | 19052887 |
| miRNA reads | 623645 | 715133 | 1113003 | 884801 |
| piRNA reads | 3074569 | 4463850 | 6045252 | 4160870 |
| Unannotated MG | 699181 | 718058 | 995007 | 1107080 |
| Unannotated UMG | 4675948 | 4721132 | 5966681 | 10891919 |
| Other | 110924 | 130059 | 154877 | 150415 |

Summary of sncRNA reads as categorised using SPORTS1.0. tsRNA, rsRNA, miRNA and piRNA read totals include both reads that can be aligned to the genome (MG) and those that cannot be aligned to the genome (UMG) but align to the respective ncRNA database.

### 7.2.3 rsRNA profiles of C57Bl/6, *Mtrr*$^{+/+}$, *Mtrr*$^{+/gt}$ and *Mtrr*$^{gt/gt}$ male sperm

Given my previous finding that there is a DMR at the *Rn45s* promoter in *Mtrr*$^{+/gt}$and *Mtrr*$^{gt/gt}$sperm and that *Rn45s* RNA is down-regulated in *Mtrr*$^{+/+}$F1, F2 and F3 generation embryos, I sought to explore the rRNAs in sperm more closely. To investigate if the high levels of rsRNA in some samples represented somatic cell contamination of sperm samples, the sncRNA categorisation profile of each sample (Figure 7.3) was compared to the respective sperm RNA BioAnalyzer profile (Figure 7.1). No correlation was observed between 18s peak size and/or presence on the BioAnalyzer profile and the percentage of total reads accounted for by rsRNAs. This suggested that the high levels of rsRNA likely did not reflect library contamination, although this is challenging to rule out. Variability in rsRNA abundance between samples may be technical, e.g. due PCR amplification biases at these regions, or biological, e.g. reflecting variability in rRNA levels in sperm. The levels of rRNA I report are in line with previous studies. Approximately 80% of reads were seen to align to portions

of the 28s and 18s rRNAs in human sperm (Johnson et al., 2011) and between 20% and 80% of transcripts in sperm of males exposed to MSUS were rRNA or mitchondrial rRNA (Gapp et al., 2018). However, some studies have reported that rRNA abundance in sperm was low with respect to testis and epididymal somatic samples (Sharma et al., 2016). Observations of the raw sncRNA counts in sperm obtained by Sharma et al. (2016) suggest that rRNA levels are approximately equivalent to those I report. Overall, the rsRNA levels appear normal, if variable, in sperm of *Mtrr*$^{gt}$males, both with respect to C57Bl/6 controls and in comparison to the literature more generally.

rRNAs are degraded in sperm to bring about transcriptional silencing leading to the generation of rRNA fragments (Johnson et al., 2011). A key feature of SPORTS1.0 was the ability to determine the rRNA subtypes (e.g. 18S, 5.8S) from which rRNA fragments were derived (Shi et al., 2018). Mapping of rsRNAs identified in sperm from C57Bl/6, *Mtrr*$^{+/+}$, *Mtrr*$^{+/gt}$and *Mtrr*$^{gt/gt}$males demonstrated that rsRNAs tend to be around 35-40 bp long, with a strong peak at 40 bp (Figure 7.5). The majority of rsRNAs were derived from the 28s rRNA, regardless of length. rsRNAs derived from the 5S and 5.8S rRNAs were the second and third most abundant rsRNA species, but accounted for a far smaller percentage of total rsRNAs (Figure 7.5). Overall, the mapping and length distribution of rsRNAs appeared equivalent across all samples, regardless of genotype, in spite of differences in overall rsRNA abundance.

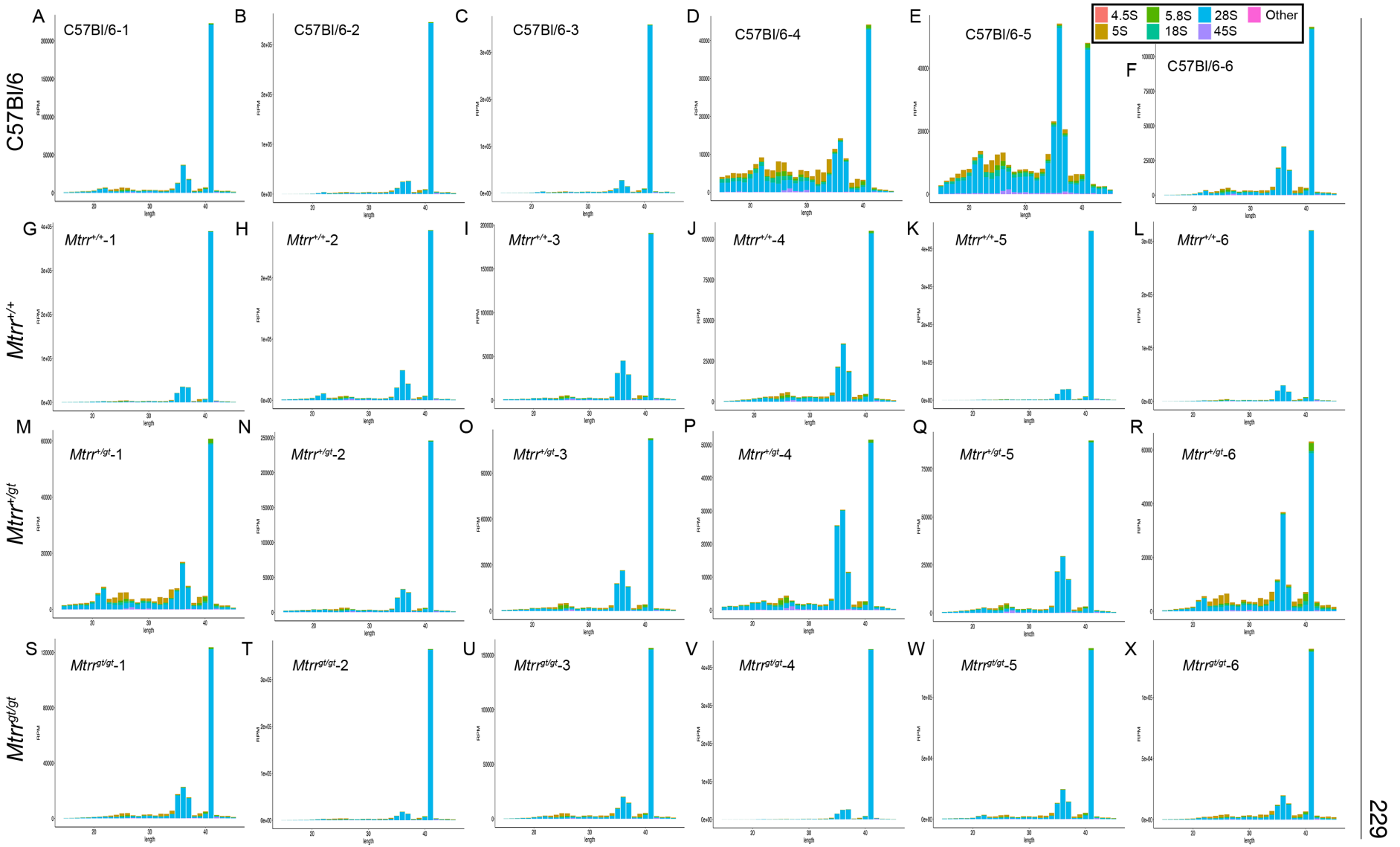

4.5S
5S
5.8S
18S
28S
45S
Other
C57Bl/6
Mtrr+/+
Mtrr+/gt
Mtrrgt/gt
A C57Bl/6-1
B C57Bl/6-2
C C57Bl/6-3
D C57Bl/6-4
E C57Bl/6-5
F C57Bl/6-6
G Mtrr+/+-1
H Mtrr+/+-2
I Mtrr+/+-3
J Mtrr+/+-4
K Mtrr+/+-5
L Mtrr+/+-6
M Mtrr+/gt-1
N Mtrr+/gt-2
O Mtrr+/gt-3
P Mtrr+/gt-4
Q Mtrr+/gt-5
R Mtrr+/gt-6
S Mtrrgt/gt-1
T Mtrrgt/gt-2
U Mtrrgt/gt-3
V Mtrrgt/gt-4
W Mtrrgt/gt-5
X Mtrrgt/gt-6
RPM
length

Figure 7.5: **rsRNA subtype analysis for C57Bl/6, *Mtrr*$^{+/+}$, *Mtrr*$^{+/gt}$ and *Mtrr*$^{gt/gt}$ male sperm.** **(A-X)** The rsRNA length distribution and subtype analysis profiles are shown for **(A-F)** C57Bl/6, **(G-L)** *Mtrr*$^{+/+}$, **(M-R)** *Mtrr*$^{+/gt}$and **(S-X)** *Mtrr*$^{gt/gt}$sperm samples. Reads per million (RPM) is plotted against read length for rsRNAs categorised by the rRNA from which they were derived: 4.5S rRNA (orange), 5S rRNA (mustard), 5.8S rRNAs (green), 18S rRNA (teal), 28S rRNA (blue), 45S rRNA (purple) and other rRNA (pink). Analysis performed using SPORTS1.0.

Furthermore, mapping information on the origin of the rsRNAs within the rRNA subtype locus was provided by SPORTS1.0 (Shi et al., 2018). Simply, the part of the rRNA from which the rsRNA fragments were derived from was assessed. The 3′ end of the 28s rRNA was the primary rsRNA generating locus in sperm (Figure 7.6). rsRNAs were also generated from the 3′ end and centre of the 5.8S rRNA and 3′ and 5′ end of the 5S rRNA (Figure 7.6). Notably, the rsRNA generating loci from the different rRNA precursors were equivalent between C57Bl/6, *Mtrr*$^{+/+}$, *Mtrr*$^{+/gt}$and *Mtrr*$^{gt/gt}$male sperm. There was some variability in the proportion of rsRNA species generated from the 5S and 5.8S rRNA loci, both between and within genotype groups. Overall, the biogenesis of rsRNAs in sperm did not appear to be affected by the *Mtrr*$^{gt}$mutation. However, differential expression analysis on rsRNAs has yet to be performed and will be required to provide substantive evidence to support or refute this conclusion.

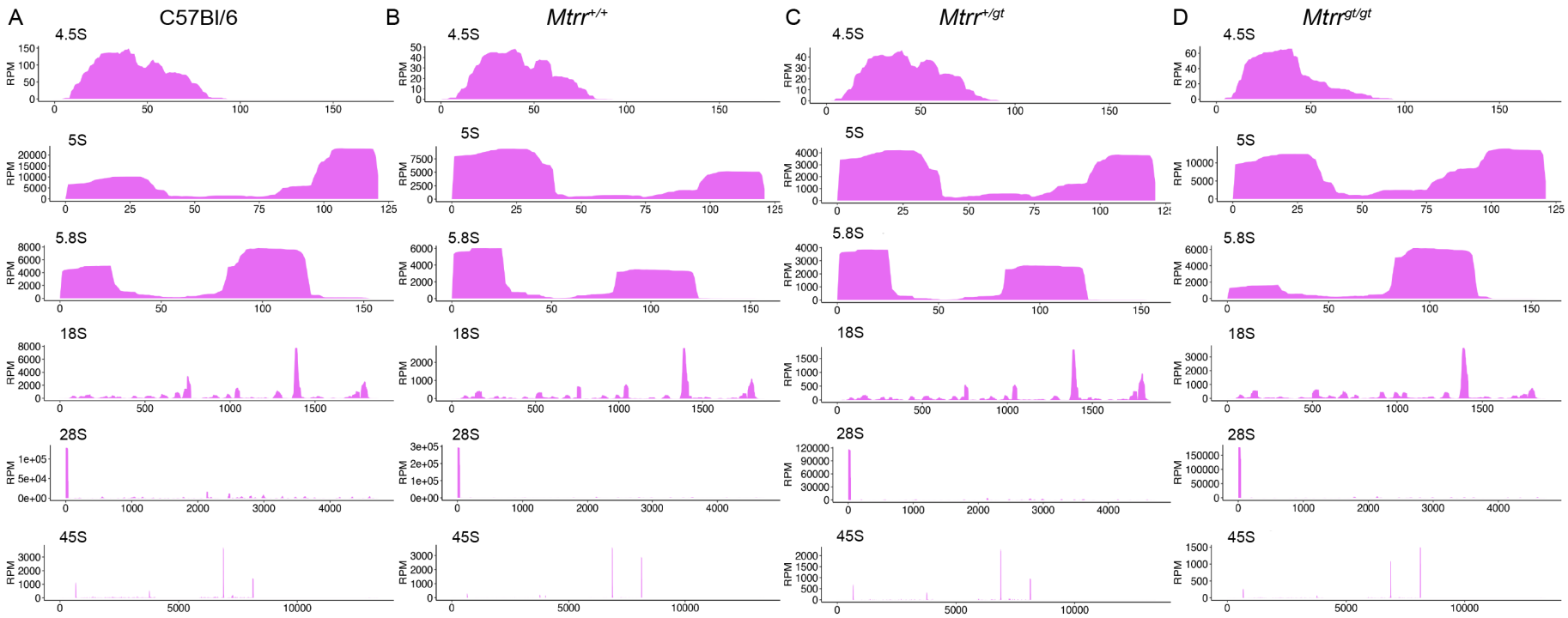



Figure 7.6: **rsRNA-generating loci from rRNA precursors (A-D)** Representative examples of rsRNA-generating loci mapped along the length of different rRNA precursors for rsRNAs identified in sperm from **(A)** C57Bl/6, **(B)** *Mtrr*$^{+/+}$, **(C)** *Mtrr*$^{+/gt}$and **(D)** *Mtrr*$^{gt/gt}$males. This analysis was performed using SPORTS1.0.

### 7.2.4 Identification of differentially expressed sncRNAs in sperm from *Mtrr*$^{+/+}$, *Mtrr*$^{+/gt}$ and *Mtrr*$^{gt/gt}$ males

Prior to performing differential sncRNA expression analysis I used principle component analysis (PCA) to assess sample clustering. The PCA-based clustering was initially performed on all sncRNAs aligned using SPORTS1.0, a total of 895 ncRNA species. Distinct clustering of samples by genotype was not observed (Figure 7.7 A). There was a slight clustering of *Mtrr*$^{gt/gt}$ sperm samples, but C57Bl/6 samples appeared particularly variable. I then performed PCA-clustering using only reads aligned as miRNAs and tsRNAs (Figure 7.7 B, C). Again, distinct sample clusters by genotype were not observed, although some slight clustering was apparent. Interestingly one *Mtrr*$^{+/gt}$ sperm sample (*Mtrr*$^{+/gt}$-6) appeared to be outlying on the miRNA PCA plot (Figure 7.7 B). This may suggest sample *Mtrr*$^{+/gt}$-6 had a distinct miRNA profile compared to the other *Mtrr*$^{+/gt}$ samples but was similar with respect to other ncRNA species. Overall, PCA-clustering highlighted that the sncRNA profiles of C57Bl/6 and *Mtrr*$^{+/+}$, *Mtrr*$^{+/gt}$ and *Mtrr*$^{gt/gt}$ male sperm were grossly similar.

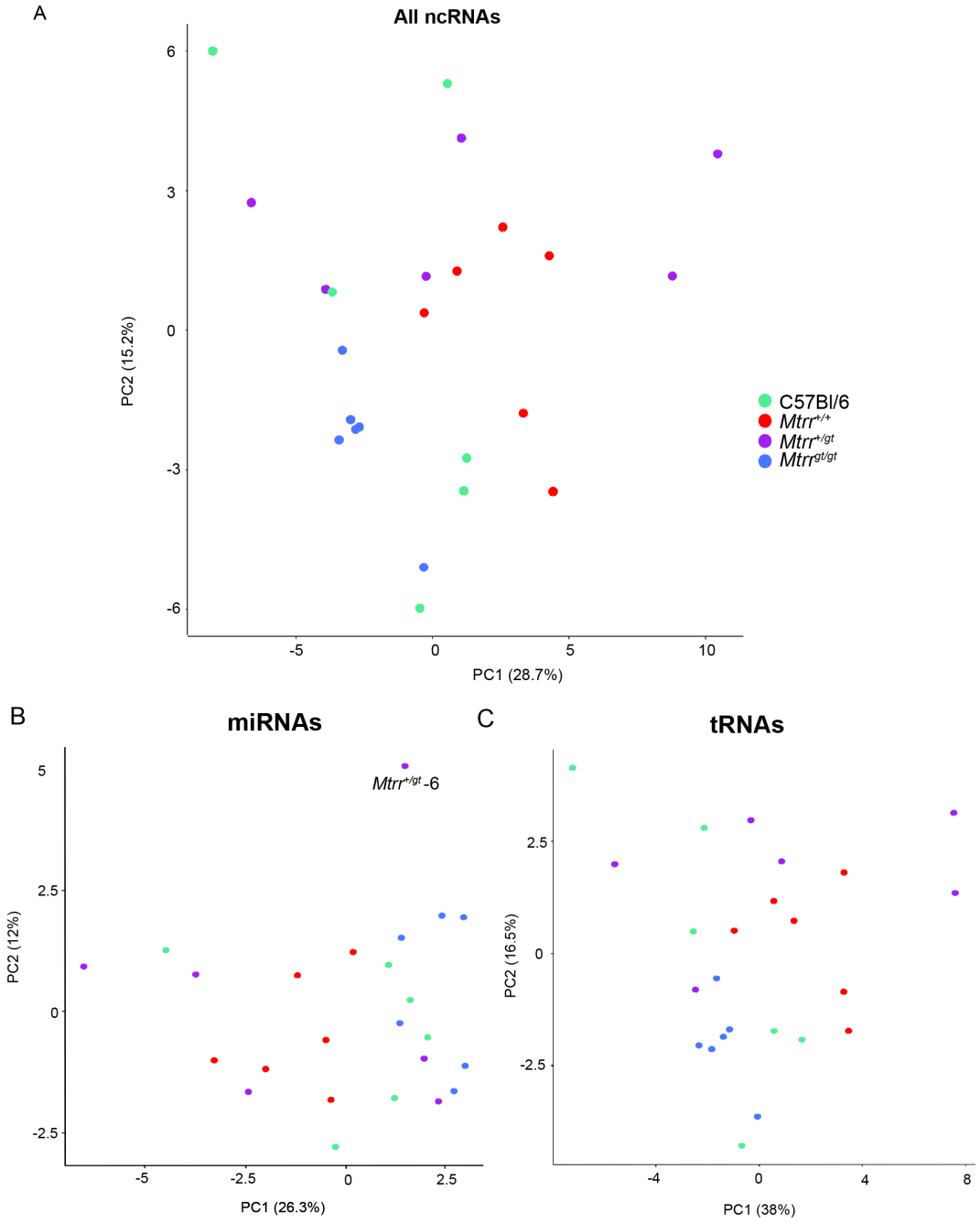



Figure 7.7: **PCA-clustering based on sncRNAs identified in sperm from C57Bl/6, *Mtrr*$^{+/+}$, *Mtrr*$^{+/gt}$and *Mtrr*$^{gt/gt}$males (A-C)** PCA based clustering of sncRNAs identified in sperm from C57Bl/6 (green), *Mtrr*$^{+/+}$(red), *Mtrr*$^{+/gt}$(purple) and *Mtrr*$^{gt/gt}$(blue) males. Data is shown for **(A)** all sncRNA species **(B)** miRNAs only and **(C)** tRNAs only. This analysis was performed using reads aligned using SPORTS1.0.

#### 7.2.4.1 tsRNAs

tRNA-derived sncRNAs have been implicated in the inheritance of metabolic phenotypes following dietary manipulation (Chen et al., 2016a; Sharma et al., 2016). Therefore I investigated if the *Mtrr*$^{gt}$mutation disrupted tsRNA expression in sperm from *Mtrr*$^{+/+}$, *Mtrr*$^{+/gt}$and *Mtrr*$^{gt/gt}$males compared to C57Bl/6 controls. The overall relative abundance of tsRNAs was equivalent between C57Bl/6, *Mtrr*$^{+/+}$, *Mtrr*$^{+/gt}$and *Mtrr*$^{gt/gt}$sperm ($p > 0.1$, two-way ANOVA on normalised total sncRNA read counts, Figure 7.4). tRNA fragments can be derived from various loci of the tRNA: 5′ terminus, 3′ terminus or 3′CCA end (Kumar et al., 2016). On average 95% of sperm tsRNAs were tRNA-5′ derived (Figure 7.8 A-D). The relative frequency of tsRNAs derived from each locus was equivalent between *Mtrr*$^{+/+}$, *Mtrr*$^{+/gt}$, *Mtrr*$^{gt/gt}$and C57Bl/6 control sperm ($p > 0.1650$, two-way ANOVA with Dunnett's multiple comparisons test). As tsRNAs derived from different loci have different biological functions (Kumar et al., 2016), this may suggest that overall tsRNAs were functionally equivalent in *Mtrr*$^{+/+}$, *Mtrr*$^{+/gt}$and *Mtrr*$^{gt/gt}$sperm compared to C57Bl/6 controls.

However, individual tRNA species can be misexpressed despite overall abundance being normal. Firstly, I performed differential expression analysis on tsRNAs aligned using SPORTS1.0, comparing tRNAs expression in *Mtrr*$^{+/+}$, *Mtrr*$^{+/gt}$and *Mtrr*$^{gt/gt}$male sperm to that in C57Bl/6 controls. No differentially expressed tsRNAs were identified in *Mtrr*$^{+/+}$sperm (Figure 7.8 E). Nineteen tsRNAs were misexpressed in *Mtrr*$^{+/gt}$sperm ($p<0.05$, Figure 7.8 F). One tRNA, 5′-tRNA-Leu-CAA, was down-regulated and eighteen tsRNAs were up-regulated (Table 7.4). The degree of up-regulation was moderate with an average log fold change of one. Two tRNAs, tRNA-Gly-CCC and tRNA-Gly-GCC (and their 5′ fragments), exhibited the strongest up-regulation. In contrast, only six tRNAs were misepxresssed in *Mtrr*$^{gt/gt}$sperm, two down-regulated and four up-regulated ($p<0.05$, Figure 7.8 G, Table 7.4). Strikingly, 5′-tRNA-Tyr-GTA showed much more pronounced upregulation (1.89 log fold change, $p<0.0001$) than other misexpressed tRNAs.

I then performed differential expression analysis on tRNAs aligned using Salmon. Again, no differentially expressed tsRNAs were identified in *Mtrr*$^{+/+}$sperm (Figure 7.8 H). Interestingly, no differentially expressed tsRNAs

were identified in *Mtrr*$^{gt/gt}$sperm either (Figure 7.8 J). Again, tsRNAs were misexpressed in *Mtrr*$^{+/gt}$sperm ($p<0.05$, Figure 7.8 I). Two tsRNAs were down-regulated and eleven were up-regulated (Table 7.4). Notably, tRNA-Gly-CCC and tRNA-Gly-GCC again exhibited strong up-regulation and tRNA-Leu-CAA was down-regulated. This supports the results obtained using reads aligned with SPORTS1.0 (Figure 7.8 F). However, the tRNAs identified as misxpressed in sperm from *Mtrr*$^{+/gt}$males identified by differential expression analysis with reads aligned using Salmon and SPORTS1.0 were not entirely concurrent. This may reflect differences in the alignment algorithms used in the two approaches. Those tRNAs found to be misexpressed using both approaches should be given priority for future analysis. How the *Mtrr*$^{gt}$allele when heterozygous, but not when homozygous, leads to a greater alteration in sperm tsRNA profiles is unclear. However, the presence of misexpressed tsRNAs in sperm from *Mtrr*$^{+/gt}$males has important implications for transgenerational inheritance in the *Mtrr*$^{gt}$model.

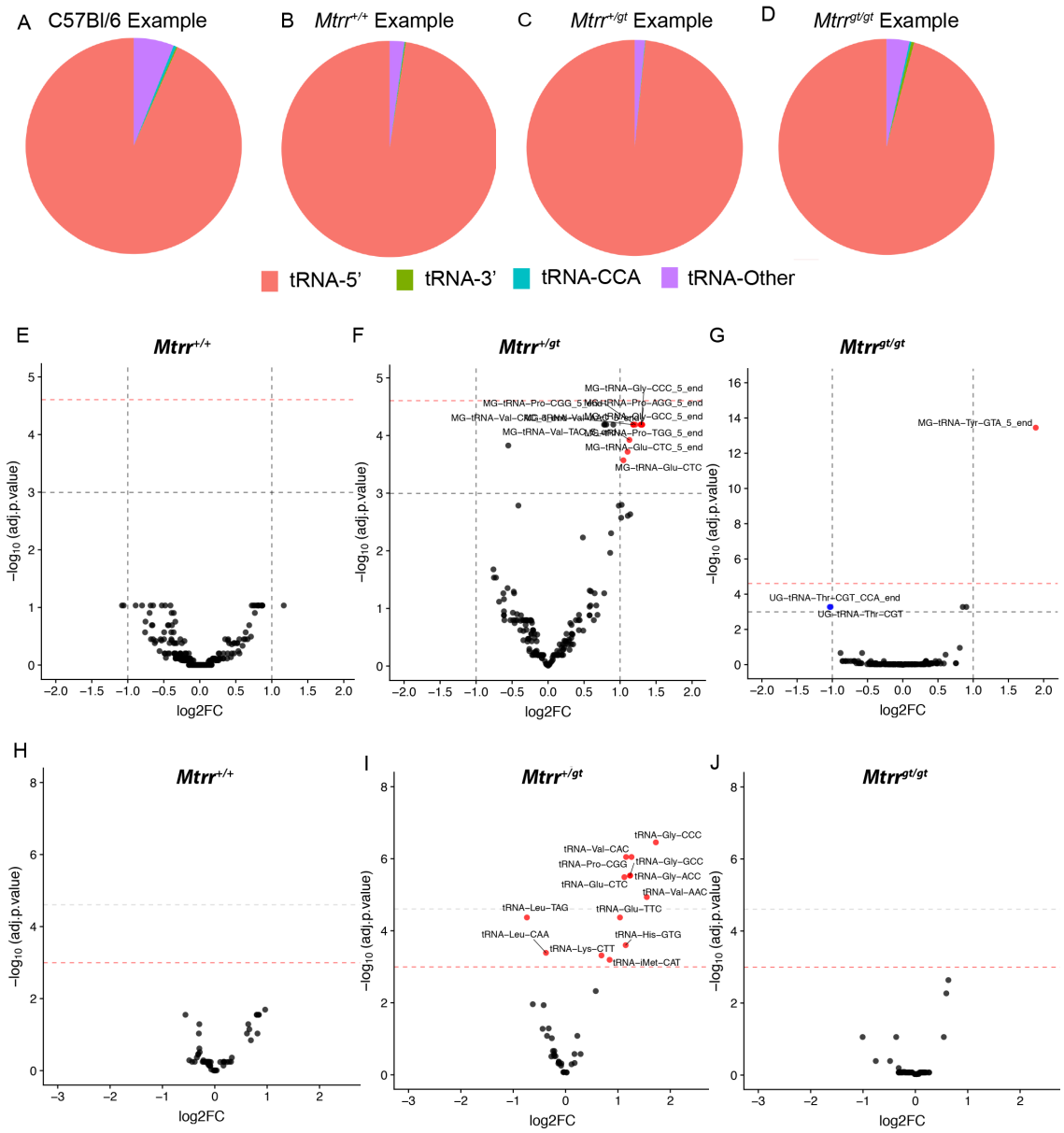



Figure 7.8: **Some tsRNAs are misexpressed in sperm from *Mtrr*$^{+/gt}$males. (A-D)** The distribution of loci from which the tsRNAs were derived; 5′ terminus (red), 3′ terminus (green), 3′CCA end (blue) or other (purple) is shown for tsRNAs in sperm from **(A)** C57Bl/6, **(B)** *Mtrr*$^{+/+}$, **(C)** *Mtrr*$^{+/gt}$and **(D)** *Mtrr*$^{gt/gt}$males. **(E-J)** Volcano plots are used to illustrate differential expression of tsRNAs (black dots) for **(E-G)** reads mapped using SPORTS1.0 in sperm from **(E)** *Mtrr*$^{+/+}$, **(F)** *Mtrr*$^{+/gt}$and **(G)** *Mtrr*$^{gt/gt}$males and for **(H-J)** reads mapped using Salmon in sperm from **(H)** *Mtrr*$^{+/+}$, **(I)** *Mtrr*$^{+/gt}$and **(J)** *Mtrr*$^{gt/gt}$males. (E-G) Grey dotted line $p< 0.05$, red dotted line $p<0.01$. (H-J) Red dotted line $p< 0.05$, grey dotted line $p<0.01$.

Table 7.4: **tRNAs misexpressed in *Mtrr*$^{+/gt}$ and *Mtrr*$^{gt/gt}$ sperm.**

| Approach | Male Genotype | tRNA Misexpressed | $Log_2$ Fold Change | Adjusted P-value |
|---|---|---|---|---|
| SPORTS1.0 | *Mtrr*$^{gt/gt}$ | 5′-tRNA-Tyr-GTA | 1.89 | 1.44E-06 |
| | | 3′-tRNA-Trp-CCA | 0.90 | 0.0378 |
| | | CCA-tRNA-Thr-CGT | -1.04 | 0.0378 |
| | | tRNA-Trp-CCA | 0.85 | 0.0378 |
| | | tRNA-Thr-CGT | -1.02 | 0.0378 |
| | | 5′-tRNA-Gly-TCC | 0.81 | 0.385 |
| SPORTS1.0 | *Mtrr*$^{+/gt}$ | 5′-tRNA-Gly-CCC | 1.29 | 0.0152 |
| | | 5′-tRNA-Gly-GCC | 1.31 | 0.0152 |
| | | 5′-tRNA-Lys-TTT | 0.91 | 0.0152 |
| | | 5′-tRNA-Pro-AGG | 1.30 | 0.0152 |
| | | 5′-tRNA-Pro-CGG | 1.21 | 0.0152 |
| | | 5′-tRNA-Pro-TGG | 1.31 | 0.0152 |
| | | 5′-tRNA-Val-AAC | 1.18 | 0.0152 |
| | | 5′-tRNA-Val-CAC | 1.18 | 0.0152 |
| | | tRNA-Gly-CCC | 1.28 | 0.0152 |
| | | tRNA-Gly-GCC | 1.30 | 0.0152 |
| | | tRNA-Pro-AGG | 0.79 | 0.0152 |
| | | tRNA-Pro-CGG | 0.81 | 0.0152 |
| | | tRNA-Pro-TGG | 0.78 | 0.0152 |
| | | tRNA-Val-AAC | 1.183 | 0.0152 |
| | | tRNA-Val-CAC | 1.18 | 0.0152 |
| | | 5′-tRNA-Val-TAC | 1.13 | 0.0198 |
| | | 5′-tRNA-Leu-CAA | -0.55 | 0.0218 |
| | | 5′-tRNA-Glu-CTC | 1.10 | 0.0243 |
| | | tRNA-Glu-CTC | 1.05 | 0.0281 |
| Salmon | *Mtrr*$^{+/gt}$ | tRNA-Gly-CCC | 1.72 | 0.0016 |
| | | tRNA-Pro-CGG | 1.16 | 0.0024 |
| | | tRNA-Val-CAC | 1.26 | 0.0024 |
| | | tRNA-Gly-ACC | 1.23 | 0.0040 |
| | | tRNA-Gly-GCC | 1.23 | 0.0040 |
| | | tRNA-Glu-CTC | 1.12 | 0.0041 |
| | | tRNA-Val-AAC | 1.55 | 0.0072 |
| | | tRNA-Glu-TTC | 1.04 | 0.0127 |
| | | tRNA-Leu-TAG | -0.74 | 0.0127 |
| | | tRNA-His-GTG | 1.15 | 0.0273 |
| | | tRNA-Leu-CAA | -0.37 | 0.0338 |
| | | tRNA-Lys-CTT | 0.69 | 0.0364 |
| | | tRNA-iMet-CAT | 0.84 | 0.0410 |

Approach indicates whether differential expression analysis was performed using reads aligned with SPORTS1.0 or Salmon

#### 7.2.4.2 miRNAs

miRNAs have been implicated in intergenerational inheritance following exposure to a range of environmental insults (Table 7.1) (Fullston et al., 2013; Grandjean et al., 2015; Gapp et al., 2014; Benito et al., 2018). I performed differential expression analysis on miRNAs aligned using SPORTS1.0. No differentially expressed miRNAs were identified in sperm from *Mtrr*$^{+/+}$, *Mtrr*$^{+/gt}$or *Mtrr*$^{gt/gt}$males (Figure 7.9). miRNA let7a-1 was downregulated in *Mtrr*$^{gt/gt}$sperm yet did not quite reach statistical significant (logFC -0.58, p=0.054). Differential expression analysis will be performed on miRNAs aligned using Salmon.

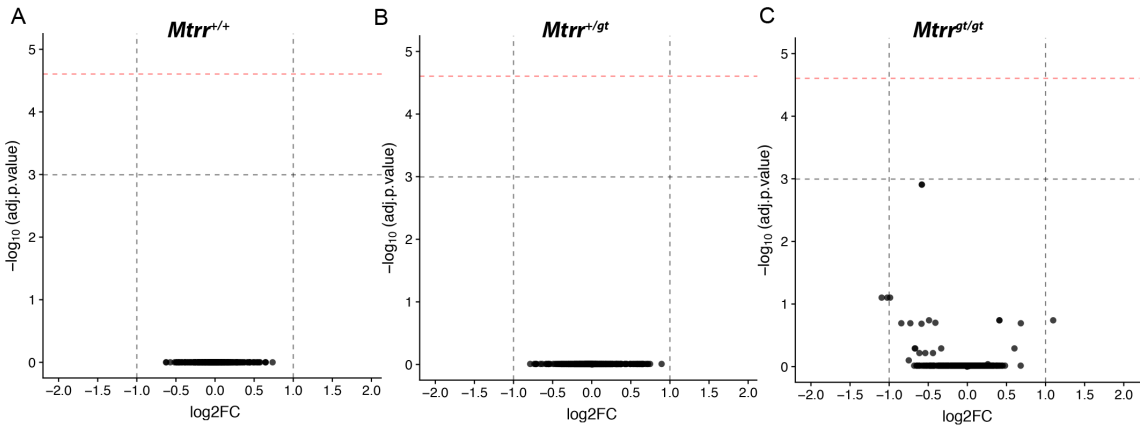



Figure 7.9: **No miRNAs are misexpressed in sperm from *Mtrr*$^{+/+}$, *Mtrr*$^{+/gt}$or *Mtrr*$^{gt/gt}$males. (A-C)** Volcano plots are used to illustrate differential expression of miRNAs (black dots) in sperm from **(A)** *Mtrr*$^{+/+}$, **(B)** *Mtrr*$^{+/gt}$and **(C)** *Mtrr*$^{gt/gt}$males. Reads mapped using SPORTS1.0. Grey dotted line p$<$ 0.05, red dotted line p$<$0.01.

#### 7.2.4.3 Other ncRNAs

RNAs that did not fall into tsRNA, miRNA or piRNA categories were also considered. These ncRNAs were referred to as "other ncRNAs" and were mapped to the Ensembl ncRNA database using Salmon. This database comprises all non-coding sequences in the genome (regardless of length) and partial alignments were allowed. A number of differentially expressed other ncRNAs were identified. Three other ncRNAs were misexpressed in *Mtrr*$^{+/+}$sperm. Two nuclear encoded rRNAs (n-R5s-109 and nR5s-29) were upregulated and Gm23840 was downregulated (Figure 7.10 A). In sperm from *Mtrr*$^{+/gt}$males,

seven other ncRNAs were misexpressed, three upregulated and four downregulated, although these were only just significant ($p<0.05$, Figure 7.10 B). These included two mitochondrial tRNAs (mt-tRNAs), three predicted genes and two RIKEN cDNAs. In sperm of $Mtrr^{gt/gt}$males, forty one other ncRNAs were misexpressed, seventeen downregulated and twenty four upregulated (Figure 7.10 C). Fourteen mt-tRNAs were up-regulated. The mt-tRNAs, mt-TS1 and mt-Tv were upregulated in both $Mtrr^{+/gt}$ and $Mtrr^{gt/gt}$ sperm and the 6430584L05Rik RIKEN cDNA was commonly down-regulated. mt-TS1 and mt-Tv are mitochondrial tRNAs for serine and valine, respectively and are encoded by the mitochondrial genome. Overall, the $Mtrr^{gt}$mutation, or parental exposure to the $Mtrr^{gt}$mutation, altered the other ncRNA profile of sperm. Intriguingly, the $Mtrr^{gt}$mutation appeared to particularly influence the expression of mt-tRNAs. This should be explored further given that inheritance of congenital malformations in the $Mtrr^{gt}$model goes down the maternal line (Padmanabhan et al., 2013). Differential expression analysis should be performed on other ncRNAs aligned using SPORTS1.0.

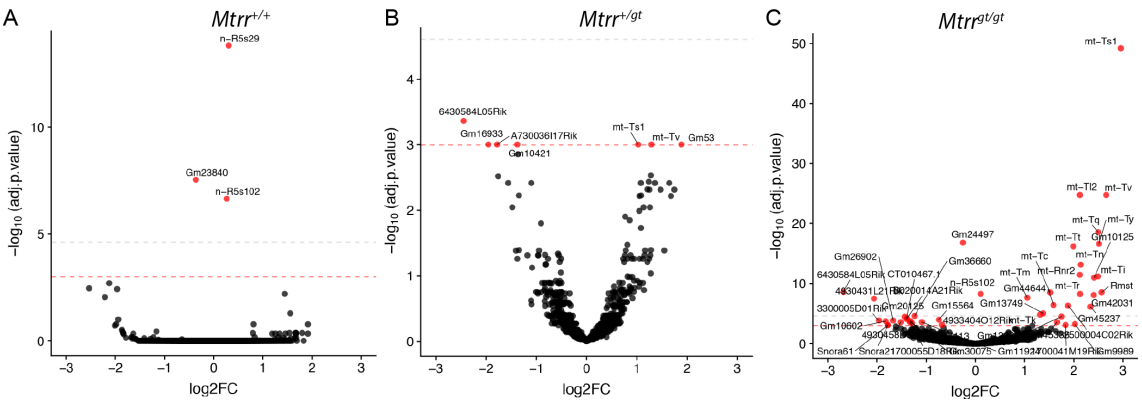



Figure 7.10: **ncRNAs are misexpressed in sperm from $Mtrr^{+/+}$, $Mtrr^{+/gt}$and $Mtrr^{gt/gt}$males. (A-C)** Volcano plots are used to illustrate differential expression of other ncRNAs (black dots) in sperm from **(A)** $Mtrr^{+/+}$, **(B)** $Mtrr^{+/gt}$and **(C)** $Mtrr^{gt/gt}$males. Red dotted line $p < 0.05$, grey dotted line $p<0.01$.

### 7.2.5 RNA modifications might be reduced in sperm from $Mtrr^{gt/gt}$ males

I observed that the expression of the RNA methyltransferase *Dnmt2* was reduced in $Mtrr^{+/+}$, $Mtrr^{+/gt}$and $Mtrr^{gt/gt}$epididymides compared to C57Bl/6 con-

trols (Chapter 3, Figure 3.7). Therefore I wanted to examine whether the modification of sncRNAs in sperm was disrupted in the *Mtrr*$^{gt}$model. Using the SPORTS1.0 analysis pipeline, sequence mismatch information can be used to identify potential RNA modifications (Shi et al., 2018). RNA modifications may lead to incorrect nucleotide incorporation during the reverse transcription step of the library preparation, leading to mismatches during read alignment. Whether all modifications would lead to errors during reverse transcription, and thus mismatches, is uncertain. Mismatch analysis calculates the percentage of unique sequences that contain significantly-enriched mismatches (either one, two or three mismatches) as a percentage total number of unique sequences (Shi et al., 2018). I performed mismatch analysis on tRNAs. Intriguingly, in *Mtrr*$^{gt/gt}$sperm there was a reduction in the number of mismatches (one, two and three) compared to C57Bl/6 controls (Figure 7.11). This suggested that tRNAs in *Mtrr*$^{gt/gt}$sperm carry fewer modifications than tRNAs in C57Bl/6 sperm. This correlated with the reduced *Dnmt2* expression observed in *Mtrr*$^{gt/gt}$epididymides (Chapter 3, Figure 3.7) and could be explained by the reduced methyl-group availability in *Mtrr*$^{gt/gt}$males for RNA methylation. However, the number of mismatches in sperm from *Mtrr*$^{+/+}$and *Mtrr*$^{+/gt}$males was equivalent to C57Bl/6 controls suggesting sperm tRNA modification levels were normal in *Mtrr*$^{+/+}$and *Mtrr*$^{+/gt}$male sperm. This was despite the reduced *Dnmt2* mRNA expression observed in *Mtrr*$^{+/+}$and *Mtrr*$^{+/gt}$epididymides. Overall, *Dnmt2* expression levels in the epididymis did not correlate with sperm inferred tRNA modification levels. This might suggest that protein expression and function of DNMT2 may be normal despite reduced transcription in *Mtrr*$^{+/+}$and *Mtrr*$^{+/gt}$epididymides. Alternatively, NSUN2, whose mRNA expression was normal in epididymides of *Mtrr*$^{+/+}$, *Mtrr*$^{+/gt}$and *Mtrr*$^{gt/gt}$males (Chapter 3, Figure 3.7), may be able to compensate for decreased DNMT2 in *Mtrr*$^{+/+}$and *Mtrr*$^{+/gt}$males but not *Mtrr*$^{gt/gt}$males. This would ensure RNA modifications were maintained in *Mtrr*$^{+/+}$and *Mtrr*$^{+/gt}$male sperm. Mismatch analysis should also be performed on miRNAs, piRNAs and rRNAs.

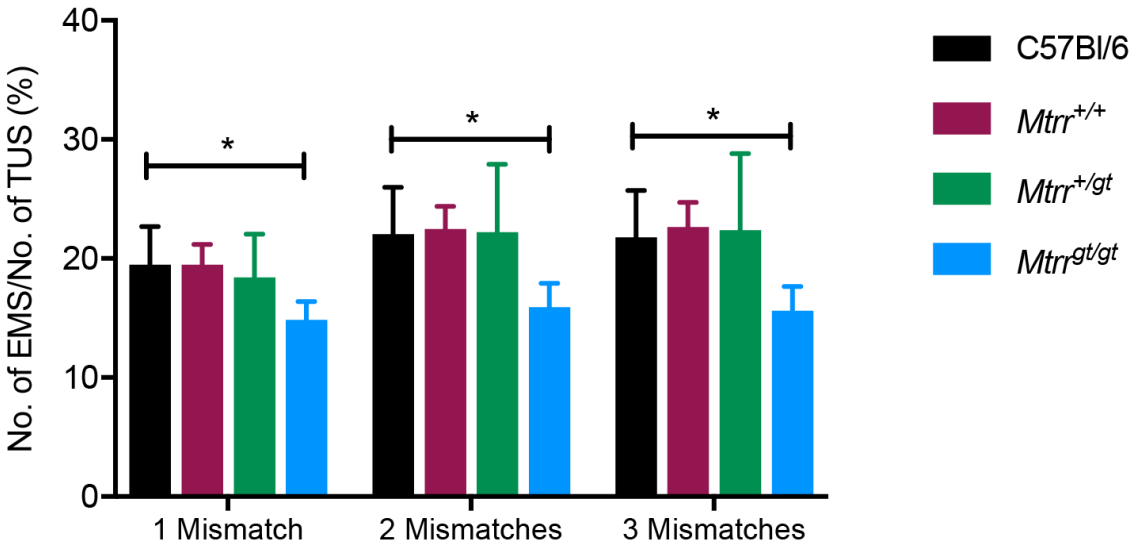



Figure 7.11: **tsRNAs mismatches in sperm from C57Bl/6, *Mtrr*$^{+/+}$, *Mtrr*$^{+/gt}$ and *Mtrr*$^{gt/gt}$ males.** The percentage of unique tsRNA sequences that contain significantly-enriched mismatches (EMS) as a percentage of the total number of unique sequences (TUS) is plotted for either one, two or three mismatches in tsRNAs identified in sperm from C57Bl/6 (black), *Mtrr*$^{+/+}$(purple), *Mtrr*$^{+/gt}$(green) and *Mtrr*$^{gt/gt}$(blue) males. One-way ANOVAs, with Dunnett's multiple comparisons test performed, * $p < 0.05$.

## 7.3 Discussion

Here I report that the overall proportions of sncRNA species in sperm were unaffected by the *Mtrr*$^{gt}$mutation or exposure to a parental *Mtrr*$^{gt}$allele. The abundance of rRNA varied considerably between samples, regardless of male genotype. This may reflect natural biological variability in rRNA expression or could be a technical artefact. I identified differential expression of some sncRNAs in sperm from *Mtrr*$^{+/+}$, *Mtrr*$^{+/gt}$and *Mtrr*$^{gt/gt}$males. Intriguingly, tRNAs were strongly misexpressed in *Mtrr*$^{+/gt}$males. While only a few tRNAs were misxpressed in *Mtrr*$^{gt/gt}$sperm, there appeared to be an overall reduction in the levels of RNA modifications on tRNAs in *Mtrr*$^{gt/gt}$sperm. Additionally, in sperm from *Mtrr*$^{gt/gt}$ males mt-tRNAs were particularly upregulated. Overall, while still incomplete, this analysis suggests that the *Mtrr*$^{gt}$mutation or exposure to a parental *Mtrr*$^{gt}$allele, can impact sncRNA populations in sperm. Those sncRNA misexpressed in *Mtrr*$^{+/gt}$males have the potential to contribute to the transgenerational inheritance of congenital malformations observed in the *Mtrr*$^{gt}$model.

Further bioinformatic analysis of this sncRNA-seq data is required. Firstly, I will perform differential expression analysis for piRNAs. Furthermore, I will complete gene ontology analysis to assess whether sncRNA species misexpressed in *Mtrr* sperm are associated with any particular biological processes that may be relevant to the inheritance of congenital malformations in the *Mtrr*$^{gt}$model. I would like to explore if misexpressed sncRNAs that map to the reference genome correlate with underlying genetic or DNA methylation changes that could account for their misexpression. The misexpression of sncRNA species in sperm should be independently validated using alternative techniques. RT-qPCR can be used to validate miRNAs. A miRNA identified as equivalently expressed between genotypes by sncRNA-seq can be used for normalisation of Ct values (Gapp et al., 2014). Validation of other ncRNAs that align to the reference genome could similarly be performed using qPCR. However, validation of tRNA misexpression is more complex and as such is often not performed (Sharma et al., 2016; Chen et al., 2016a). Northern blotting could be used for tsRNA validation (Pekarsky et al., 2016). Once validated, functional studies can be performed.

Unlike other next-generation sequencing data analysis, there is no 'gold-standard' analysis pipeline for sperm sncRNA-seq data. For that reason I used two parallel approaches to categorise and align sncRNA reads; 1) using the recently published SPORTS1.0 pipeline (Shi et al., 2018) and 2) mapping reads using Salmon (Patro et al., 2017). I then aimed to perform differential expression analysis twice for each sncRNA species, once using reads aligned using SPORTS1.0 and and once using reads aligned using Salmon. Both analyses have yet to be performed for most sncRNA species. However, this analysis was performed for tRNAs. While some tRNAs were identified as differentially expressed in both analyses, some discrepancies between the findings were noted. I should perform a systematic comparison of the findings from each approach. A third, newly published sncRNA analysis pipeline, TEsmall (O'Neill et al., 2018), could also be used to perform sncRNA alignment. Comparing my findings thus far with the findings from those generated using TEsmall may help indicate which of two original approaches was superior, provided TEsmall gives results I feel are reliable.

The variability in abundance of rRNA within the sperm ncRNA libraries had the ability to massively distort differential expression analysis. As ncRNA read mapping is performed against individual ncRNA subtype databases, the relative abundance of each ncRNA species needs to be taken into account to avoid skewed differential expression analysis. For example, in those samples with a high abundance of rsRNA, other sncRNAs would appear less abundant based on a comparison to a sample with the same total read count but less abundant rsRNA. Therefore, rsRNA reads should be removed from all sncRNA-seq libraries and the remaining reads renormalised to the rsRNA depleted total read count. In this way the relative abundance of each sncRNA species and the total number of reads per library is accounted for. All subsequent differential expression analysis should be performed on rsRNA depleted renormalised libraries. However, this was not performed for the differential expression analysis I reported (Section 7.2.4). Interestingly, tRNAs were found to be misexpressed much more highly in sperm from $Mtrr^{+/gt}$males than in sperm from $Mtrr^{gt/gt}$males (Figure 7.8). Additionally there was no overlap between the tsRNAs misexpressed in $Mtrr^{+/gt}$and $Mtrr^{gt/gt}$males. This seems surprising

if tsRNA expression is a result of the *Mtrr*$^{gt}$allele. However, sncRNA libraries for *Mtrr*$^{+/gt}$male sperm had the lowest proportion of rRNAs (Figure 7.4). As differential expression analysis was not performed on rRNA depleted and renormalised libraries, the finding of upregulation of tsRNAs in *Mtrr*$^{+/gt}$sperm may be an artefact rather than a true biological phenomenon.

The overall sncRNA profiles of sperm I describe here are similar to those reported in other studies (Shi et al., 2018; Zhang et al., 2018; Gapp et al., 2018). Interestingly I report tRNA-Gly-CCC and tRNA-Gly-GCC as some of the most strongly upregulated tsRNAs, regardless of the read alignment approach prior to differential expression analysis (Figure 7.8). These specific tsRNAs were also strongly upregulated as a result of low-protein diet (Sharma et al., 2016). This might suggest that these tsRNAs are particularly responsive to adverse conditions in sperm, regardless of cause. However, there were also differentially expressed sncRNAs identified in sperm from *Mtrr*$^{+/gt}$and *Mtrr*$^{gt/gt}$males, e.g. 5'-tRNA-Tyr-GTA, that have not been reported in other epigenetic inheritance paradigms suggesting model specific effects.

I observed that tRNAs from *Mtrr*$^{gt/gt}$males likely have fewer RNA modifications than C57Bl/6 controls. However, it appeared that the levels of RNA modification were normal on *Mtrr*$^{+/+}$and *Mtrr*$^{+/gt}$sperm tRNAs compared to controls. The expression of *Dnmt2* was reduced in *Mtrr*$^{+/+}$, *Mtrr*$^{+/gt}$and *Mtrr*$^{gt/gt}$epididymides compared to C57Bl/6 controls (Chapter 3, Figure 3.7). It will be necessary to confirm by western blotting that the decreased *Dnmt2* mRNA levels leads to a decrease in DNMT2 protein. The mismatch analysis provided by SPORTS1.0 does not provide any information regarding what the RNA modifications themselves are. The decrease in RNA modifications observed in sperm from *Mtrr*$^{gt/gt}$males may not result from decreased RNA methylation, applied by DNMT2, but from a decrease in the abundance of other RNA modifications e.g. acetylation (Frye et al., 2016). Lipid chromatography-tandem mass spectroscopy (LC-MS/MS) analysis in combination with antibody pull-down methods should be performed to identify which specific sperm ncRNA modifications are less abundant (Zhang et al., 2018). RNA modifications can alter RNA stability, allowing them to persist for longer in the post-fertilisation embryo, and could alter their 3D structure and interaction capacity

(Chen et al., 2016a). As such altered RNA modification profiles could have just as many functional consequences with respect to phenotypic inheritance as differential sncRNA expression.

Fundamentally, I would like to assess if misexpression of sperm sncRNAs is functionally relevant to the inheritance of phenotypes in the *Mtrr*$^{gt}$model. As a first step to address this I will assess if sncRNAs misexpressed in sperm from *Mtrr*$^{+/gt}$males are also misexpressed in *Mtrr*$^{+/+}$offspring tissues. The expression of any genes known to be regulated by misexpressed sncRNAs could also be assessed by RT-qPCR in *Mtrr*$^{+/+}$offspring tissues, including early embryos and embryos at E10.5, over several generations. Secondly I would like to perform microinjection studies using sncRNAs derived from *Mtrr*$^{+/gt}$male sperm. sncRNAs from *Mtrr*$^{+/gt}$sperm (or a size selected portion of total sperm ncRNAs) would be injected into a fertilised C57Bl/6 oocytes and the resultant blastocysts transferred to a recipient pseudopregnant female. Firstly I would assess phenotypes at E10.5 of injected embryos. This would indicate if the injected sncRNAs are directly able to cause embryonic phenotypes. Subsequently, I would then take adult females derived by microinjection and cross them to C57Bl/6 males. I would assess if congenital malformations are present in the litters of the microinjection derived females at E10.5. If present, this would suggest that sperm sncRNAs are sufficient to facilitate transgenerational inheritance in the *Mtrr*$^{gt}$model. It may also been interesting to perform transcriptomic analysis of F1 early embryos (pre-blastocyst) following microinjection to determine if *Mtrr*$^{+/gt}$sperm sncRNAs can induced transcriptional changes. These microinjection studies could also be performed using synthetic ncRNAs which may provide information regarding the importance of RNA modifications. Microinjection studies are however not without confounding factors. Fertilised oocytes already contain sperm RNAs, thus sperm RNAs injected in are over and above the RNA payload normally present. This may influence the activity and outcome of these sncRNAs. Ideally, one would like to restore normal expression of misexpressed RNAs in sperm. Offspring phenotypes should be rescued in offspring derived from sperm in which aberrant ncRNA expression has been corrected. This experiment would be technically challenging to perform. An

alternative, but no less complex approach, might be to generate synthetic germ cells, in which the pool of sncRNAs can be tightly controlled (Zheng 2015).

Overall, the data presented in this chapter demonstrate that abnormal folate metabolism could disrupt germline sncRNA profiles. Together with my previous finding that sperm DNA methylation profiles are also altered by the *Mtrr*$^{gt}$mutation (Chapter 5), this suggest that the sperm epigenome is highly susceptible to abnormal folate metabolism. Whether misexpression of sperm sncRNAs contributes to the inheritance of congenital malformations in the *Mtrr*$^{gt}$mice remains to be investigated.

# Chapter 8

# Discussion

## 8.1 Summary

In this thesis I set out to determine the mechanism behind transgenerational inheritance in *Mtrr*$^{gt}$mice. I hypothesised that gametic epigenetic inheritance was responsible for the congenital malformations observed in the wildtype grandprogeny of a *Mtrr*$^{+/gt}$maternal grandparent. However, I also considered whether genetic instability resulting from the *Mtrr*$^{gt}$mutation or abnormal germ cell morphology may also contribute to the transgenerational inheritance previously reported (Padmanabhan et al., 2013). I observed that testes morphology, spermatogenesis and sperm parameters were normal in *Mtrr*$^{+/gt}$males (Chapter 3). As a result I conclude that general sperm abnormalities are unlikely to be a factor responsible for TEI in the *Mtrr*$^{gt}$model. I observed that the frequency of SVs and SNPs identified in *Mtrr*$^{gt/gt}$embryos was similar to C57Bl/6 controls (Chapter 4). Therefore, I feel it is unlikely that genetic instability is a factor responsible for the inheritance of phenotypes in the *Mtrr*$^{gt}$model. However, I cannot completely exclude that *Mtrr*$^{gt}$line specific genetic variants (SVs and SNPs) or genetic background effects associated with the 129/P2 DNA at the *Mtrr* locus contribute to the phenotypes observed in the *Mtrr*$^{gt}$model (Chapter 4).

I demonstrated that intrinsic or parental exposure to the *Mtrr*$^{gt}$allele was sufficient to disrupt the sperm methylome. DNA methylation patterns were altered in sperm from *Mtrr*$^{+/+}$, *Mtrr*$^{+/gt}$ and *Mtrr*$^{gt/gt}$ males (Chapter 5). The majority of DMRs were not associated with known genetic variants (SVs or SNPs). Differential methylation identified in sperm from *Mtrr*$^{+/gt}$males was not maintained in a range of F1 and F2 generation tissues (Chapter 6). This indicated that DNA methylation may not have a role in conferring long-term memory of abnormal folate metabolism. Generally, the expression of genes near sperm DMRs was equivalent to C57Bl/6 controls in offspring tissues. However, I observed tissue specific downregulation of *Hira* and *Rn45s* in F2 and F3 generation embryos (Chapter 6). This suggested that differential DNA methylation at sperm DMRs may function as an initiator or demarcator of gene misexpression at specific loci. I also observed dysregulation of several sncRNA species including tsRNAs in *Mtrr*$^{+/gt}$sperm (Chapter 7). Whether these contribute to the inheritance of congenital malformations and epigenetic insta-

bility in the *Mtrr*$^{gt}$model remains to be explored. Furthermore, downregulation of *Dnmt2* in the epididymis of *Mtrr*$^{+/gt}$males suggested a role for sperm RNA methylation in the *Mtrr*$^{gt}$model (Chapter 3). Overall, I have yet to fully elucidate the mechanism of TEI in the *Mtrr*$^{gt}$model. However, I have identified plausible epigenetic candidates, have likely ruled out genetic instability and abnormal sperm morphology as contributing factors and have begun to explore potential genetic effects. This brings us closer to understanding of the molecular processes underpinning the phenomenon.

## 8.2 Hypotheses for transgenerational inheritance in the *Mtrr*$^{gt}$ model

Here I present several hypotheses for mechanisms that may be responsible for transgenerational inheritance in the *Mtrr*$^{gt}$model based on the findings and conclusions presented in this thesis. Indeed, it may be that a contribution from multiple mechanisms, both genetic and epigenetic, is required in order to bring about the complex and variable phenotypes observed in *Mtrr*$^{gt}$mice.

### 8.2.1 Genetic background and transgenerational inheritance in the *Mtrr*$^{gt}$model

I acknowledge that genetic background effects could contribute to the inheritance of phenotypes over multiple generations in *Mtrr*$^{gt}$mice. I observed that an approximately 20 Mb region flanking the *Mtrr* gene was likely a 129/P2 chromosomal segment. This is a relic of the 129/P2 ESCs from which the *Mtrr*$^{gt}$line was derived prior to backcrossing into the C57Bl/6 genetic background. I show that mRNA expression of a small subset of genes at the *Mtrr* locus is generally not altered in *Mtrr*$^{gt/gt}$embryos (Chapter 4). However, 129/P2 variants may not affect transcription of these genes but may alter splicing, translation or protein function. Therefore, with the analysis performed so far I cannot confidently exclude that this region does not play a role in phenotypic inheritance in the *Mtrr*$^{gt}$model.

Nevertheless, there is evidence that the *Mtrr* locus may not be functionally relevant to the inheritance of phenotypes observed in the *Mtrr*$^{gt}$model. Beyond

the F1 generation of the *Mtrr*$^{+/gt}$maternal grandparent pedigrees all individuals are wildtype (Chapter 1, Figure 1.6). These individuals will have C57Bl/6 genetic background at the *Mtrr* locus. Additionally, there was no significant abnormalities in the F1 generation hybrid conceptuses at E10.5 derived from the reciprocal crosses between C57Bl/6 and 129/P2 mice (N. Padmanabhan and E. Watson, unpublished data). However, an interaction between the *Mtrr*$^{gt}$mutation itself and the 129/P2 DNA surrounding it in the F0 generation may be required for transgenerational inheritance to occur in the *Mtrr*$^{gt}$model.

It has long been acknowledged that genetic background can influence phenotypes in genetically modified mice (Doetschman, 2009). Abnormal folate metabolism resulting from genetic mutations in folate metabolism enzymes can lead to differing phenotypes and/or phenotypic penetrance dependent on genetic background. For example there is a variable testes phenotype in BALB/c and C57Bl/6 *Mthfr*$^{-/-}$ mice (Chan et al., 2010). Genetic background effects have also been reported to have an influence in many epigenetic inheritance paradigms, for example in the inheritance of the *Axin-fused* and *A*$^{vy}$ metastable epialleles (Rakyan et al., 2003) and following *in-utero* vinclozolin exposure in mice (Guerrero-Bosagna et al., 2012). Perhaps particularly relevant to the *Mtrr*$^{gt}$model are those epigenetic inheritance paradigms in which a genetic mutation that instigates intergenerational inheritance was generated in one genetic background prior to backcrossing into another, as in *Mtrr*$^{gt}$mice. The *A*$^{vy}$ mouse line was founded from an individual of C3H/HeJ genetic background over half a century ago and extensively backcrossed to C57Bl/6 (Dickies, 1962). While sequencing of littermates was used to exclude genetic effects contributing to the *A*$^{vy}$ phenotype, variants at the C3H/HeJ region were excluded from this analysis (Oey et al., 2015). This region should be explored more closely in *A*$^{vy}$ mice. More recently, *Dnmt2* was linked to intergenerational transmission of metabolic disorder following high-fat diet exposure in mice. The inheritance of phenotypes was rescued in *Dnmt2*$^{-/-}$ mice. However, these mice were generated in a mixed 129X1:C57Bl/6 genetic background and were backcrossed for only five generations into the C57BL/6NCrl strain (Zhang et al., 2018). More extensive backcrossing of this line would need to be performed to ensure that genetic background at the *Dnmt2* allele is not contributing to the phenotypic

rescue. Overall, a greater understanding of the mechanisms of inheritance in these models may come from exploring potential genetic background effects further.

*Mtrr*$^{gt}$mice provide an excellent model in which to probe genetic background effects, and in doing so tackle an area long neglected in both the folate metabolism and TEI fields. To conclusively rule out a genetic background effect I would need to derive a new *Mtrr* mouse line in a pure isogenic C57Bl/6 background using CRISP-cas9. Only if transgenerational inheritance was observed in this mouse line could I confirm with certainty that a genetic background effect was not a factor in the *Mtrr*$^{gt}$model. It would also be valuable to determine if supplementation of the diet of *Mtrr*$^{+/gt}$mice with methionine (from conception to breeding) was sufficient to lead to an absence of congenital malformations in their *Mtrr*$^{+/+}$grandprogeny. This would confirm that it is the disruption of folate metabolism specifically that is responsible for the phenotypic inheritance.

### 8.2.2 Genetic variants unique to the *Mtrr*$^{gt}$colony and transgenerational inheritance

I also identified some genetic variants unique to the *Mtrr*$^{gt}$colony. These could contribute to the TEI observed in *Mtrr*$^{gt}$mice. The genetic variants unique to the *Mtrr*$^{gt}$colony are likely *de novo* mutations that have arisen in the *Mtrr*$^{gt}$line or mutations present in the C57Bl/6 mice used for backcrossing the *Mtrr*$^{gt}$allele. Backcrossing of the *Mtrr*$^{gt}$line was performed at the University of Calgary, Canada. The *Mtrr*$^{gt}$colony was moved to the UK and a new C57Bl/6 control colony was bought in. Thus the C57Bl/6 mice used as controls and those used for backcrossing the *Mtrr*$^{gt}$allele are not genetically identical. Ideally, the *Mtrr*$^{gt/gt}$genome should have been compared to whole genome sequencing data from the original C57Bl/6 mice used for backcrossing when determining if the *Mtrr*$^{gt}$allele led to genetic instability. However, since the *Mtrr*$^{gt}$line was established and backcrossed over fifteen years ago some natural genetic divergence may have occurred complicating this analysis if performed now. Furthermore, it will be important to sequence the genomes of *Mtrr*$^{+/+}$individuals to determine if genetic variants thought to be common within the *Mtrr*$^{gt}$colony, such as

the tandem duplication on chromosome 19, are indeed present in *Mtrr*$^{+/+}$mice (Chapter 4). Validating the presence of common *Mtrr*$^{gt}$variants in *Mtrr*$^{+/+}$tissues using locus specific approaches, such as Sanger sequencing, southern blotting or dosage qPCR, could also be used if a whole-genome approach were prohibitively expensive. These approaches could also be used to retrospectively genotype frozen F1 generation maternal tissues for *Mtrr*$^{gt}$specific variants. The presence or absence of these variants in the mother could then be correlated with the severity of phenotypes in her *Mtrr*$^{+/+}$F2 generation offspring to give an indication as to whether these variants contribute to the phenotypes observed in the *Mtrr*$^{gt}$model.

### 8.2.3 sncRNA mediated TEI and the *Mtrr*$^{gt}$model

I hypothesise that misexpression of sncRNAs could contribute to the inheritance of phenotypes over multiple generations in *Mtrr*$^{gt}$mice. *Mtrr* deficiency leads to misexpression of tsRNAs and other ncRNAs in sperm from *Mtrr*$^{+/gt}$males (Chapter 7, Figures 7.8 and 7.10). Some of those tsRNAs that are misexpressed in *Mtrr*$^{+/gt}$male sperm have been shown to lead to phenotypic inheritance in mice fed a low-protein diet (Sharma et al., 2016). Furthermore I observed a decrease in the level of RNA modification on tRNAs in sperm from *Mtrr*$^{gt/gt}$males (Chapter 7, Figure 7.11) and downregulation of *Dnmt2* in the epididymis of *Mtrr*$^{+/+}$, *Mtrr*$^{+/gt}$and *Mtrr*$^{gt/gt}$males (Chapter 3, Figure 3.7). This suggests that sncRNA modifications may also play a role in the inheritance of phenotypes in the *Mtrr*$^{gt}$model.

sncRNAs, potentially with abnormal modifications, might directly regulate the expression or activity of genes in the post-fertilisation embryo. For example, tRNA fragments have been shown to repress genes associated with the endogenous retroelement MERVL in ESCs and zygotes (Sharma et al., 2016). A bioinformatic analysis could be performed to investigate if any sncRNAs misexpressed in *Mtrr*$^{+/gt}$sperm could play a role in regulating gene expression. Candidates could be further investigated with approaches such as RNA interference, e.g. interfering with tRNA function using locked nucleic acid containing oligonucleotides (Sharma et al., 2016), or RNA-fluorescent *in-situ* hybridisation to demonstrate RNA-RNA interaction. It will be exciting to

determine if there is an interaction between sncRNAs and gene misexpression in the *Mtrr*$^{gt}$model.

Alternatively sncRNAs may act in concert with other epigenetic mechanisms to facilitate inheritance over multiple generations. Therefore, analysis of histone modifications in sperm of the *Mtrr*$^{gt}$model will be essential. A low-protein diet was shown to alter H3K27me3 levels and tsRNAs in sperm (Carone et al., 2010; Sharma et al., 2016). Furthermore, an interaction between miRNAs and histone modifications is observed in an intriguing paramutant example. miR-124 injection into fertilised oocytes led to overgrowth of the offspring and increased expression of *Sox9* which persisted into the F2 generation (Grandjean et al., 2009). This coincided with a distinct pattern of histone modifications being established at the *Sox9* promoter (Grandjean et al., 2009). This suggests that sncRNA can guide chromatin remodelling in embryos. Therefore, the chromatin landscape of *Mtrr*$^{+/+}$F1 and F2 generation embryos should also be investigated on a genome-wide scale to explore if sncRNA directed chromatin remodelling is important for phenotypic inheritance in the *Mtrr*$^{gt}$model.

### 8.2.4 A hypothesis for HIRA mediated epigenetic instability and TEI in the *Mtrr*$^{gt}$model

I observed tissue specific downregulation of *Hira* in *Mtrr*$^{+/+}$F2 and F3 generation embryos (Chapter 6, Figure 6.6). HIRA is a histone chaperone for the histone variant H3.3 (Tang et al., 2006). In mice, in the post-fertilisation embryo, maternally supplied HIRA is required for H3.3 incorporation into the paternal pronucleus (Torres-Padilla et al., 2006; Lin et al., 2014). Depletion of HIRA in oocytes leads to reduced fertility and chromatin structure abnormalities (Nashun et al., 2015). There is remarkable similarity between *Hira* knockout phenotypes and those seen in the *Mtrr*$^{gt}$model (Padmanabhan et al., 2013; Roberts et al., 2002). Therefore, I propose that HIRA may play an important role in the *Mtrr*$^{gt}$model.

Here I put forward a hypothesis for how HIRA may be involved in the epigenetic dysregulation and inheritance of congenital malformations observed in *Mtrr*$^{+/gt}$maternal grandfather pedigree (Figure 8.1). Sperm from *Mtrr*$^{+/gt}$ males have a DMR downstream of the *Hira* gene. Assuming sperm are tran-

scriptionally silent, *Hira* would not be expressed in sperm. In the C57Bl/6 oocytes, *Hira* would be expressed normally, be entirely functional and would have no abnormal epigenetic patterns associated with it (Figure 8.1). *Mtrr*$^{+/+}$F1 blastocysts are derived from *Mtrr*$^{+/gt}$sperm and C57Bl/6 oocytes. As HIRA and H3.3 in the zygote are maternally derived, H3.3 incorporation into the paternal pronucleus post-fertilisation should occur normally in the F1 blastocyst. This would lead to normal development in *Mtrr*$^{+/+}$F1 embryos derived from *Mtrr*$^{+/gt}$males (Figure 8.1). This hypothesis therefore provides a plausible explanation for the absence of phenotypes in the F1 generation of the maternal grandfather pedigree which has been a perplexing feature of the *Mtrr*$^{gt}$model (Padmanabhan et al., 2013). Phenotypes are present in the F1 generation in the *Mtrr*$^{+/gt}$maternal grandmother pedigree (Padmanabhan et al., 2013). This would be predicted by this hypothesis owing to dysregulation of HIRA in the F0 oocyte. I hypothesise *Mtrr*$^{+/gt}$male derived epimutations (DNA methylation, ncRNAs, histones etc.) may lead to epigenetic dysregulation in F1 embryos. This could drive abnormal *Hira* expression in F1 generation embryos, including in the developing PGCs (Figure 8.1). This would potentially lead to *Hira* misexpression, abnormal HIRA activity and chromatin defects in F1 oocytes. *Mtrr*$^{+/+}$F2 blastocysts are derived from these F1 oocytes and C57Bl/6 sperm (Figure 8.1). Following fertilisation, maternal HIRA abnormalities might lead to aberrant H3.3 incorporation into the paternal pronucleus. This may lead to epigenetic dysregulation, alterations in histone modification patterns and/or abnormal pericentromeric chromatin formation (Szenker et al., 2011). Consequently there may be transcriptional changes and an increased frequency of congenital malformations in the F2 embryos. Phenotypic variability could result from stochastic differences in the location of H3.3 (mis)incorporation into the paternal pronucleus as a result of abnormal HIRA activity or from variable HIRA insufficiency in F1 oocytes. A similar scenario would play out in the F3 generation and beyond.

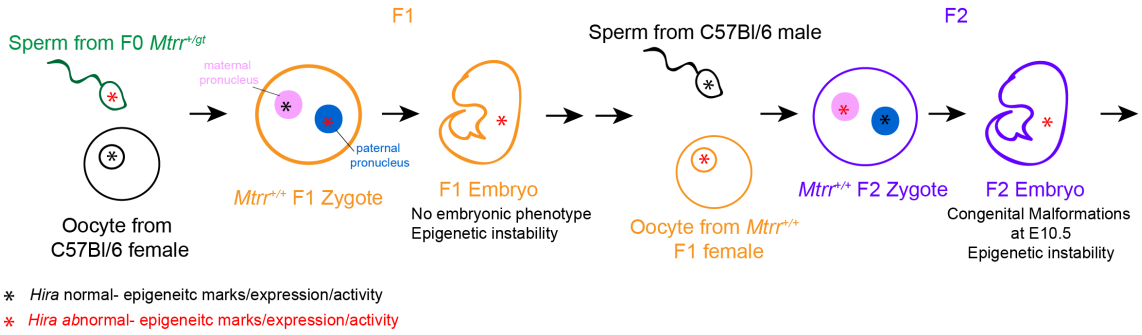



Figure 8.1: **A schematic of our hypothesis for the role of HIRA in trans-generational inheritance of phenotypes in the *Mtrr$^{gt}$* model.** *Mtrr*$^{+/+}$F1 blastocysts (orange) are derived from *Mtrr*$^{+/gt}$male sperm (green) and C57Bl/6 oocytes (black). The *Mtrr*$^{+/gt}$sperm carry the *Hira* DMR (demarcated by a red asterisks). In C57Bl/6 oocytes, HIRA is epigenetically, transcriptionally and functionally normal (demarcated by a black asterisks). As HIRA and H3.3 are maternally derived, H3.3 incorporation into the paternal pronucleus (blue) post-fertilisation occurs normally in the F1 blastocyst. However, *Mtrr*$^{+/gt}$male derived epimutations (DNA methylation, ncRNAs, histones etc.) could lead to epigenetic dysregulation and could drive abnormal *Hira* expression and activity in F1 generation embryos. This would lead to abnormal *Hira* expression and chromatin defects in the F1 oocytes. *Mtrr*$^{+/+}$F2 blastocysts (purple) are derived from these F1 oocytes and C57Bl/6 sperm. Following fertilisation, maternal HIRA abnormalities would lead to aberrant H3.3 incorporation into the paternal pronucleus. This may lead to an increased frequency of congenital malformations and epigenetic instability in the F2 embryos.

However, the hypothesis does not entirely fit the data. I did not see a decrease in *Hira* expression in F1 embryos, although *Hira* expression was highly variable (Chapter 6, Figure 6.6 K). *Hira* methylation also appeared variable in sperm (Chapter 6, Figure 6.6 B) although this needs to be verified. One could speculate that the variability in *Hira* expression in F1 embryos may reflect variability in the methylation of the *Hira* DMR in sperm. A correlation between sperm methylation and F1 embryo *Hira* expression could be assessed using father-daughter pairs. *Hira* misexpression in F1 embryos may occur in a tissue specific manner, thus may not be observed using whole embryo analysis. Additionally, *Hira* mRNA was upregulated in *Mtrr*$^{gt/gt}$embryos (Chapter 6, Figure 6.6 J). This may result from differences in the epigenetic dysregulation resulting from intrinsic *Mtrr*$^{gt}$allele exposure compared to parental or ancestral *Mtrr*$^{gt}$allele exposure. Intriguingly, *Hira* expression was down-regulated in

both phenotypically normal F2 and F3 embryos as well as those with congenital malformations. This suggests that while these embryos appear phenotypically normal they may have a degree of molecular abnormality. Indeed, phenotypically normal male F2 embryos at E10.5 derived from a *Mtrr*$^{+/gt}$maternal grandfather weighed less than C57Bl/6 controls despite appearing overtly normal (Padmanabhan et al., 2017). Furthermore, embryos that are phenotypically normal at E10.5 may present with phenotypes later in gestation. Indeed, foetal and placental phenotypes are observed at E14.5 and E18.5 (E. Watson, personal communication). However, this finding may also preclude that the misexpression of *Hira* is solely responsible for the embryonic phenotypes observed. *Hira* misexpression in concert with other epigenetic factors or against a specific molecular or epigenetic backdrop may act to bring about the presence of congenital malformations in the *Mtrr*$^{gt}$model. Furthermore, there is little evidence to date for epigenetic instability in *Mtrr*$^{+/+}$F1 or F2 generation embryos (Padmanabhan et al., 2013). Global DNA methylation levels, determined using a DNA methylation quantification kit based method, were normal in *Mtrr*$^{+/+}$F2 embryos despite global hypomethylation in placentas (Padmanabhan et al., 2013). However, tissue-specific epigenetic instability within the embryo may be masked using whole embryo analysis. More accurate and sensitive techniques for assessing global DNA methylation levels (e.g. mass spectrometry) and genome-wide high-resolution approaches (e.g. whole-genome bisulfite sequencing) may yet reveal dysregulation of DNA methylation in embryos.

A substantial amount of further experimentation will be required to test this hypothesis in the maternal grandfather pedigree. Firstly I need to consider the F0 generation. Verification of the A10 DMR, downstream of *Hira*, will be required in *Mtrr*$^{+/gt}$sperm from additional individuals due to the variability in CpG methylation levels previously observed. I also need to confirm if *Hira* is indeed not transcribed in sperm. If expressed, there may be a role for HIRA in regulating sperm chromatin structure. Next I should consider the *Mtrr*$^{+/+}$F1 generation. Are there functionally relevant changes in chromatin structure and HIRA activity in the F1 generation? It will be essential to perform ChIP-seq experiments (and/or immunohistochemistry) on *Mtrr*$^{+/+}$F1 generation zygotes and embryos at E10.5 to assess H3.3 incorporation. This should be equivalent

to C57Bl/6 controls. It must be determined if *Hira* is misexpressed in certain cell types of the $Mtrr^{+/+}$F1 generation embryo, particularly PGCs. RT-qPCR, immunohistochemistry and/or ChIP-seq should then be used to determine *Hira* mRNA expression, HIRA protein levels and H3.3 levels in F1 generation oocytes. If the hypothesis we present is correct, differences should be detected between F1 oocytes and C57Bl/6 control oocytes. The epigenome of F1 generation embryos and oocytes, including histone modification, DNA methylation and sncRNA profiles, should be examined to determine if there is indeed epigenetic instability. A similar analysis to that performed in the F1 generation should then be performed in F2 and F3 generation tissues to identify functionally relevant changes in chromatin structure and epigenetic instability. We would like to establish if oocyte specific disruption of HIRA is required for the phenotypes in the F2 generation. As such, the presence or absence of phenotypes in F2 generation embryos at E10.5 derived from $Mtrr^{+/+}$F1 sperm and C57Bl/6 oocytes should be assessed. This cross has previously not been performed and will demonstrate the effect of paternal grandpaternal *Mtrr* deficiency.

This hypothesis should also be tested and explored outside the $Mtrr^{+/gt}$ maternal grandfather pedigree. Firstly, the role of *Hira* could be investigated in the maternal grandmother pedigree (Chapter 1, Figure 1.6). In particular, HIRA and H3.3 levels in oocytes of $Mtrr^{+/gt}$maternal grandmothers (F0) should be assessed. Next, we need to establish what drives *Hira* misexpression in F2 and F3 generation embryos. To begin to explore the role of epigenetic instability in establishing the misexpression of *Hira*, *Hira* expression could be assessed in *Dnmt* knock-out mice (e.g $Dnmt3L^{-/-}$ mice or $Dnmt3a^{-/-}/Dnmt3b^{-/-}$ double knock-outs). Interestingly, $Dnmt3L^{-/-}$ mice also have similar embryonic phenotypes to $Hira^{-/-}$ mice (Hata et al., 2002). Additionally, miRNA-124 has been shown to modulate *Hira* expression in neurites (Yu et al., 2008). We did not observe miRNA-124 misexpression in $Mtrr^{gt}$sperm, but perhaps *Hira* may be regulated by different sncRNA species in sperm or the early embryo. The regulation of *Hira* by sncRNAs should be explored *in vitro*. Candidate regulatory sncRNAs identified bioinformationally could be depleted using RNAi-based approaches. It would be interesting to determine if *Hira* mRNA misexpression or abnormal

HIRA function is a common consequence of low-methyl group availability or a *Mtrr*-specific effect. *Hira* expression should be assessed in other abnormal folate metabolism models, in mice fed a folate-deficient diet or in embryonic stem cells grown in folate-deficient medium. Furthermore, whether transgenerational inheritance of developmental phenotypes occurs in the wildtype progeny and grandprogeny of $Hira^{+/-}$ mice has not been established and should be investigated. These experiments should help elucidate if HIRA does indeed play a role in the inheritance of embryonic abnormalities in the $Mtrr^{gt}$model.

### 8.2.5 rDNA mediated transgenerational inheritance in the $Mtrr^{gt}$ model

We recognise that genetic and epigenetic variation associated with rDNA could possibly contribute to the inheritance of phenotypes observed in $Mtrr^{gt}$mice. *Rn45s* is the pre-ribosomal RNA transcript encoded in the genome. Interestingly, we observed differential methylation at the *Rn45s* promoter in $Mtrr^{+/gt}$and $Mtrr^{gt/gt}$sperm and tissue specific downregulation of *Rn45s* in F1, F2 and F3 embryos (Chapter 6, Figure 6.5). Interestingly, a DMR associated with *Rn45s* was also identified in sperm of males exposed to *in utero* undernutrition (Radford et al., 2014). rDNA is associated with copy number variation and epigenetic variation between inbred individuals (Shea et al., 2015). There are some who believe, although with scant evidence, that all transgenerational epigenetic inheritance may result from unmapped induced polymorphisms in the ribosomal DNA (Bughio and Maggert, 2018). $Mtrr^{gt}$mice could be an ideal model to elucidate the link between rDNA and TEI further. However, studying the rDNA locus is challenging even with modern molecular genetic approaches (Bughio and Maggert, 2018).

Intriguingly, there are several studies that suggest a possible link between HIRA and ribosomal-RNA. HIRA has been linked to the repression of rRNA transcription in *S. cerevisiae* (Neumuller et al., 2013). Conversely, HIRA-dependent rRNA transcription has been shown to be vital for the first cleavage division in mouse zygotes (Lin et al., 2014). This raised a possibility that the decreased *Rn45s* expression observed in $Mtrr^{+/+}$F2 and F3 generation embryos was a downstream consequence of decreased *Hira* expression. However the

downregulation of *Rn45s* in F1 embryos in which *Hira* mRNA expression is equivalent to C57Bl/6 controls challenges this prediction. Although, if HIRA activity or function were abnormal despite normal mRNA expression in F1 embryos, this could perhaps lead to decreased *Rn45s* expression. Assessing *Rn45s* expression in $Hira^{-/-}$ embryos could help determine if *Hira* does play a role in the regulation of *Rn45s*. Probing this link between ribosomal-RNA and HIRA could help explicate the mechanisms of TEI in the $Mtrr^{gt}$model

## 8.3 Epigenetic inheritance via the maternal line

This study has focused on inheritance initiated by *Mtrr* heterozygosity in the maternal grandfather. Inheritance of developmental abnormalities also occurs in the $Mtrr^{+/gt}$maternal grandmother pedigree. Studying the maternal grandmother pedigree is more challenging. Embryo transfer and IVF experiments are required to demonstrate that the inheritance of congenital malformations is independent of the maternal uterine environment (Padmanabhan et al., 2013). As we have identified differential methylation in *Mtrr* sperm, it would be interesting to perform a similar analysis in oocytes. This would allow us to ascertain if the $Mtrr^{gt}$mutation leads to similar alterations in epigenetic patterns in both male and female germ cells. Although the methylation profiles of oocytes and sperm are quite distinct, the phenotypes and phenotypic frequencies seen in the wildtype grandprogeny of the $Mtrr^{gt}$transgenerational pedigrees are similar regardless of whether the maternal grandparent is male or female. This might suggest differential DNA methylation, would be similar in sperm and oocytes. Likewise, ncRNA and histone modifications should also be profiled in oocytes. While technically more challenging than studying the sperm epigenome, comparing the epigenetic profiles of sperm and oocytes in the $Mtrr^{gt}$model may highlight common molecular pathways that mediate inheritance of developmental phenotypes.

## 8.4 Variability in the $Mtrr^{gt}$model

A strikingly feature of the $Mtrr^{gt}$model is the large degree of phenotypic variability between individuals. For example, a $Mtrr^{+/+}$F2 generation litter at E10.5,

derived form a *Mtrr*$^{+/gt}$grandparent, may be composed of entirely phenotypically normal embryos, entirely of embryos with a range of developmental phenotypes or a mixture of both (Padmanabhan et al., 2013). This phenotypic variability has been attributed to epigenetic variability between individuals. *Mtrr* deficiency may lead to stochastic epigenetic changes across the genome. However, at the sperm DMRs we identified, methylation appeared to be highly correlated between individuals at all bar a few exceptional loci (Chapter 5, Figures 5.8 and 5.9). The approach used to identify DMRs is however most likely to find those regions commonly differentially methylated. Other regions of the genome may show more variability in DNA methylation.

Epigenetic variability may also not be apparent within a population of cells but may become clear at a single cell level. Indeed, the percentage methylation at a CpG site detected using bisulfite pyrosequencing can crudely indicate variability in methylation, either between the maternal and paternal alleles or between cells within a population. This is because DNA methylation can be considered binary (Prochenka et al., 2015). Each CpG site can be either methylated or unmethylated. For example, in a somatic tissue, 50% methylation at a CpG site could represent one methylated and one unmethylated allele in all cells (as seen at imprinted loci). However, 50% methylation at a CpG site may also represent that 50% of cells in a population had both alleles methylated while the other 50% of cells had both alleles unmethylated. Germ cells are haploid. Therefore the level of germ cell DNA methylation represents varying proportions of germ cells with either a methylated or unmethylated cytosine at each locus. Therefore, as we identified CpG sites in sperm with neither 0% nor 100% methylation, there was variability in the methylation status within the sperm population of an individual male. Each sperm will become one individual. Therefore, if germ cell DNA methylation is able to influence offspring phenotypes, variable methylation might be expected to lead to variable penetrance of phenotypes across littermates (Shea et al., 2015). Perhaps this may account for the variable occurrence of developmental phenotypes in the *Mtrr*$^{gt}$model. Variability in histone modifications and ncRNA profiles may also be present and should be assessed in the *Mtrr*$^{gt}$model. Single-cell epigenome analysis approaches could be used to investigate this. Ideally, the epigenomic

profile of a single sperm and the offspring derived from it would be determined. As this is not possible, correlating single cell epigenome information to phenotypes will remain challenging.

## 8.5 The *Mtrr*$^{gt}$model as a TEI paradigm

In this thesis we demonstrate that differential methylation identified in sperm is not maintained in offspring tissues and is likely reprogrammed in the blastocyst. This supports a growing body of evidence that suggests that DNA methylation is not directly inherited between generations (Radford et al., 2014; Iqbal et al., 2015; Blewitt et al., 2006; Kazachenka et al., 2018). As such, our data argues against so called replicative inheritance (Chapter 1, Figure 1.2). The reconstruction model of inheritance suggests that epigenetic marks responsible for a phenotype undergo normal reprogramming in the germline and zygote but are then reestablished in each successive generation (Jablonka, 2013). However, we do not observe a reestablishment of abnormal DNA methylation at loci differentially methylated in sperm in offspring tissues. This does not preclude that differential methylation may be established at other genomic loci in offspring tissues. Indeed epigenetic instability, in the form of global DNA hypomethylation and locus specific DNA dysregulation, has been observed in *Mtrr*$^{+/+}$F2 placentas at E10.5 derived from a *Mtrr*$^{+/gt}$grandparent (Padmanabhan et al., 2013). Locus specific whole-methylome analysis of F1 and F2 tissues should be performed. Therefore, our data does not entirely support the hypothesis of reconstructive inheritance. We suggest widening the scope of reconstructive inheritance to include general inheritance of epigenetic instability over several generations. This would best fit the *Mtrr*$^{gt}$model. The lack of overlap between epigenetic changes identified in F1 rats exposed to vinclozolin *in utero* and the unexposed F3 generation also supports the idea that there is a general inheritance of epigenetic instability in TEI models.

We tried to conduct this research in such as way as to minimise confounding factors and to address criticisms generally levied at TEI models (outlined in Chapter 1). Firstly, we directly address the potential for genetic inheritance in the *Mtrr*$^{gt}$model using whole genome sequencing. We believe this is the most thorough analysis of a potential genetic contribution that has been performed

in any TEI model. Previously, the potential for a genetic contribution in TEI models has only been considered in $A^{vy}$ mice, by sequencing a pair of littermates (Oey et al., 2015), and in F3 generation sperm of rats ancestrally exposed to vinclozolin using a genome-wide tiling array approach to identify CNVs (Skinner et al., 2015). Importantly, we were also able to use the whole genome sequencing data to determine if differential DNA methylation identified in sperm was likely secondary to genetic mutations. Secondly, we chose to use non-pooled samples for MeDIP-seq and sncRNA-seq analysis in order to allow us to identify inter-individual variability. While this could still be explored further, we show that generally inter-indervidual variability in DNA methylation is low at DMRs. Thirdly, sperm samples for MeDIP-seq and sncRNA-seq were collected from littermate and non-littermate males allowing littermate effects to be considered. We have not extensively explored potential littermate effects, however no overt bias as a result of littermate status was observed in the data. Fourthly, when considering inheritance of DNA methylation, we assessed the unexposed F2 and F3 generations thus considering a true transgenerational inheritance effect. Finally, we have begun an examination of interaction between epigenetic mechanisms by profiling DNA methylation and ncRNAs in the $Mtrr^{gt}$model. Histone modifications should be assessed to complete the full picture of the epigenetic landscape of sperm in the $Mtrr^{gt}$model.

However, we did not successfully exclude all confounders in this study. All F1 and F2 generations studied were conceived naturally. Male mate choice effects and seminal fluid composition may influence offspring outcomes (Drickamer et al., 2003; Watkins et al., 2018). This could be excluded by using IVF to generate all F1 and F2 generations. However, we are conscious that *in vitro* manipulation techniques may lead to alterations in the epigenetic landscape of the offspring and thus introduce a further confounding factor (Ventura-Junc et al., 2015; Canovas et al., 2017). Furthermore, we have yet to accurately correlate any epigenetic changes identified in the $Mtrr^{gt}$model to the developmental phenotypes observed. Performing experiments to assess if *Hira* misexpression leads to functionally relevant changes in chromatin structure may be the first step to addressing this question. Likewise, microinjection studies of misexpressed sncRNA should be performed. In the future targeted epimutagenesis

could be used to demonstrate a causal link between any epimutations identified and offspring phenotypes. Systems such as TALE-TET1-fusions (Maeder et al., 2013) and CRISPR-Cas9-based actelytransferase enzymes (Hilton et al., 2015) could be used to generate or remove an epimutation. Recapitulation or rescue of the phenotypes expected would suggest a causal role for that epigenetic mark.

## 8.6 Using the *Mtrr*$^{gt}$model to understand the role of folate in offspring health

The findings of our study add to the large body of conflicting evidence exploring the role of folate on male fertility, genetic stability, DNA methylation, and offspring health. We observed a mild testes phenotype and normal fertility in the *Mtrr*$^{gt}$ model. This is consistent with a model of dietary folate deficiency in C57Bl/6 mice (Lambrot et al., 2013), but a milder testes phenotype than observed in *Mthfr* knock-out mice (Kelly et al., 2005; Chan et al., 2010). We report that the *Mtrr*$^{gt}$mutation likely does not lead to genetic instability, although further genetic analysis is required. This finding adds weight to evidence that folate deficiency does not lead to increased mutation rates (Choi et al., 1998; Voutounou et al., 2012) but directly conflicts other reports (Blount et al., 1997; Swayne et al., 2012). Furthermore, our data supports existing evidence that folate is required for DNA methylation in sperm (Lambrot et al., 2013; Chan et al., 2010). Overall, there may be some commonality between the impact of the *Mtrr*$^{gt}$mutation, dietary folate-deficiency and other models of abnormal metabolism on testes morphology and function, genetic stability and DNA methylation. However, complexities and conflicts arise from exposure or mutation specific effects and technical variation. This makes drawing general conclusions about the role of folate impossible from a single study.

We identify several other novel epigenetic candidates that may play a role in phenotypic inheritance in *Mtrr*$^{gt}$mice that should be explored in other models of abnormal folate metabolism or dietary folate deficiency. sncRNAs are misexpressed in sperm of the *Mtrr*$^{gt}$model and we observed that the RNA methyltransferase *Dnmt2* is down-regulated in the epididymis. sncRNAs, po-

tentially with specific sncRNA modifications, can be taken up by sperm from the epididymis (Vojtech et al., 2014). Expression of ncRNAs, and their modification profiles, should be examined in folate-deficient mice and other models of abnormal folate metabolism. We propose that *Hira* misexpression may be a crucial factor driving phenotypes in the $Mtrr^{gt}$model. It will be important to identify if *Hira* is also dyresgulated in other models of dietary folate deficiency or abnormal folate metabolism. Furthermore, it will be interesting to examine if transgenerational inheritance occurs following dietary folate deficiency or in the wildtype offspring generations of models in which folate metabolism is disrupted by genetic mutations.

## 8.7 Implications of TEI following abnormal folate metabolism

Folate deficiency if a relatively common problem among women of child bearing age, particularly in countries without mandatory fortification or widespread use of supplementation. For example, in Belize almost 50% of women of child bearing age had folate insufficiency (Rosenthal et al., 2017). Furthermore, polymorphisms in folate metabolism enzymes are prevalent in some human populations, although their frequency varies with ethnicity and geographical location (Hiraoka and Kagawa, 2017). It will be vital to perform epidemiological studies on folate-deficient populations and those with polymorphisms in genes such as *Mthfr* and *Mtrr* to determine if transgenerational inheritance of developmental abnormalities occurs in human populations. It is also coming to a time when we can begin to assess the impact of folic acid fortification programmes established in many countries in the late 1990s. The grandchildren of those women first consuming folic acid fortified foods, who were exposed to fortification *in utero* as germ cells, are beginning to be born. It will be interesting to observe if there is a greater reduction in the occurrence of birth defects than would be expected with direct fortification/supplementation of the F1 generation alone owing to a grand-parental effect. Further studies should also be conducted exploring the role of male folate status on offspring health in humans. In mice, there is evidence that abnormal folate metabolism in the

maternal grandfather or paternal folate deficiency can impact offspring health (Padmanabhan et al., 2013; Lambrot et al., 2013). Perhaps studies could evaluate if folate supplementation should be recommended to males whose partners are trying to conceive. With the recent announcement that the UK government will begin consultations on the fortification of flour with folic acid (October 2018), now may be a prime time to begin these investigations. The impact of a fortification programme in the UK, if introduced, should be monitored closely over several generations. Developing a complete picture of how folate impacts offspring health over multiple generations may impact how we act to mitigate the inheritance of disease in the future.

# Bibliography

A, Z. C., Yang, Y., Zhang, S. Z., Li, N., and Zhang, W. Single nucleotide polymorphism C677T in the methylenetetrahydrofolate reductase gene might be a genetic risk factor for infertility for Chinese men with azoospermia or severe oligozoospermia. *Asian J Androl*, 2007. 9(1): 57–62.

Agarwal, A., Virk, G., Ong, C., and du Plessis, S. S. Effect of oxidative stress on male reproduction. *The world journal of men's health*, 2014. 32(1): 1–17.

Aiken, C. E. and Ozanne, S. E. Transgenerational developmental programming. *Hum Reprod Update*, 2014. 20(1): 63–75.

Aiken, C. E., Tarry-Adkins, J. L., and Ozanne, S. E. Transgenerational effects of maternal diet on metabolic and reproductive ageing. *Mamm Genome*, 2016.

Ajay, S. S., Parker, S. C., Abaan, H. O., Fajardo, K. V., and Margulies, E. H. Accurate and comprehensive sequencing of personal genomes. *Genome Res*, 2011. 21(9): 1498–505.

Almouzni, G. and Cedar, H. Maintenance of Epigenetic Information. *Cold Spring Harb Perspect Biol*, 2016. 8(5).

Amouroux, R., Nashun, B., Shirane, K., Nakagawa, S., Hill, P. W. S., DSouza, Z., Nakayama, M., Matsuda, M., Turp, A., Ndjetehe, E., Encheva, V., Kudo, N. R., Koseki, H., Sasaki, H., and Hajkova, P. De novo DNA methylation drives 5hmC accumulation in mouse zygotes. *Nature Cell Biology*, 2016. 18: 225.

Anawalt, B. D. Approach to Male Infertility and Induction of Spermatogenesis. *The Journal of Clinical Endocrinology and Metabolism*, 2013. 98(9): 3532–3542.

Anway, M. D., Cupp, A. S., Uzumcu, M., and Skinner, M. K. Epigenetic transgenerational actions of endocrine disruptors and male fertility. *Science*, 2005. 308(5727): 1466–9.

Anway, M. D., Leathers, C., and Skinner, M. K. Endocrine disruptor vinclozolin induced epigenetic transgenerational adult-onset disease. *Endocrinology*, 2006. 147(12): 5515–23.

Ash, J. A., Jiang, X., Malysheva, O. V., Fiorenza, C. G., Bisogni, A. J., Levitsky, D. A., Strawderman, M. S., Caudill, M. A., Stover, P. J., and Strupp, B. J. Dietary and genetic manipulations of folate metabolism differentially affect neocortical functions in mice. *Neurotoxicology and teratology*, 2013. 38: 79–91.

Avendao, C. and Menndez, J. C. *Chapter 2 - Antimetabolites*, pages 9–52. Elsevier, Amsterdam, 2008.

Barau, J., Teissandier, A., Zamudio, N., Roy, S., Nalesso, V., Hrault, Y., Guillou, F., and Bourchis, D. The DNA methyltransferase DNMT3C protects male germ cells from transposon activity. *Science*, 2016. 354: 909–912.

Barnett, M. P. G., Bermingham, E. N., Young, W., Bassett, S. A., Hesketh, J. E., Maciel-Dominguez, A., McNabb, W. C., and Roy, N. C. Low folate and selenium in the mouse maternal diet alters liver gene expression patterns in the offspring after weaning. *Nutrients*, 2015. 7(5): 3370–3386.

Beck, D., Sadler-Riggleman, I., and Skinner, M. K. Generational comparisons (F1 versus F3) of vinclozolin induced epigenetic transgenerational inheritance of sperm differential DNA methylation regions (epimutations) using MeDIP-Seq. *Environ Epigenet*, 2017. 3(3).

Ben Maamar, M., Sadler-Riggleman, I., Beck, D., McBirney, M., Nilsson, E., Klukovich, R., Xie, Y., Tang, C., Yan, W., and Skinner, M. K. Alterations in sperm DNA methylation, non-coding RNA expression, and histone retention mediate vinclozolin-induced epigenetic transgenerational inheritance of disease. *Environmental Epigenetics*, 2018. 4(2): dvy010–dvy010.

Benito, E., Kerimoglu, C., Ramachandran, B., Pena-Centeno, T., Jain, G., Stilling, R. M., Islam, M. R., Capece, V., Zhou, Q., Edbauer, D., Dean, C., and Fis-

cher, A. RNA-Dependent Intergenerational Inheritance of Enhanced Synaptic Plasticity after Environmental Enrichment. *Cell Rep*, 2018. 23(2): 546–554.

Bezold, G., Lange, M., and Peter, R. U. Homozygous methylenetetrahydrofolate reductase C677T mutation and male infertility. *N Engl J Med*, 2001. 344(15): 1172–3.

Bird, A. DNA methylation patterns and epigenetic memory. *Genes Dev*, 2002. 16(1): 6–21.

Bistulfi, G., Vandette, E., Matsui, S., and Smiraglia, D. J. Mild folate deficiency induces genetic and epigenetic instability and phenotype changes in prostate cancer cells. *BMC Biol*, 2010. 8: 6.

Blake, G. E. T., Rakoczy, J., and Watson, E. D. *Chapter 26 - Epigenetics of Transgenerational Inheritance of Disease*, volume 6, pages 805–836. Academic Press, 2018.

Blake, G. E. T. and Watson, E. D. Unravelling the complex mechanisms of transgenerational epigenetic inheritance. *Current Opinion in Chemical Biology*, 2016. 33: 101–107.

Blewitt, M. E., Vickaryous, N. K., Paldi, A., Koseki, H., and Whitelaw, E. Dynamic Reprogramming of DNA Methylation at an Epigenetically Sensitive Allele in Mice. *PLOS Genetics*, 2006. 2(4): e49.

Blount, B. C., Mack, M. M., Wehr, C. M., MacGregor, J. T., Hiatt, R. A., Wang, G., Wickramasinghe, S. N., Everson, R. B., and Ames, B. N. Folate deficiency causes uracil misincorporation into human DNA and chromosome breakage: implications for cancer and neuronal damage. *Proc Natl Acad Sci U S A*, 1997. 94(7): 3290–5.

Bohacek, J. and Mansuy, I. M. A guide to designing germline-dependent epigenetic inheritance experiments in mammals. *Nat Methods*, 2017. 14(3): 243–249.

Borgel, J., Guibert, S., Li, Y., Chiba, H., Schubeler, D., Sasaki, H., Forne, T., and Weber, M. Targets and dynamics of promoter DNA methylation during early mouse development. *Nat Genet*, 2010. 42(12): 1093–100.

Bouckenheimer, J., Fauque, P., Lecellier, C.-H., Bruno, C., Commes, T., Lematre, J.-M., De Vos, J., and Assou, S. Differential long non-coding RNA expression profiles in human oocytes and cumulus cells. *Scientific Reports*, 2018. 8(1): 2202.

Bourc'his, D. and Bestor, T. H. Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature*, 2004. 431(7004): 96–9.

Boxmeer, J. C., Smit, M., Utomo, E., Romijn, J. C., Eijkemans, M. J. C., Lindemans, J., Laven, J. S. E., Macklon, N. S., Steegers, E. A. P., and Steegers-Theunissen, R. P. M. Low folate in seminal plasma is associated with increased sperm DNA damage. *Fertility and Sterility*, 2009. 92(2): 548–556.

Brar, G. A. and Weissman, J. S. Ribosome profiling reveals the what, when, where and how of protein synthesis. *Nat Rev Mol Cell Biol*, 2015. 16(11): 651–64.

Brunner, A. M., Nanni, P., and Mansuy, I. M. Epigenetic marking of sperm by post-translational modification of histones and protamines. *Epigenetics Chromatin*, 2014. 7(1): 2.

Bughio, F. and Maggert, K. A. The peculiar genetics of the ribosomal DNA blurs the boundaries of transgenerational epigenetic inheritance. *Chromosome Research*, 2018.

Bull, C. F., Mayrhofer, G., O'Callaghan, N. J., Au, A. Y., Pickett, H. A., Low, G. K., Zeegers, D., Hande, M. P., and Fenech, M. F. Folate deficiency induces dysfunctional long and short telomeres; both states are associated with hypomethylation and DNA damage in human WIL2-NS cells. *Cancer Prev Res (Phila)*, 2014. 7(1): 128–38.

Bull, C. F., Mayrhofer, G., Zeegers, D., Mun, G. L., Hande, M. P., and Fenech, M. F. Folate deficiency is associated with the formation of complex nuclear anomalies in the cytokinesis-block micronucleus cytome assay. *Environ Mol Mutagen*, 2012. 53(4): 311–23.

Bygren, L. O., Tinghog, P., Carstensen, J., Edvinsson, S., Kaati, G., Pembrey, M. E., and Sjostrom, M. Change in paternal grandmothers' early food supply

influenced cardiovascular mortality of the female grandchildren. *BMC Genet*, 2014. 15: 12.

Cabelof, D. C., Raffoul, J. J., Nakamura, J., Kapoor, D., Abdalla, H., and Heydari, A. R. Imbalanced base excision repair in response to folate deficiency is accelerated by polymerase beta haploinsufficiency. *J Biol Chem*, 2004. 279(35): 36504–13.

Calado, R. T. and Dumitriu, B. Telomere dynamics in mice and humans. *Seminars in hematology*, 2013. 50(2): 165–174.

Callicott, R. J. and Womack, J. E. Real-time PCR assay for measurement of mouse telomeres. *Comp Med*, 2006. 56(1): 17–22.

Canovas, S., Ross, P. J., Kelsey, G., and Coy, P. DNA Methylation in Embryo Development: Epigenetic Impact of ART (Assisted Reproductive Technologies). *Bioessays*, 2017. 39(11).

Carone, B. R., Fauquier, L., Habib, N., Shea, J. M., Hart, C. E., Li, R., Bock, C., Li, C., Gu, H., Zamore, P. D., Meissner, A., Weng, Z., Hofmann, H. A., Friedman, N., and Rando, O. J. Paternally induced transgenerational environmental reprogramming of metabolic gene expression in mammals. *Cell*, 2010. 143(7): 1084–96.

Carone, B. R., Hung, J. H., Hainer, S. J., Chou, M. T., Carone, D. M., Weng, Z., Fazzio, T. G., and Rando, O. J. High-resolution mapping of chromatin packaging in mouse embryonic stem cells and sperm. *Dev Cell*, 2014. 30(1): 11–22.

Carrell, D. T. and Aston, K. I. *Spermatogenesis*, volume 927 of *Methods in Molecular Biology*. Humana Press, New York, 2013.

Casas, E. and Vavouri, T. Sperm epigenomics: challenges and opportunities. *Front Genet*, 2014. 5: 330.

Cawthon, R. M. Telomere measurement by quantitative PCR. *Nucleic Acids Research*, 2002. 30(10): e47–e47.

Champagne, F. A. and Meaney, M. J. Stress during gestation alters postpartum maternal care and the development of the offspring in a rodent model. *Biol Psychiatry*, 2006. 59(12): 1227–35.

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

Chan, P. P. and Lowe, T. M. GtRNAdb: a database of transfer RNA genes detected in genomic sequence. *Nucleic Acids Research*, 2009. 37(Database Issue): D93–D97.

Chan, P. P. and Lowe, T. M. GtRNAdb 2.0: an expanded database of transfer RNA genes identified in complete and draft genomes. *Nucleic Acids Research*, 2016. 44(Database issue): D184–D189.

Chanson, A., Sayd, T., Rock, E., Chambon, C., Sante-Lhoutellier, V., Potier de Courcy, G., and Brachet, P. Proteomic analysis reveals changes in the liver protein pattern of rats exposed to dietary folate deficiency. *J Nutr*, 2005. 135(11): 2524–9.

Chen, Q., Yan, M., Cao, Z., Li, X., Zhang, Y., Shi, J., Feng, G. H., Peng, H., Zhang, X., Zhang, Y., Qian, J., Duan, E., Zhai, Q., and Zhou, Q. Sperm tsRNAs contribute to intergenerational inheritance of an acquired metabolic disorder. *Science*, 2015.

Chen, Q., Yan, W., and Duan, E. Epigenetic inheritance of acquired traits through sperm RNAs and sperm RNA modifications. *Nat Rev Genet*, 2016a. 17(12): 733–743.

Chen, X., Schulz-Trieglaff, O., Shaw, R., Barnes, B., Schlesinger, F., Kllberg, M., Cox, A. J., Kruglyak, S., and Saunders, C. T. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics*, 2016b. 32(8): 1220–1222.

Choi, C. S., Gonzales, E. L., Kim, K. C., Yang, S. M., Kim, J. W., Mabunga, D. F., Cheong, J. H., Han, S. H., Bahn, G. H., and Shin, C. Y. The transgenerational

inheritance of autism-like phenotypes in mice exposed to valproic acid during pregnancy. *Sci Rep*, 2016. 6: 36250.

Choi, S. W., Kim, Y. I., Weitzel, J. N., and Mason, J. B. Folate depletion impairs DNA excision repair in the colon of the rat. *Gut*, 1998. 43(1): 93–9.

Christensen, K. E., Zada, Y. F., Rohlicek, C. V., Andelfinger, G. U., Michaud, J. L., Bigras, J. L., Richter, A., Dube, M. P., and Rozen, R. Risk of congenital heart defects is influenced by genetic variation in folate metabolism. *Cardiol Young*, 2013. 23(1): 89–98.

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Lu, X., and Ruden, D. M. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly*, 2012. 6(2): 80–92.

Cossetti, C., Lugini, L., Astrologo, L., Saggio, I., Fais, S., and Spadafora, C. Soma-to-germline transmission of RNA in mice xenografted with human tumour cells: possible transport by exosomes. *PLoS One*, 2014. 9(7): e101629.

Crichton, J. H., Dunican, D. S., Maclennan, M., Meehan, R. R., and Adams, I. R. Defending the genome from the enemy within: mechanisms of retrotransposon suppression in the mouse germline. *Cell Mol Life Sci*, 2014. 71(9): 1581–605.

Crider, K. S., Bailey, L. B., and Berry, R. J. Folic acid food fortification-its history, effect, concerns, and future directions. *Nutrients*, 2011. 3(3): 370–384.

Cropley, J. E., Eaton, S. A., Aiken, A., Young, P. E., Giannoulatou, E., Ho, J. W., Buckland, M. E., Keam, S. P., Hutvagner, G., Humphreys, D. T., Langley, K. G., Henstridge, D. C., Martin, D. I., Febbraio, M. A., and Suter, C. M. Male-lineage transmission of an acquired metabolic phenotype induced by grand-paternal obesity. *Mol Metab*, 2016. 5(8): 699–708.

Cropley, J. E., Suter, C. M., Beckman, K. B., and Martin, D. I. CpG methylation of a silent controlling element in the murine Avy allele is incomplete and unresponsive to methyl donor supplementation. *PLoS One*, 2010. 5(2): e9055.

Czeizel, A. E. and Duds, I. Prevention of the First Occurrence of Neural-Tube Defects by Periconceptional Vitamin Supplementation. *New England Journal of Medicine*, 1992. 327(26): 1832–1835.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R., and Genomes Project Analysis, G. The variant call format and VCFtools. *Bioinformatics*, 2011. 27(15): 2156–2158.

Daskalos, A., Nikolaidis, G., Xinarianos, G., Savvari, P., Cassidy, A., Zakopoulou, R., Kotsinas, A., Gorgoulis, V., Field, J. K., and Liloglou, T. Hypomethylation of retrotransposable elements correlates with genomic instability in non-small cell lung cancer. *Int J Cancer*, 2009. 124(1): 81–7.

Davis, T. L., Yang, G. J., McCarrey, J. R., and Bartolomei, M. S. The H19 methylation imprint is erased and re-established differentially on the parental alleles during male germ cell development. *Hum Mol Genet*, 2000. 9(19): 2885–94.

Daxinger, L., Oey, H., Isbel, L., Whitelaw, N. C., Youngson, N. A., Spurling, A., Vonk, K. K., and Whitelaw, E. Hypomethylation of ERVs in the sperm of mice haploinsufficient for the histone methyltransferase Setdb1 correlates with a paternal effect on phenotype. *Sci Rep*, 2016. 6: 25004.

Daxinger, L. and Whitelaw, E. Understanding transgenerational epigenetic inheritance via the gametes in mammals. *Nat Rev Genet*, 2012. 13(3): 153–62.

de Castro Barbosa, T., Ingerslev, L. R., Alm, P. S., Versteyhe, S., Massart, J., Rasmussen, M., Donkin, I., Sjogren, R., Mudry, J. M., Vetterli, L., Gupta, S., Krook, A., Zierath, J. R., and Barres, R. High-fat diet reprograms the epigenome of rat spermatozoa and transgenerationally affects metabolism of the offspring. *Mol Metab*, 2016. 5(3): 184–97.

Deaton, A. M. and Bird, A. CpG islands and the regulation of transcription. *Genes Dev*, 2011. 25(10): 1010–22.

Del Sal, G., Ruaro, M. E., Philipson, L., and Schneider, C. The growth arrest-specific gene, gas1, is involved in growth suppression. *Cell*, 1992. 70(4): 595–607.

Deng, L., Elmore, C. L., Lawrance, A. K., Matthews, R. G., and Rozen, R. Methionine synthase reductase deficiency results in adverse reproductive outcomes and congenital heart defects in mice. *Mol Genet Metab*, 2008. 94(3): 336–42.

Dias, B. G. and Ressler, K. J. Parental olfactory experience influences behavior and neural structure in subsequent generations. *Nat Neurosci*, 2014. 17(1): 89–96.

Dickies, M. M. A new viable yellow mutation in the house mouse. *J Hered*, 1962. 53: 84–6.

Ding, Z., Mangino, M., Aviv, A., Consortium, U. K., Spector, T., and Durbin, R. Estimating telomere length from whole genome sequence data. *Nucleic Acids Research*, 2014. 42(9): e75–e75.

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T. R. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, 2013. 29(1): 15–21.

Dobosy, J. R., Fu, V. X., Desotelle, J. A., Srinivasan, R., Kenowski, M. L., Almassi, N., Weindruch, R., Svaren, J., and Jarrard, D. F. A methyl-deficient diet modifies histone methylation and alters Igf2 and H19 repression in the prostate. *Prostate*, 2008. 68(11): 1187–95.

Doetschman, T. Influence of genetic background on genetically engineered mouse phenotypes. *Methods in molecular biology (Clifton, N.J.)*, 2009. 530: 423–433.

Dolinoy, D. C., Huang, D., and Jirtle, R. L. Maternal nutrient supplementation counteracts bisphenol A-induced DNA hypomethylation in early development. *Proc Natl Acad Sci U S A*, 2007. 104(32): 13056–61.

Drenckhahn, J. D., Schwarz, Q. P., Gray, S., Laskowski, A., Kiriazis, H., Ming, Z., Harvey, R. P., Du, X. J., Thorburn, D. R., and Cox, T. C. Compensatory growth of healthy cardiac cells in the presence of diseased cells restores tissue homeostasis during heart development. *Dev Cell*, 2008. 15(4): 521–33.

Drickamer, L. C., Gowaty, P. A., and Wagner, D. M. Free mutual mate preferences in house mice affect reproductive success and offspring performance. *Animal Behaviour*, 2003. 65(1): 105–114.

Du, J., Johnson, L. M., Jacobsen, S. E., and Patel, D. J. DNA methylation pathways and their crosstalk with histone methylation. *Nature reviews. Molecular cell biology*, 2015. 16(9): 519–532.

Duthie, S. J. Folate and cancer: how DNA damage, repair and methylation impact on colon carcinogenesis. *J Inherit Metab Dis*, 2011. 34(1): 101–9.

Duthie, S. J., Grant, G., and Narayanan, S. Increased uracil misincorporation in lymphocytes from folate-deficient rats. *Br J Cancer*, 2000a. 83(11): 1532–7.

Duthie, S. J. and Hawdon, A. DNA instability (strand breakage, uracil misincorporation, and defective repair) is increased by folic acid depletion in human lymphocytes in vitro. *FASEB J*, 1998. 12(14): 1491–7.

Duthie, S. J., Narayanan, S., Brand, G. M., and Grant, G. DNA stability and genomic methylation status in colonocytes isolated from methyl-donor-deficient rats. *Eur J Nutr*, 2000b. 39(3): 106–11.

Duthie, S. J., Narayanan, S., Brand, G. M., Pirie, L., and Grant, G. Impact of folate deficiency on DNA stability. *J Nutr*, 2002. 132(8 Suppl): 2444S–2449S.

Ebisch, I. M., van Heerde, W. L., Thomas, C. M., van der Put, N., Wong, W. Y., and Steegers-Theunissen, R. P. C677T methylenetetrahydrofolate reductase polymorphism interferes with the effects of folic acid and zinc sulfate on sperm concentration. *Fertil Steril*, 2003. 80(5): 1190–4.

Eckart, S., Hortnagl, H., Kronenberg, G., Gertz, K., Horster, H., Endres, M., and Hellweg, R. Reduced nerve growth factor levels in stress-related brain regions of folate-deficient mice. *Neuroscience*, 2013. 245: 129–35.

Eckhardt, F., Lewin, J., Cortese, R., Rakyan, V. K., Attwood, J., Burger, M., Burton, J., Cox, T. V., Davies, R., Down, T. A., Haefliger, C., Horton, R., Howe, K., Jackson, D. K., Kunde, J., Koenig, C., Liddle, J., Niblett, D., Otto, T., Pettett, R., Seemann, S., Thompson, C., West, T., Rogers, J., Olek, A., Berlin, K., and Beck, S. DNA methylation profiling of human chromosomes 6, 20 and 22. *Nat Genet*, 2006. 38(12): 1378–85.

Ek, J. Plasma and red cell folate values in newborn infants and their mothers in relation to gestational age. *The Journal of Pediatrics*, 1980. 97(2): 288–292.

Elmore, C. L., Wu, X., Leclerc, D., Watson, E. D., Bottiglieri, T., Krupenko, N. I., Krupenko, S. A., Cross, J. C., Rozen, R., Gravel, R. A., and Matthews, R. G. Metabolic derangement of methionine and folate metabolism in mice deficient in methionine synthase reductase. *Mol Genet Metab*, 2007. 91(1): 85–97.

Eric Tang, M.-H., Varadan, V., Kamalakaran, S., Zhang, M. Q., Dimitrova, N., and Hicks, J. Major Chromosomal Breakpoint Intervals in Breast Cancer Co-Localize with Differentially Methylated Regions. *Frontiers in Oncology*, 2012. 2: 197.

Erkek, S., Hisano, M., Liang, C. Y., Gill, M., Murr, R., Dieker, J., Schubeler, D., van der Vlag, J., Stadler, M. B., and Peters, A. H. Molecular determinants of nucleosome retention at CpG-rich sequences in mouse spermatozoa. *Nat Struct Mol Biol*, 2013. 20(7): 868–75.

Fang, H., Wu, Y., Narzisi, G., O'Rawe, J. A., Barron, L. T., Rosenbaum, J., Ronemus, M., Iossifov, I., Schatz, M. C., and Lyon, G. J. Reducing INDEL calling errors in whole genome and exome sequencing data. *Genome Med*, 2014. 6(10): 89.

Farrer, R. A., Henk, D. A., MacLean, D., Studholme, D. J., and Fisher, M. C. Using False Discovery Rates to Benchmark SNP-callers in next-generation sequencing projects. *Scientific Reports*, 2013. 3: 1512.

Farris, W., Mansourian, S., Chang, Y., Lindsley, L., Eckman, E. A., Frosch, M. P., Eckman, C. B., Tanzi, R. E., Selkoe, D. J., and Guenette, S. Insulin-degrading

enzyme regulates the levels of insulin, amyloid beta-protein, and the beta-amyloid precursor protein intracellular domain in vivo. *Proc Natl Acad Sci U S A*, 2003. 100(7): 4162–7.

Faulk, C., Kim, J. H., Jones, T. R., McEachin, R. C., Nahar, M. S., Dolinoy, D. C., and Sartor, M. A. Bisphenol A-associated alterations in genome-wide DNA methylation and gene expression patterns reveal sequence-dependent and non-monotonic effects in human fetal liver. *Environmental Epigenetics*, 2015. 1(1): dvv006–dvv006.

Ficz, G., Branco, M. R., Seisenberger, S., Santos, F., Krueger, F., Hore, T. A., Marques, C. J., Andrews, S., and Reik, W. Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature*, 2011. 473(7347): 398–402.

Fontaine, D. A. and Davis, D. B. Attention to Background Strain Is Essential for Metabolic Research: C57BL/6 and the International Knockout Mouse Consortium. *Diabetes*, 2016. 65(1): 25–33.

Forges, T., Monnier-Barbarino, P., Alberto, J. M., Gueant-Rodriguez, R. M., Daval, J. L., and Gueant, J. L. Impact of folate and homocysteine metabolism on human reproductive health. *Hum Reprod Update*, 2007. 13(3): 225–38.

Fouse, S. D., Nagarajan, R. O., and Costello, J. F. Genome-scale DNA methylation analysis. *Epigenomics*, 2010. 2(1): 105–17.

Freeman, J. L., Perry, G. H., Feuk, L., Redon, R., McCarroll, S. A., Altshuler, D. M., Aburatani, H., Jones, K. W., Tyler-Smith, C., Hurles, M. E., Carter, N. P., Scherer, S. W., and Lee, C. Copy number variation: new insights in genome diversity. *Genome Res*, 2006. 16(8): 949–61.

Frye, M., Jaffrey, S. R., Pan, T., Rechavi, G., and Suzuki, T. RNA modifications: what have we learned and where are we headed? *Nature Reviews Genetics*, 2016. 17: 365.

Fullston, T., Ohlsson Teague, E. M., Palmer, N. O., DeBlasio, M. J., Mitchell, M., Corbett, M., Print, C. G., Owens, J. A., and Lane, M. Paternal obesity

initiates metabolic disturbances in two generations of mice with incomplete penetrance to the F2 generation and alters the transcriptional profile of testis and sperm microRNA content. *Faseb j*, 2013. 27(10): 4226–43.

Gahurova, L., Tomizawa, S. I., Smallwood, S. A., Stewart-Morgan, K. R., Saadeh, H., Kim, J., Andrews, S. R., Chen, T., and Kelsey, G. Transcription and chromatin determinants of de novo DNA methylation timing in oocytes. *Epigenetics Chromatin*, 2017. 10: 25.

Gapp, K., Bohacek, J., Grossmann, J., Brunner, A. M., Manuella, F., Nanni, P., and Mansuy, I. M. Potential of Environmental Enrichment to Prevent Transgenerational Effects of Paternal Trauma. *Neuropsychopharmacology*, 2016.

Gapp, K., van Steenwyk, G., Germain, P. L., Matsushima, W., Rudolph, K. L. M., Manuella, F., Roszkowski, M., Vernaz, G., Ghosh, T., Pelczar, P., Mansuy, I. M., and Miska, E. A. Alterations in sperm long RNA contribute to the epigenetic inheritance of the effects of postnatal trauma. *Molecular Psychiatry*, 2018.

Gapp, K., von Ziegler, L., Tweedie-Cullen, R. Y., and Mansuy, I. M. Early life epigenetic programming and transmission of stress-induced traits in mammals: how and when can environmental factors influence traits and their transgenerational inheritance? *Bioessays*, 2014. 36(5): 491–502.

Garner, J. L., Niles, K. M., McGraw, S., Yeh, J. R., Cushnie, D. W., Hermo, L., Nagano, M. C., and Trasler, J. M. Stability of DNA Methylation Patterns in Mouse Spermatogonia Under Conditions of MTHFR Deficiency and Methionine Supplementation. *Biology of Reproduction*, 2013. 89(5): 125.

Gebert, C., Kunkel, D., Grinberg, A., and Pfeifer, K. H19 imprinting control region methylation requires an imprinted environment only in the male germ line. *Mol Cell Biol*, 2010. 30(5): 1108–15.

Ghandour, H., Lin, B.-F., Choi, S.-W., Mason, J. B., and Selhub, J. Folate Status and Age Affect the Accumulation of l-Isoaspartyl Residues in Rat Liver Proteins. *The Journal of Nutrition*, 2002. 132(6): 1357–1360.

Gillich, A., Bao, S., Grabole, N., Hayashi, K., Trotter, M. W., Pasque, V., Magnusdottir, E., and Surani, M. A. Epiblast stem cell-based system reveals reprogramming synergy of germline factors. *Cell Stem Cell*, 2012. 10(4): 425–39.

Glasson, S. S., Askew, R., Sheppard, B., Carito, B., Blanchet, T., Ma, H. L., Flannery, C. R., Peluso, D., Kanki, K., Yang, Z., Majumdar, M. K., and Morris, E. A. Deletion of active ADAMTS5 prevents cartilage degradation in a murine model of osteoarthritis. *Nature*, 2005. 434(7033): 644–8.

Golshan Iranpour, F. and Rezazadeh Valojerdi, M. The epididymal sperm viability, motility and DNA integrity in dead mice maintained at 4-6oC. *Iranian Journal of Reproductive Medicine*, 2013. 11(3): 195–200.

Grandjean, V., Fourre, S., De Abreu, D. A., Derieppe, M. A., Remy, J. J., and Rassoulzadegan, M. RNA-mediated paternal heredity of diet-induced obesity and metabolic disorders. *Sci Rep*, 2015. 5: 18193.

Grandjean, V., Gounon, P., Wagner, N., Martin, L., Wagner, K. D., Bernex, F., Cuzin, F., and Rassoulzadegan, M. The miR-124-Sox9 paramutation: RNA-mediated epigenetic control of embryonic and adult growth. *Development*, 2009. 136(21): 3647–55.

Greer, E. L., Becker, B., Latza, C., Antebi, A., and Shi, Y. Mutation of C. elegans demethylase spr-5 extends transgenerational longevity. *Cell Res*, 2016. 26(2): 229–38.

Greer, E. L., Maures, T. J., Ucar, D., Hauswirth, A. G., Mancini, E., Lim, J. P., Benayoun, B. A., Shi, Y., and Brunet, A. Transgenerational epigenetic inheritance of longevity in Caenorhabditis elegans. *Nature*, 2011. 479(7373): 365–71.

Gregory, D. J., Kobzik, L., Yang, Z., McGuire, C. C., and Fedulov, A. V. Transgenerational transmission of asthma risk after exposure to environmental particles during pregnancy. *Am J Physiol Lung Cell Mol Physiol*, 2017. 313(2): L395–l405.

Griswold, M. D. Spermatogenesis: The Commitment to Meiosis. *Physiol Rev*, 2016. 96(1): 1–17.

Group, M. V. S. R. Prevention of neural tube defects: Results of the Medical Research Council Vitamin Study. *The Lancet*, 1991. 338(8760): 131–137.

Grozdanov, P. N. and Karagyozov, L. The mouse ribosomal DNA amplification promoting sequence 1 is highly methylated and repeated in the genome. *Zeitschrift fur Naturforschung - Section C Journal of Biosciences*, 2002. 57(9-10): 897–901.

Gu, T. P., Guo, F., Yang, H., Wu, H. P., Xu, G. F., Liu, W., Xie, Z. G., Shi, L., He, X., Jin, S. G., Iqbal, K., Shi, Y. G., Deng, Z., Szabo, P. E., Pfeifer, G. P., Li, J., and Xu, G. L. The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. *Nature*, 2011. 477(7366): 606–10.

Guerrero-Bosagna, C., Covert, T. R., Haque, M. M., Settles, M., Nilsson, E. E., Anway, M. D., and Skinner, M. K. Epigenetic transgenerational inheritance of vinclozolin induced mouse adult onset disease and associated sperm epigenome biomarkers. *Reprod Toxicol*, 2012. 34(4): 694–707.

Guo, F., Li, X., Liang, D., Li, T., Zhu, P., Guo, H., Wu, X., Wen, L., Gu, T.-P., Hu, B., Walsh, C., Li, J., Tang, F., and Xu, G.-L. Active and Passive Demethylation of Male and Female Pronuclear DNA in the Mammalian Zygote. *Cell Stem Cell*, 2014. 15(4): 447–459.

Hackett, J. A., Sengupta, R., Zylicz, J. J., Murakami, K., Lee, C., Down, T. A., and Surani, M. A. Germline DNA demethylation dynamics and imprint erasure through 5-hydroxymethylcytosine. *Science*, 2013. 339(6118): 448–52.

Hajkova, P., Jeffries, S. J., Lee, C., Miller, N., Jackson, S. P., and Surani, M. A. Genome-wide reprogramming in the mouse germ line entails the base excision repair pathway. *Science*, 2010. 329(5987): 78–82.

Hansson, V., Weddington, S. C., Naess, O., Attramadal, A., French, F. S., Kotite, N., and Nayfeh, S. N. Testicular androgen binding protein (ABP) - a parameter of Sertoli cell secretory function. *Curr Top Mol Endocrinol*, 1975. 2: 323–36.

Haque, M. M., Nilsson, E. E., Holder, L. B., and Skinner, M. K. Genomic Clustering of differential DNA methylated regions (epimutations) associated with the epigenetic transgenerational inheritance of disease and phenotypic variation. *BMC Genomics*, 2016. 17: 418.

Hata, K., Okano, M., Lei, H., and Li, E. Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice. *Development*, 2002. 129(8): 1983–93.

Heard, E. and Martienssen, R. A. Transgenerational epigenetic inheritance: myths and mechanisms. *Cell*, 2014. 157(1): 95–109.

Heid, M. K., Bills, N. D., Hinrichs, S. H., and Clifford, A. J. Folate deficiency alone does not produce neural tube defects in mice. *J Nutr*, 1992. 122(4): 888–94.

Heijmans, B. T., Tobi, E. W., Stein, A. D., Putter, H., Blauw, G. J., Susser, E. S., Slagboom, P. E., and Lumey, L. H. Persistent epigenetic differences associated with prenatal exposure to famine in humans. *Proc Natl Acad Sci U S A*, 2008. 105(44): 17046–9.

Herbert, V. Experimental nutritional folate deficiency in man. *Trans Assoc Am Physicians*, 1962. 75: 307–20.

Herrera, E., Samper, E., and Blasco, M. A. Telomere shortening in mTR-/- embryos is associated with failure to close the neural tube. *Embo j*, 1999. 18(5): 1172–81.

Hildebrandt, M. R., Germain, D. R., Monckton, E. A., Brun, M., and Godbout, R. Ddx1 knockout results in transgenerational wild-type lethality in mice. *Sci Rep*, 2015. 5: 9829.

Hill, P. W. S., Leitch, H. G., Requena, C. E., Sun, Z., Amouroux, R., Roman-Trufero, M., Borkowska, M., Terragni, J., Vaisvila, R., Linnett, S., Bagci, H., Dharmalingham, G., Haberle, V., Lenhard, B., Zheng, Y., Pradhan, S., and Hajkova, P. Epigenetic reprogramming enables the transition from primordial germ cell to gonocyte. *Nature*, 2018. 555(7696): 392–396.

Hilton, I. B., D'Ippolito, A. M., Vockley, C. M., Thakore, P. I., Crawford, G. E., Reddy, T. E., and Gersbach, C. A. Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nat Biotechnol*, 2015. 33(5): 510–7.

Hiraoka, M. and Kagawa, Y. Genetic polymorphisms and folate status. *Congenital anomalies*, 2017. 57(5): 142–149.

Hisano, M., Erkek, S., Dessus-Babus, S., Ramos, L., Stadler, M. B., and Peters, A. H. Genome-wide chromatin analysis in mature mouse and human spermatozoa. *Nat Protoc*, 2013. 8(12): 2449–70.

Ho, P. I., Ashline, D., Dhitavat, S., Ortiz, D., Collins, S. C., Shea, T. B., and Rogers, E. Folate deprivation induces neurodegeneration: roles of oxidative stress and increased homocysteine. *Neurobiol Dis*, 2003. 14(1): 32–42.

Hoivik, E. A., Bjanesoy, T. E., Mai, O., Okamoto, S., Minokoshi, Y., Shima, Y., Morohashi, K.-i., Boehm, U., and Bakke, M. DNA Methylation of Intronic Enhancers Directs Tissue-Specific Expression of Steroidogenic Factor 1/Adrenal 4 Binding Protein (SF-1/Ad4BP). *Endocrinology*, 2011. 152(5): 2100–2112.

Holliday, R. and Grigg, G. W. DNA methylation and mutation. *Mutat Res*, 1993. 285(1): 61–7.

Holoch, D. and Moazed, D. RNA-mediated epigenetic regulation of gene expression. *Nat Rev Genet*, 2015. 16(2): 71–84.

Howard, D. R., Brown, J. M., Todd, S., and Gregory, W. M. Recommendations on multiple testing adjustment in multi-arm trials with a shared control group. *Stat Methods Med Res*, 2018. 27(5): 1513–1530.

Hrelia, P., Fimognari, C., Maffei, F., Vigagni, F., Mesirca, R., Pozzetti, L., Paolini, M., and Cantelli Forti, G. The genetic and non-genetic toxicity of the fungicide Vinclozolin. *Mutagenesis*, 1996. 11(5): 445–53.

Huang da, W., Sherman, B. T., and Lempicki, R. A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res*, 2009a. 37(1): 1–13.

Huang da, W., Sherman, B. T., and Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*, 2009b. 4(1): 44–57.

Hussain, S., Tuorto, F., Menon, S., Blanco, S., Cox, C., Flores, J. V., Watt, S., Kudo, N. R., Lyko, F., and Frye, M. The mouse cytosine-5 RNA methyltransferase NSun2 is a component of the chromatoid body and required for testis differentiation. *Mol Cell Biol*, 2013. 33(8): 1561–70.

Hwang, I.-Y., Kwak, S., Lee, S., Kim, H., Lee, S., Kim, J.-H., Kim, Y., Jeon, Y., Chung, D., Jin, X., Park, S., Jang, H., Cho, E.-J., and Youn, H.-D. Psat1-Dependent Fluctuations in Alpha-Ketoglutarate Affect the Timing of ESC Differentiation. *Cell Metabolism*, 2016. 24(3): 494–501.

Imbard, A., Benoist, J.-F., and Blom, H. J. Neural tube defects, folic acid and methylation. *International journal of environmental research and public health*, 2013. 10(9): 4352–4389.

Iqbal, K., Tran, D. A., Li, A. X., Warden, C., Bai, A. Y., Singh, P., Wu, X., Pfeifer, G. P., and Szabo, P. E. Deleterious effects of endocrine disruptors are corrected in the mammalian germline by epigenome reprogramming. *Genome Biol*, 2015. 16: 59.

Jablonka, E. Epigenetic inheritance and plasticity: The responsive germline. *Prog Biophys Mol Biol*, 2013. 111(2-3): 99–107.

Jacob, R. A., Gretz, D. M., Taylor, P. C., James, S. J., Pogribny, I. P., Miller, B. J., Henning, S. M., and Swendseid, M. E. Moderate folate depletion increases plasma homocysteine and decreases lymphocyte DNA methylation in postmenopausal women. *J Nutr*, 1998. 128(7): 1204–12.

Jenkins, T. G., James, E. R., Alonso, D. F., Hoidal, J. R., Murphy, P. J., Hotaling, J. M., Cairns, B. R., Carrell, D. T., and Aston, K. I. Cigarette smoking significantly alters sperm DNA methylation patterns. *Andrology*, 2017.

Jimenez-Chillaron, J. C., Isganaitis, E., Charalambous, M., Gesta, S., Pentinat-Pelegrin, T., Faucette, R. R., Otis, J. P., Chow, A., Diaz, R., Ferguson-Smith, A.,

and Patti, M.-E. Intergenerational Transmission of Glucose Intolerance and Obesity by In Utero Undernutrition in Mice. *Diabetes*, 2009. 58(2): 460–468.

Johnson, G. D., Sendler, E., Lalancette, C., Hauser, R., Diamond, M. P., and Krawetz, S. A. Cleavage of rRNA ensures translational cessation in sperm at fertilization. *Mol Hum Reprod*, 2011. 17(12): 721–6.

Kaati, G., Bygren, L. O., and Edvinsson, S. Cardiovascular and diabetes mortality determined by nutrition during parents' and grandparents' slow growth period. *Eur J Hum Genet*, 2002. 10(11): 682–8.

Kankel, M. W., Ramsey, D. E., Stokes, T. L., Flowers, S. K., Haag, J. R., Jeddeloh, J. A., Riddle, N. C., Verbsky, M. L., and Richards, E. J. Arabidopsis MET1 cytosine methyltransferase mutants. *Genetics*, 2003. 163(3): 1109–22.

Kayagaki, N., Warming, S., Lamkanfi, M., Vande Walle, L., Louie, S., Dong, J., Newton, K., Qu, Y., Liu, J., Heldens, S., Zhang, J., Lee, W. P., Roose-Girma, M., and Dixit, V. M. Non-canonical inflammasome activation targets caspase-11. *Nature*, 2011. 479(7371): 117–21.

Kazachenka, A., Bertozzi, T. M., Sjoberg-Herrera, M. K., Walker, N., Gardner, J., Gunning, R., Pahita, E., Adams, S., Adams, D., and Ferguson-Smith, A. C. Identification, Characterization, and Heritability of Murine Metastable Epialleles: Implications for Non-genetic Inheritance. *Cell*, 2018. 175(5): 1259–1271.

Keane, T. M., Goodstadt, L., Danecek, P., White, M. A., Wong, K., Yalcin, B., Heger, A., Agam, A., Slater, G., Goodson, M., Furlotte, N. A., Eskin, E., Nellaker, C., Whitley, H., Cleak, J., Janowitz, D., Hernandez-Pliego, P., Edwards, A., Belgard, T. G., Oliver, P. L., McIntyre, R. E., Bhomra, A., Nicod, J., Gan, X., Yuan, W., van der Weyden, L., Steward, C. A., Bala, S., Stalker, J., Mott, R., Durbin, R., Jackson, I. J., Czechanski, A., Guerra-Assuncao, J. A., Donahue, L. R., Reinholdt, L. G., Payseur, B. A., Ponting, C. P., Birney, E., Flint, J., and Adams, D. J. Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature*, 2011. 477(7364): 289–94.

Kelly, T. L., Neaga, O. R., Schwahn, B. C., Rozen, R., and Trasler, J. M. Infertility in 5,10-methylenetetrahydrofolate reductase (MTHFR)-deficient male mice is

partially alleviated by lifetime dietary betaine supplementation. *Biol Reprod*, 2005. 72(3): 667–77.

Kiani, J., Grandjean, V., Liebers, R., Tuorto, F., Ghanbarian, H., Lyko, F., Cuzin, F., and Rassoulzadegan, M. RNA-mediated epigenetic heredity requires the cytosine methyltransferase Dnmt2. *PLoS Genet*, 2013. 9(5): e1003498.

Kim, S., Gunesdogan, U., Zylicz, J. J., Hackett, J. A., Cougot, D., Bao, S., Lee, C., Dietmann, S., Allen, G. E., Sengupta, R., and Surani, M. A. PRMT5 protects genomic integrity during global DNA demethylation in primordial germ cells and preimplantation embryos. *Mol Cell*, 2014. 56(4): 564–79.

Kipling, D. and Cooke, H. J. Hypervariable ultra-long telomeres in mice. *Nature*, 1990. 347(6291): 400–2.

Klosin, A., Casas, E., Hidalgo-Carcedo, C., Vavouri, T., and Lehner, B. Transgenerational transmission of environmental information in C. elegans. *Science*, 2017. 356(6335): 320–323.

Knott, A. B. and Bossy-Wetzel, E. Nitric oxide in health and disease of the nervous system. *Antioxid Redox Signal*, 2009. 11(3): 541–54.

Kobayashi, H., Sakurai, T., Imai, M., Takahashi, N., Fukuda, A., Yayoi, O., Sato, S., Nakabayashi, K., Hata, K., Sotomaru, Y., Suzuki, Y., and Kono, T. Contribution of intragenic DNA methylation in mouse gametic DNA methylomes to establish oocyte-specific heritable marks. *PLoS Genet*, 2012. 8(1): e1002440.

Kobayashi, H., Sakurai, T., Miura, F., Imai, M., Mochiduki, K., Yanagisawa, E., Sakashita, A., Wakai, T., Suzuki, Y., Ito, T., Matsui, Y., and Kono, T. High-resolution DNA methylome analysis of primordial germ cells identifies gender-specific reprogramming in mice. *Genome Res*, 2013. 23(4): 616–27.

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

Kooistra, M., Trasler, J. M., and Baltz, J. M. Folate Transport in Mouse Cumulus-Oocyte Complexes and Preimplantation Embryos. *Biology of Reproduction*, 2013. 89(3): 63, 1–9.

Kota, S. K. and Feil, R. Epigenetic transitions in germ cell development and meiosis. *Dev Cell*, 2010. 19(5): 675–86.

Kowluru, R. A., Shen, Y., and Mishra, M. Dynamic DNA methylation of matrix metalloproteinase-9 in the development of diabetic retinopathy. *Laboratory investigation; a journal of technical methods and pathology*, 2016. 96(10): 1040–1049.

Kozel, P. J., Friedman, R. A., Erway, L. C., Yamoah, E. N., Liu, L. H., Riddle, T., Duffy, J. J., Doetschman, T., Miller, M. L., Cardell, E. L., and Shull, G. E. Balance and hearing deficits in mice with a null mutation in the gene encoding plasma membrane Ca2+-ATPase isoform 2. *J Biol Chem*, 1998. 273(30): 18693–6.

Kozomara, A. and Griffiths-Jones, S. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Research*, 2014. 42(D1): D68–D73.

Krawetz, S. A. Paternal contribution: new insights and future challenges. *Nat Rev Genet*, 2005. 6(8): 633–42.

Kuan, P. F. and Chiang, D. Y. Integrating prior knowledge in multiple testing under dependence with applications to detecting differential DNA methylation. *Biometrics*, 2012. 68(3): 774–83.

Kumar, P., Kuscu, C., and Dutta, A. Biogenesis and Function of Transfer RNA-Related Fragments (tRFs). *Trends in Biochemical Sciences*, 2016. 41(8): 679–689.

Kumar, S. and Subramanian, S. Mutation rates in mammalian genomes. *Proc Natl Acad Sci U S A*, 2002. 99(2): 803–8.

Kwan, E. X., Wang, X. S., Amemiya, H. M., Brewer, B. J., and Raghuraman, M. K. rDNA Copy Number Variants Are Frequent Passenger Mutations in

Saccharomyces cerevisiae Deletion Collections and de Novo Transformants. *G3: Genes—Genomes—Genetics*, 2016. 6(9): 2829.

Kwong, W. Y., Adamiak, S. J., Gwynn, A., Singh, R., and Sinclair, K. D. Endogenous folates and single-carbon metabolism in the ovarian follicle, oocyte and pre-implantation embryo. *Reproduction*, 2010. 139(4): 705–15.

La Salle, S., Mertineit, C., Taketo, T., Moens, P. B., Bestor, T. H., and Trasler, J. M. Windows for sex-specific methylation marked by DNA methyltransferase expression profiles in mouse germ cells. *Developmental Biology*, 2004. 268(2): 403–415.

Lambrot, R., Xu, C., Saint-Phar, S., Chountalos, G., Cohen, T., Paquet, M., Suderman, M., Hallett, M., and Kimmins, S. Low paternal dietary folate alters the mouse sperm epigenome and is associated with negative pregnancy outcomes. *Nat Commun*, 2013. 4: 2889.

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

Largaespada, D. A. Transposon Mutagenesis in Mice. *Methods in molecular biology (Clifton, N.J.)*, 2009. 530: 379–390.

Larsen, L., Scheike, T., Jensen, T. K., Bonde, J. P., Ernst, E., Hjollund, N. H., Zhou, Y., Skakkebk, N. E., Giwercman, A., and The Danish First Pregnancy Planner Study, T. Computer-assisted semen analysis parameters as predictors for fertility of men from the general population. *Human Reproduction*, 2000. 15(7): 1562–1567.

Lawrence, M., Daujat, S., and Schneider, R. Lateral Thinking: How Histone Modifications Regulate Gene Expression. *Trends in Genetics*, 2016. 32(1): 42–56.

Le Scouarnec, S. and Gribble, S. M. Characterising chromosome rearrangements: recent technical advances in molecular cytogenetics. *Heredity*, 2012. 108(1): 75–85.

Lee, C. S., May, N. R., and Fan, C.-M. Transdifferentiation of the Ventral Retinal Pigmented Epithelium to Neural Retina in the Growth Arrest Specific Gene 1 Mutant. *Developmental Biology*, 2001. 236(1): 17–29.

Lee, H.-C., Jeong, Y.-M., Lee, S. H., Cha, K. Y., Song, S.-H., Kim, N. K., Lee, K. W., and Lee, S. Association study of four polymorphisms in three folate-related enzyme genes with non-obstructive male infertility. *Human Reproduction*, 2006. 21(12): 3162–3170.

Lee, H. J., Hore, T. A., and Reik, W. Reprogramming the methylome: erasing memory and creating diversity. *Cell Stem Cell*, 2014. 14(6): 710–9.

Lee, T.-L., Xiao, A., and Rennert, O. M. Identification of novel long noncoding RNA transcripts in male germ cells. *Methods in molecular biology (Clifton, N.J.)*, 2012. 825: 105–114.

Lev Maor, G., Yearim, A., and Ast, G. The alternative role of DNA methylation in splicing regulation. *Trends Genet*, 2015. 31(5): 274–80.

Li, H. and Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 2009. 25(14): 1754–60.

Li, Y. and Jiao, J. Histone chaperone HIRA regulates neural progenitor cell proliferation and neurogenesis via Beta-catenin. *The Journal of Cell Biology*, 2017. 216(7): 1975.

Liebers, R., Rassoulzadegan, M., and Lyko, F. Epigenetic regulation by heritable RNA. *PLoS Genet*, 2014. 10(4): e1004296.

Lienhard, M., Grimm, C., Morkel, M., Herwig, R., and Chavez, L. MEDIPS: genome-wide differential coverage analysis of sequencing data derived from DNA enrichment experiments. *Bioinformatics*, 2014. 30(2): 284–286.

Lin, C. J., Koh, F. M., Wong, P., Conti, M., and Ramalho-Santos, M. Hira-mediated H3.3 incorporation is required for DNA replication and ribosomal RNA transcription in the mouse zygote. *Dev Cell*, 2014. 30(3): 268–79.

Liu, Y., Liu, C., Yamada, Y., and Fan, C. M. Growth arrest specific gene 1 acts as a region-specific mediator of the Fgf10/Fgf8 regulatory loop in the limb. *Development*, 2002. 129(22): 5289–300.

Livak, K. J. and Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods*, 2001. 25(4): 402–8.

Loukinov, D. I., Pugacheva, E., Vatolin, S., Pack, S. D., Moon, H., Chernukhin, I., Mannan, P., Larsson, E., Kanduri, C., Vostrov, A. A., Cui, H., Niemitz, E. L., Rasko, J. E., Docquier, F. M., Kistler, M., Breen, J. J., Zhuang, Z., Quitschke, W. W., Renkawitz, R., Klenova, E. M., Feinberg, A. P., Ohlsson, R., Morse, H. C., r., and Lobanenkov, V. V. BORIS, a novel male germ-line-specific protein associated with epigenetic reprogramming events, shares the same 11-zinc-finger domain with CTCF, the insulator protein involved in reading imprinting marks in the soma. *Proc Natl Acad Sci U S A*, 2002. 99(10): 6806–11.

Love, M. I., Huber, W., and Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 2014. 15(12): 550.

Lucifero, D., Mertineit, C., Clarke, H. J., Bestor, T. H., and Trasler, J. M. Methylation Dynamics of Imprinted Genes in Mouse Germ Cells. *Genomics*, 2002. 79(4): 530–538.

Lumey, L. H., Stein, A. D., and Ravelli, A. C. Timing of prenatal starvation in women and birth weight in their first and second born offspring: the Dutch Famine Birth Cohort study. *Eur J Obstet Gynecol Reprod Biol*, 1995. 61(1): 23–30.

Luo, C., Hajkova, P., and Ecker, J. R. Dynamic DNA methylation: In the right place at the right time. *Science*, 2018. 361(6409): 1336–1340.

Lusis, A. J., Yu, J., and Wang, S. S. The problem of passenger genes in transgenic mice. *Arterioscler Thromb Vasc Biol*, 2007. 27(10): 2100–3.

Ma, F., Li, W., Tang, R., Liu, Z., Ouyang, S., Cao, D., Li, Y., and Wu, J. Long Non-Coding RNA Expression Profiling in Obesity Mice with Folic Acid Supplement. *Cellular Physiology and Biochemistry*, 2017. 42(1): 416–426.

MacFarlane, A. J., Perry, C. A., Girnary, H. H., Gao, D., Allen, R. H., Stabler, S. P., Shane, B., and Stover, P. J. Mthfd1 is an essential gene in mice and alters biomarkers of impaired one-carbon metabolism. *J Biol Chem*, 2009. 284(3): 1533–9.

Maeder, M. L., Angstman, J. F., Richardson, M. E., Linder, S. J., Cascio, V. M., Tsai, S. Q., Ho, Q. H., Sander, J. D., Reyon, D., Bernstein, B. E., Costello, J. F., Wilkinson, M. F., and Joung, J. K. Targeted DNA demethylation and activation of endogenous genes using programmable TALE-TET1 fusion proteins. *Nat Biotechnol*, 2013. 31(12): 1137–42.

Mahajan, V. S., Demissie, E., Mattoo, H., Viswanadham, V., Varki, A., Morris, R., and Pillai, S. Striking Immune Phenotypes in Gene-Targeted Mice Are Driven by a Copy-Number Variant Originating from a Commercially Available C57BL/6 Strain. *Cell Rep*, 2016. 15(9): 1901–9.

Maloney, C. A., Hay, S. M., and Rees, W. D. Folate deficiency during pregnancy impacts on methyl metabolism without affecting global DNA methylation in the rat fetus. *Br J Nutr*, 2007. 97(6): 1090–8.

Manikkam, M., Tracey, R., Guerrero-Bosagna, C., and Skinner, M. K. Plastics derived endocrine disruptors (BPA, DEHP and DBP) induce epigenetic transgenerational inheritance of obesity, reproductive disease and sperm epimutations. *PLoS One*, 2013. 8(1): e55387.

Marcho, C., Cui, W., and Mager, J. Epigenetic dynamics during preimplantation development. *Reproduction*, 2015. 150(3): R109–20.

Marmorstein, R. and Trievel, R. C. Histone Modifying Enzymes: Structures, Mechanisms, and Specificities. *Biochim Biophys Acta*, 2009. 1789(1): 58–68.

Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, 2011. 17(1).

Matherly, L. H. Molecular and cellular biology of the human reduced folate carrier. *Prog Nucleic Acid Res Mol Biol*, 2001. 67: 131–62.

McCarrey, J. R., Lehle, J. D., Raju, S. S., Wang, Y., Nilsson, E. E., and Skinner, M. K. Tertiary Epimutations - A Novel Aspect of Epigenetic Transgenerational Inheritance Promoting Genome Instability. *PLoS One*, 2016. 11(12): e0168038.

McGee, J., Goodyear, R. J., McMillan, D. R., Stauffer, E. A., Holt, J. R., Locke, K. G., Birch, D. G., Legan, P. K., White, P. C., Walsh, E. J., and Richardson, G. P. The very large G-protein-coupled receptor VLGR1: a component of the ankle link complex required for the normal development of auditory hair bundles. *J Neurosci*, 2006. 26(24): 6543–53.

McGrath, J. and Solter, D. Completion of mouse embryogenesis requires both the maternal and paternal genomes. *Cell*, 1984. 37(1): 179–83.

McKay, J. A. and Mathers, J. C. Maternal folate deficiency and metabolic dysfunction in offspring. *Proc Nutr Soc*, 2015. 75(1): 90–95.

McKay, J. A., Wong, Y. K., Relton, C. L., Ford, D., and Mathers, J. C. Maternal folate supply and sex influence gene-specific DNA methylation in the fetal gut. *Mol Nutr Food Res*, 2011. 55(11): 1717–23.

McKay, J. A., Xie, L., Adriaens, M., Evelo, C. T., Ford, D., and Mathers, J. C. Maternal folate depletion during early development and high fat feeding from weaning elicit similar changes in gene expression, but not in DNA methylation, in adult offspring. *Mol Nutr Food Res*, 2016.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M. A. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*, 2010. 20.

McStay, B. and Grummt, I. The Epigenetics of rRNA Genes: From Molecular to Chromosome Biology. *Annual Review of Cell and Developmental Biology*, 2008. 24(1): 131–157.

Messerschmidt, D. M. Should I stay or should I go: protection and maintenance of DNA methylation at imprinted genes. *Epigenetics*, 2012. 7(9): 969–75.

Messerschmidt, D. M., Knowles, B. B., and Solter, D. DNA methylation dynamics during epigenetic reprogramming in the germline and preimplantation embryos. *Genes Dev*, 2014. 28(8): 812–28.

Meyer, L. R., Zweig, A. S., Hinrichs, A. S., Karolchik, D., Kuhn, R. M., Wong, M., Sloan, C. A., Rosenbloom, K. R., Roe, G., Rhead, B., Raney, B. J., Pohl, A., Malladi, V. S., Li, C. H., Lee, B. T., Learned, K., Kirkup, V., Hsu, F., Heitner, S., Harte, R. A., Haeussler, M., Guruvadoo, L., Goldman, M., Giardine, B. M., Fujita, P. A., Dreszer, T. R., Diekhans, M., Cline, M. S., Clawson, H., Barber, G. P., Haussler, D., and Kent, W. J. The UCSC Genome Browser database: extensions and updates 2013. *Nucleic Acids Res*, 2013. 41(Database issue): D64–9.

Mfady, D. S., Sadiq, M. F., Khabour, O. F., Fararjeh, A. S., Abu-Awad, A., and Khader, Y. Associations of variants in MTHFR and MTRR genes with male infertility in the Jordanian population. *Gene*, 2014. 536(1): 40–4.

Mikael, L. G., Deng, L., Paul, L., Selhub, J., and Rozen, R. Moderately high intake of folic acid has a negative impact on mouse embryonic development. *Birth Defects Research Part A: Clinical and Molecular Teratology*, 2012. 97(1): 47–52.

Mikael, L. G., Pancer, J., Jiang, X., Wu, Q., Caudill, M., and Rozen, R. Low dietary folate and methylenetetrahydrofolate reductase deficiency may lead to pregnancy complications through modulation of ApoAI and IFN-gamma in spleen and placenta, and through reduction of methylation potential. *Mol Nutr Food Res*, 2013. 57(4): 661–70.

Miller, L. L., Henderson, J., Northstone, K., Pembrey, M., and Golding, J. Do grandmaternal smoking patterns influence the etiology of childhood asthma? *Chest*, 2014. 145(6): 1213–8.

Miltenberger, R. J., Mynatt, R. L., Wilkinson, J. E., and Woychik, R. P. The role of the agouti gene in the yellow obese syndrome. *J Nutr*, 1997. 127(9): 1902S–1907S.

Miska, E. A. and Ferguson-Smith, A. C. Transgenerational inheritance: Models and mechanisms of nonDNA sequencebased inheritance. *Science*, 2016. 354(6308): 59.

Momb, J., Lewandowski, J. P., Bryant, J. D., Fitch, R., Surman, D. R., Vokes, S. A., and Appling, D. R. Deletion of Mthfd1l causes embryonic lethality and neural tube and craniofacial defects in mice. *Proc Natl Acad Sci U S A*, 2013. 110(2): 549–54.

Montjean, D., Benkhalifa, M., Dessolle, L., Cohen-Bacrie, P., Belloc, S., Siffroi, J. P., Ravel, C., Bashamboo, A., and McElreavey, K. Polymorphisms in MTHFR and MTRR genes associated with blood plasma homocysteine concentration and sperm counts. *Fertil Steril*, 2011. 95(2): 635–40.

Morgan, H. D., Sutherland, H. G. E., Martin, D. I. K., and Whitelaw, E. Epigenetic inheritance at the agouti locus in the mouse. *Nature Genetics*, 1999. 23(3): 314–318.

Moss, T., Mars, J.-C., Tremblay, M. G., and Sabourin-Felix, M. The chromatin landscape of the ribosomal RNA genes in mouse and human. *Chromosome Research*, 2019. 27(1): 31–40.

Murphy, L. E., Mills, J. L., Molloy, A. M., Qian, C., Carter, T. C., Strevens, H., Wide-Swensson, D., Giwercman, A., and Levine, R. J. Folate and vitamin B(12) in idiopathic male infertility. *Asian Journal of Andrology*, 2011. 13(6): 856–861.

Nadler, A., Kav-Venaki, S., and Gleitman, B. Transgenerational effects of the holocaust: externalization of aggression in second generation of holocaust survivors. *J Consult Clin Psychol*, 1985. 53(3): 365–9.

Nakamura, T., Arai, Y., Umehara, H., Masuhara, M., Kimura, T., Taniguchi, H., Sekimoto, T., Ikawa, M., Yoneda, Y., Okabe, M., Tanaka, S., Shiota, K.,

and Nakano, T. PGC7/Stella protects against DNA demethylation in early embryogenesis. *Nat Cell Biol*, 2007. 9(1): 64–71.

Nashun, B., Hill, P. W., Smallwood, S. A., Dharmalingam, G., Amouroux, R., Clark, S. J., Sharma, V., Ndjetehe, E., Pelczar, P., Festenstein, R. J., Kelsey, G., and Hajkova, P. Continuous Histone Replacement by Hira Is Essential for Normal Transcriptional Regulation and De Novo DNA Methylation during Mouse Oogenesis. *Mol Cell*, 2015. 60(4): 611–25.

Nersisyan, L. and Arakelyan, A. Computel: Computation of Mean Telomere Length from Whole-Genome Next-Generation Sequencing Data. *PLoS ONE*, 2015. 10(4): e0125201.

Neumuller, R. A., Gross, T., Samsonova, A. A., Vinayagam, A., Buckner, M., Founk, K., Hu, Y., Sharifpoor, S., Rosebrock, A. P., Andrews, B., Winston, F., and Perrimon, N. Conserved regulators of nucleolar size revealed by global phenotypic analyses. *Sci Signal*, 2013. 6(289): ra70.

Ni, W., Li, H., Wu, A., Zhang, P., Yang, H., Yang, X., Huang, X., and Jiang, L. Lack of association between genetic polymorphisms in three folate-related enzyme genes and male infertility in the Chinese population. *Journal of Assisted Reproduction and Genetics*, 2015. 32(3): 369–374.

Nilsson, E., King, S. E., McBirney, M., Kubsad, D., Pappalardo, M., Beck, D., Sadler-Riggleman, I., and Skinner, M. K. Vinclozolin induced epigenetic transgenerational inheritance of pathologies and sperm epimutation biomarkers for specific diseases. *PLoS One*, 2018. 13(8): e0202662.

Oakes, C. C., La Salle, S., Smiraglia, D. J., Robaire, B., and Trasler, J. M. Developmental acquisition of genome-wide DNA methylation occurs prior to meiosis in male germ cells. *Dev Biol*, 2007. 307(2): 368–79.

O'Doherty, A. M. and McGettigan, P. A. Epigenetic processes in the male germline. *Reprod Fertil Dev*, 2015. 27(5): 725–38.

Oey, H., Isbel, L., Hickey, P., Ebaid, B., and Whitelaw, E. Genetic and epigenetic variation among inbred mouse littermates: identification of inter-individual differentially methylated regions. *Epigenetics Chromatin*, 2015. 8: 54.

O'Neill, C. Endogenous folic acid is essential for normal development of preimplantation embryos. *Hum Reprod*, 1998. 13(5): 1312–6.

O'Neill, K., Liao, W.-W., Patel, A., and Hammell, M. G. TEsmall Identifies Small RNAs Associated With Targeted Inhibitor Resistance in Melanoma. *Frontiers in Genetics*, 2018. 9: 461.

Ostermeier, G. C., Dix, D. J., Miller, D., Khatri, P., and Krawetz, S. A. Spermatozoal RNA profiles of normal fertile men. *Lancet*, 2002. 360(9335): 772–7.

Ostermeier, G. C., Miller, D., Huntriss, J. D., Diamond, M. P., and Krawetz, S. A. Reproductive biology: delivering spermatozoan RNA to the oocyte. *Nature*, 2004. 429(6988): 154.

Oswald, J., Engemann, S., Lane, N., Mayer, W., Olek, A., Fundele, R., Dean, W., Reik, W., and Walter, J. Active demethylation of the paternal genome in the mouse zygote. *Curr Biol*, 2000. 10(8): 475–8.

Otani, J., Nankumo, T., Arita, K., Inamoto, S., Ariyoshi, M., and Shirakawa, M. Structural basis for recognition of H3K4 methylation status by the DNA methyltransferase 3A ATRX-DNMT3-DNMT3L domain. *EMBO Rep*, 2009. 10(11): 1235–41.

Padmanabhan, N., Jia, D., Geary-Joo, C., Wu, X., Ferguson-Smith, A. C., Fung, E., Bieda, M. C., Snyder, F. F., Gravel, R. A., Cross, J. C., and Watson, E. D. Mutation in folate metabolism causes epigenetic instability and transgenerational effects on development. *Cell*, 2013. 155(1): 81–93.

Padmanabhan, N., Menelaou, K., Gao, J., Anderson, A., Blake, G. E. T., Li, T., Daw, B. N., and Watson, E. D. Abnormal folate metabolism causes age-, sex- and parent-of-origin-specific haematological defects in mice. *J Physiol*, 2018. 596(18): 4341–4360.

Padmanabhan, N., Rakoczy, J., Kondratowicz, M., Menelaou, K., Blake, G. E. T., and Watson, E. D. Multigenerational analysis of sex-specific phenotypic differences at midgestation caused by abnormal folate metabolism. *Environmental Epigenetics*, 2017. 3(4).

Painter, R. C., Osmond, C., Gluckman, P., Hanson, M., Phillips, D. I., and Roseboom, T. J. Transgenerational effects of prenatal exposure to the Dutch famine on neonatal adiposity and health in later life. *Bjog*, 2008. 115(10): 1243–9.

Park, C. B., Asin-Cayuela, J., Camara, Y., Shi, Y., Pellegrini, M., Gaspari, M., Wibom, R., Hultenby, K., Erdjument-Bromage, H., Tempst, P., Falkenberg, M., Gustafsson, C. M., and Larsson, N. G. MTERF3 is a negative regulator of mammalian mtDNA transcription. *Cell*, 2007. 130(2): 273–85.

Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., and Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods*, 2017. 14(4): 417–419.

Peirce, E. and Breed, W. A comparative study of sperm production in two species of Australian arid zone rodents (Pseudomys australis, Notomys alexis) with marked differences in testis size. *Reproduction*, 2001. 121(2): 239–47.

Pekarsky, Y., Balatti, V., Palamarchuk, A., Rizzotto, L., Veneziano, D., Nigita, G., Rassenti, L. Z., Pass, H. I., Kipps, T. J., Liu, C.-G., and Croce, C. M. Dysregulation of a family of short noncoding RNAs, tsRNAs, in human cancer. *Proceedings of the National Academy of Sciences of the United States of America*, 2016. 113(18): 5071–5076.

Pembrey, M. E. Male-line transgenerational responses in humans. *Hum Fertil (Camb)*, 2010. 13(4): 268–71.

Pikor, L., Thu, K., Vucic, E., and Lam, W. The detection and implication of genome instability in cancer. *Cancer and Metastasis Reviews*, 2013. 32(3-4): 341–352.

Pogribny, I. P., Kutanzi, K., Melnyk, S., de Conti, A., Tryndyak, V., Montgomery, B., Pogribna, M., Muskhelishvili, L., Latendresse, J. R., James, S. J., Beland, F. A., and Rusyn, I. Strain-dependent dysregulation of one-carbon metabolism in male mice is associated with choline- and folate-deficient diet-induced liver injury. *Faseb j*, 2013. 27(6): 2233–43.

Prakash, S. K., Cormier, T. A., McCall, A. E., Garcia, J. J., Sierra, R., Haupt, B., Zoghbi, H. Y., and Van Den Veyver, I. B. Loss of holocytochrome c-type synthetase causes the male lethality of X-linked dominant microphthalmia with linear skin defects (MLS) syndrome. *Hum Mol Genet*, 2002. 11(25): 3237–48.

Prochenka, A., Pokarowski, P., Gasperowicz, P., Kosiska, J., Stawiski, P., Zbie-Piekarska, R., Splnicka, M., Branicki, W., and Poski, R. A cautionary note on using binary calls for analysis of DNA methylation. *Bioinformatics*, 2015. 31(9): 1519–1520.

Quenneville, S., Verde, G., Corsinotti, A., Kapopoulou, A., Jakobsson, J., Offner, S., Baglivo, I., Pedone, P. V., Grimaldi, G., Riccio, A., and Trono, D. In embryonic stem cells, ZFP57/KAP1 recognize a methylated hexanucleotide to affect chromatin and DNA methylation of imprinting control regions. *Mol Cell*, 2011. 44(3): 361–72.

Quinlan, A. R. and Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 2010. 26(6): 841–842.

Radford, E. J., Ito, M., Shi, H., Corish, J. A., Yamazawa, K., Isganaitis, E., Seisenberger, S., Hore, T. A., Reik, W., Erkek, S., Peters, A. H. F. M., Patti, M.-E., and Ferguson-Smith, A. C. In utero undernourishment perturbs the adult sperm methylome and intergenerational metabolism. *Science*, 2014. 345(6198): 785–793.

Rakoczy, J., Padmanabhan, N., Krzak, A. M., Kieckbusch, J., Cindrova-Davies, T., and Watson, E. D. Dynamic expression of TET1, TET2, and TET3 dioxygenases in mouse and human placentas throughout gestation. *Placenta*, 2017. 59: 46–56.

Rakyan, V. K., Chong, S., Champ, M. E., Cuthbert, P. C., Morgan, H. D., Luu, K. V., and Whitelaw, E. Transgenerational inheritance of epigenetic states at the murine Axin(Fu) allele occurs after maternal and paternal transmission. *Proc Natl Acad Sci U S A*, 2003. 100(5): 2538–43.

Rameix-Welti, M. A., Le Goffic, R., Herve, P. L., Sourimant, J., Remot, A., Riffault, S., Yu, Q., Galloux, M., Gault, E., and Eleouet, J. F. Visualizing the replication of respiratory syncytial virus in cells and in living mice. *Nat Commun*, 2014. 5: 5104.

Rando, O. J. Intergenerational Transfer of Epigenetic Information in Sperm. *Cold Spring Harb Perspect Med*, 2016. 6(5).

Rasmussen, K. D. and Helin, K. Role of TET enzymes in DNA methylation, development, and cancer. *Genes Dev*, 2016. 30(7): 733–50.

Ratnam, M., Marquardt, H., Duhring, J. L., and Freisheim, J. H. Homologous membrane folate binding proteins in human placenta: cloning and sequence of a cDNA. *Biochemistry*, 1989. 28(20): 8249–54.

Ravelli, A. C., van Der Meulen, J. H., Osmond, C., Barker, D. J., and Bleker, O. P. Obesity at the age of 50 y in men and women exposed to famine prenatally. *Am J Clin Nutr*, 1999. 70(5): 811–6.

Ray, J. G. and Laskin, C. A. Folic acid and homocyst(e)ine metabolic defects and the risk of placental abruption, pre-eclampsia and spontaneous pregnancy loss: A systematic review. *Placenta*, 1999. 20(7): 519–29.

Ray, P. F., Toure, A., Metzler-Guillemain, C., Mitchell, M. J., Arnoult, C., and Coutton, C. Genetic abnormalities leading to qualitative defects of sperm morphology or function. *Clin Genet*, 2017. 91(2): 217–232.

Ray-Gallet, D., Ricketts, M. D., Sato, Y., Gupta, K., Boyarchuk, E., Senda, T., Marmorstein, R., and Almouzni, G. Functional activity of the H3.3 histone chaperone complex HIRA requires trimerization of the HIRA subunit. *Nature Communications*, 2018. 9: 3103.

Rechavi, O., Houri-Ze'evi, L., Anava, S., Goh, W. S., Kerk, S. Y., Hannon, G. J., and Hobert, O. Starvation-induced transgenerational inheritance of small RNAs in C. elegans. *Cell*, 2014. 158(2): 277–87.

Rechavi, O. and Lev, I. Principles of Transgenerational Small RNA Inheritance in Caenorhabditis elegans. *Curr Biol*, 2017. 27(14): R720–r730.

Reik, W., Dean, W., and Walter, J. Epigenetic reprogramming in mammalian development. *Science*, 2001. 293(5532): 1089–93.

Reik, W. and Surani, M. A. Germline and Pluripotent Stem Cells. *Cold Spring Harb Perspect Biol*, 2015. 7(11).

Ribeiro, A., Golicz, A., Hackett, C. A., Milne, I., Stephen, G., Marshall, D., Flavell, A. J., and Bayer, M. An investigation of causes of false positive single nucleotide polymorphisms using simulated reads from a small eukaryote genome. *BMC Bioinformatics*, 2015. 16(1): 382.

Richmond, R. C., Sharp, G. C., Herbert, G., Atkinson, C., Taylor, C., Bhattacharya, S., Campbell, D., Hall, M., Kazmi, N., Gaunt, T., McArdle, W., Ring, S., Davey Smith, G., Ness, A., and Relton, C. L. The long-term impact of folic acid in pregnancy on offspring DNA methylation: follow-up of the Aberdeen Folic Acid Supplementation Trial (AFAST). *International Journal of Epidemiology*, 2018. 47(3): 928–937.

Roberts, C., Sutherland, H. F., Farmer, H., Kimber, W., Halford, S., Carey, A., Brickman, J. M., Wynshaw-Boris, A., and Scambler, P. J. Targeted Mutagenesis of the Hira Gene Results in Gastrulation Defects and Patterning Abnormalities of Mesoendodermal Derivatives Prior to Early Embryonic Lethality. *Molecular and Cellular Biology*, 2002. 22(7): 2318–2328.

Robertson, K. D. and A.Jones, P. DNA methylation: past, present and future directions. *Carcinogenesis*, 2000. 21(3): 461–467.

Robledo, C. A., Yeung, E., Mendola, P., Sundaram, R., Maisog, J., Sweeney, A. M., Barr, D. B., and Louis, G. M. B. Preconception maternal and paternal exposure to persistent organic pollutants and birth size: the LIFE study. *Environmental health perspectives*, 2015. 123(1): 88–94.

Rosenthal, J., Largaespada, N., Bailey, L. B., Cannon, M., Alverson, C. J., Ortiz, D., Kauwell, G. P., Sniezek, J., Figueroa, R., Daly, R., and Allen, P. Folate Deficiency Is Prevalent in Women of Childbearing Age in Belize and Is Negatively Affected by Coexisting Vitamin B-12 Deficiency: Belize National Micronutrient Survey 2011. *The Journal of nutrition*, 2017. 147(6): 1183–1193.

Saben, J. L., Boudoures, A. L., Asghar, Z., Thompson, A., Drury, A., Zhang, W., Chi, M., Cusumano, A., Scheaffer, S., and Moley, K. H. Maternal Metabolic Syndrome Programs Mitochondrial Dysfunction via Germline Changes across Three Generations. *Cell Reports*, 2016. 16(1): 1–8.

Saleh, R. N. M., Dilg, D., Abou Zeid, A. A., Hashad, D. I., Scambler, P. J., and Chapgier, A. L. A. HIRA directly targets the enhancers of selected cardiac transcription factors during in vitro differentiation of mouse embryonic stem cells. *Molecular Biology Reports*, 2018. 45(5): 1001–1011.

Santos, F., Hendrich, B., Reik, W., and Dean, W. Dynamic reprogramming of DNA methylation in the early mouse embryo. *Dev Biol*, 2002. 241(1): 172–82.

Santos, F., Peat, J., Burgess, H., Rada, C., Reik, W., and Dean, W. Active demethylation in mouse zygotes involves cytosine deamination and base excision repair. *Epigenetics Chromatin*, 2013. 6(1): 39.

Sasaki, E., Susa, K., Mori, T., Isobe, K., Araki, Y., Inoue, Y., Yoshizaki, Y., Ando, F., Mori, Y., Mandai, S., Zeniya, M., Takahashi, D., Nomura, N., Rai, T., Uchida, S., and Sohara, E. KLHL3 Knockout Mice Reveal the Physiological Role of KLHL3 and the Pathophysiology of Pseudohypoaldosteronism Type II Caused by Mutant KLHL3. *Mol Cell Biol*, 2017. 37(7).

Scholl, T. O. and Johnson, W. G. Folic acid: influence on the outcome of pregnancy. *Am J Clin Nutr*, 2000. 71(5 Suppl): 1295s–303s.

Schulte, R. T., Ohl, D. A., Sigman, M., and Smith, G. D. Sperm DNA damage in male infertility: etiologies, assays, and outcomes. *Journal of assisted reproduction and genetics*, 2010. 27(1): 3–12.

Schuster, A., Skinner, M. K., and Yan, W. Ancestral vinclozolin exposure alters the epigenetic transgenerational inheritance of sperm small noncoding RNAs. *Environ Epigenet*, 2016. 2(1).

Seelige, R., Natsch, C., Marz, S., Jing, D., Frye, M., Butz, S., and Vestweber, D. Cutting edge: Endothelial-specific gene ablation of CD99L2 impairs leukocyte extravasation in vivo. *J Immunol*, 2013. 190(3): 892–6.

Seisenberger, S., Andrews, S., Krueger, F., Arand, J., Walter, J., Santos, F., Popp, C., Thienpont, B., Dean, W., and Reik, W. The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells. *Mol Cell*, 2012. 48(6): 849–62.

Seki, Y., Hayashi, K., Itoh, K., Mizugaki, M., Saitou, M., and Matsui, Y. Extensive and orderly reprogramming of genome-wide chromatin modifications associated with specification and early development of germ cells in mice. *Dev Biol*, 2005. 278(2): 440–58.

Senner, C. E., Krueger, F., Oxley, D., Andrews, S., and Hemberger, M. DNA methylation profiles define stem cell identity and reveal a tight embryonic-extraembryonic lineage boundary. *Stem Cells*, 2012. 30(12): 2732–45.

Shane, B. and Stokstad, E. L. Vitamin B12-folate interrelationships. *Annu Rev Nutr*, 1985. 5: 115–41.

Sharma, U., Conine, C. C., Shea, J. M., Boskovic, A., Derr, A. G., Bing, X. Y., Belleannee, C., Kucukural, A., Serra, R. W., Sun, F., Song, L., Carone, B. R., Ricci, E. P., Li, X. Z., Fauquier, L., Moore, M. J., Sullivan, R., Mello, C. C., Garber, M., and Rando, O. J. Biogenesis and function of tRNA fragments during sperm maturation and fertilization in mammals. *Science*, 2016. 351(6271): 391–6.

Sharpe, R. M. Perinatal Determinants of Adult Testis Size and Function. *The Journal of Clinical Endocrinology Metabolism*, 2006. 91(7): 2503–2505.

Shaul, Y. D. and Seger, R. The MEK/ERK cascade: From signaling specificity to diverse functions. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 2007. 1773(8): 1213–1226.

Shea, J. M., Serra, R. W., Carone, B. R., Shulha, H. P., Kucukural, A., Ziller, M. J., Vallaster, M. P., Gu, H., Tapper, A. R., Gardner, P. D., Meissner, A., Garber, M., and Rando, O. J. Genetic and Epigenetic Variation, but Not Diet, Shape the Sperm Methylome. *Dev Cell*, 2015. 35(6): 750–8.

Shen, L., Kondo, Y., Guo, Y., Zhang, J., Zhang, L., Ahmed, S., Shu, J., Chen, X., Waterland, R. A., and Issa, J. P. Genome-wide profiling of DNA methylation

reveals a class of normally methylated CpG island promoters. *PLoS Genet*, 2007. 3(10): 2023–36.

Shi, J., Ko, E.-A., Sanders, K. M., Chen, Q., and Zhou, T. SPORTS1.0: A Tool for Annotating and Profiling Non-coding RNAs Optimized for rRNA- and tRNA-derived Small RNAs. *Genomics, Proteomics Bioinformatics*, 2018. 16(2): 144–151.

Shindo, T., Doi, S., Nakashima, A., Sasaki, K., Arihiro, K., and Masaki, T. TGF-1 promotes expression of fibrosis-related genes through the induction of histone variant H3.3 and histone chaperone HIRA. *Scientific Reports*, 2018. 8(1): 14060.

Short, A. K., Yeshurun, S., Powell, R., Perreau, V. M., Fox, A., Kim, J. H., Pang, T. Y., and Hannan, A. J. Exercise alters mouse sperm small noncoding RNAs and induces a transgenerational modification of male offspring conditioned fear and anxiety. *Translational Psychiatry*, 2017. 7: e1114.

Shu, J. H., Zhang, B., Feng, G. X., Gan, X. Y., Zhou, H., Zhou, L., and Liu, Y. Influence of sperm morphology on the outcomes and neonatal status in IVF-ET. *Zhonghua Nan Ke Xue*, 2010. 16(10): 897–900.

Sie, K. K., Li, J., Ly, A., Sohn, K. J., Croxford, R., and Kim, Y. I. Effect of maternal and postweaning folic acid supplementation on global and gene-specific DNA methylation in the liver of the rat offspring. *Mol Nutr Food Res*, 2013. 57(4): 677–85.

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., and Kimmins, S. Disruption of histone methylation in developing sperm impairs offspring health transgenerationally. *Science*, 2015.

Silva, C., Keating, E., and Pinto, E. The impact of folic acid supplementation on gestational and long term health: Critical temporal windows, benefits and risks. *Porto Biomedical Journal*, 2017. 2(6): 315–332.

Simon, M. M., Greenaway, S., White, J. K., Fuchs, H., Gailus-Durner, V., Wells, S., Sorg, T., Wong, K., Bedu, E., Cartwright, E. J., Dacquin, R., Djebali, S.,

Estabel, J., Graw, J., Ingham, N. J., Jackson, I. J., Lengeling, A., Mandillo, S., Marvel, J., Meziane, H., Preitner, F., Puk, O., Roux, M., Adams, D. J., Atkins, S., Ayadi, A., Becker, L., Blake, A., Brooker, D., Cater, H., Champy, M.-F., Combe, R., Danecek, P., di Fenza, A., Gates, H., Gerdin, A.-K., Golini, E., Hancock, J. M., Hans, W., Hlter, S. M., Hough, T., Jurdic, P., Keane, T. M., Morgan, H., Mller, W., Neff, F., Nicholson, G., Pasche, B., Roberson, L.-A., Rozman, J., Sanderson, M., Santos, L., Selloum, M., Shannon, C., Southwell, A., Tocchini-Valentini, G. P., Vancollie, V. E., Westerberg, H., Wurst, W., Zi, M., Yalcin, B., Ramirez-Solis, R., Steel, K. P., Mallon, A.-M., de Angelis, M. H., Herault, Y., and Brown, S. D. M. A comparative phenotypic and genomic analysis of C57BL/6J and C57BL/6N mouse strains. *Genome biology*, 2013. 14(7): R82–R82.

Sims, D., Sudbery, I., Ilott, N. E., Heger, A., and Ponting, C. P. Sequencing depth and coverage: key considerations in genomic analyses. *Nat Rev Genet*, 2014. 15(2): 121–132.

Sinclair, K. D., Allegrucci, C., Singh, R., Gardner, D. S., Sebastian, S., Bispham, J., Thurston, A., Huntley, J. F., Rees, W. D., Maloney, C. A., Lea, R. G., Craigon, J., McEvoy, T. G., and Young, L. E. DNA methylation, insulin resistance, and blood pressure in offspring determined by maternal periconceptional B vitamin and methionine status. *Proceedings of the National Academy of Sciences*, 2007. 104(49): 19351.

Singh, K. and Jaiswal, D. One-carbon metabolism, spermatogenesis, and male infertility. *Reprod Sci*, 2013. 20(6): 622–30.

Skinner, M. K. and Guerrero-Bosagna, C. Role of CpG deserts in the epigenetic transgenerational inheritance of differential DNA methylation regions. *BMC Genomics*, 2014. 15: 692.

Skinner, M. K., Guerrero-Bosagna, C., and Haque, M. M. Environmentally induced epigenetic transgenerational inheritance of sperm epimutations promote genetic mutations. *Epigenetics*, 2015. 10(8): 762–71.

Skinner, M. K., Manikkam, M., Haque, M. M., Zhang, B., and Savenkova, M. I. Epigenetic transgenerational inheritance of somatic transcriptomes and epigenetic control regions. *Genome Biol*, 2012. 13(10): R91.

Slotkin, R. K. and Martienssen, R. Transposable elements and the epigenetic regulation of the genome. *Nat Rev Genet*, 2007. 8(4): 272–85.

Smallwood, S. A., Tomizawa, S., Krueger, F., Ruf, N., Carli, N., Segonds-Pichon, A., Sato, S., Hata, K., Andrews, S. R., and Kelsey, G. Dynamic CpG island methylation landscape in oocytes and preimplantation embryos. *Nat Genet*, 2011. 43(8): 811–4.

Smith, C. L., Blake, J. A., Kadin, J. A., Richardson, J. E., and Bult, C. J. Mouse Genome Database (MGD)-2018: knowledgebase for the laboratory mouse. *Nucleic Acids Res*, 2018. 46(D1): D836–d842.

Smith, L. B. and Walker, W. H. The regulation of spermatogenesis by androgens. *Semin Cell Dev Biol*, 2014. 30: 2–13.

Smith, Z. D., Chan, M. M., Mikkelsen, T. S., Gu, H., Gnirke, A., Regev, A., and Meissner, A. A unique regulatory phase of DNA methylation in the early mammalian embryo. *Nature*, 2012. 484(7394): 339–44.

Soppe, W. J., Jacobsen, S. E., Alonso-Blanco, C., Jackson, J. P., Kakutani, T., Koornneef, M., and Peeters, A. J. The late flowering phenotype of fwa mutants is caused by gain-of-function epigenetic alleles of a homeodomain gene. *Mol Cell*, 2000. 6(4): 791–802.

Stadler, M. B., Murr, R., Burger, L., Ivanek, R., Lienert, F., Scholer, A., van Nimwegen, E., Wirbelauer, C., Oakeley, E. J., Gaidatzis, D., Tiwari, V. K., and Schubeler, D. DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature*, 2011. 480(7378): 490–5.

Staunstrup, N. H., Starnawska, A., Nyegaard, M., Christiansen, L., Nielsen, A. L., Borglum, A., and Mors, O. Genome-wide DNA methylation profiling with MeDIP-seq using archived dried blood spots. *Clin Epigenetics*, 2016. 8: 81.

Stewart, K. R., Veselovska, L., and Kelsey, G. Establishment and functions of DNA methylation in the germline. *Epigenomics*, 2016. 8(10): 1399–1413.

Stewart, K. R., Veselovska, L., Kim, J., Huang, J., Saadeh, H., Tomizawa, S., Smallwood, S. A., Chen, T., and Kelsey, G. Dynamic changes in histone modifications precede de novo DNA methylation in oocytes. *Genes Dev*, 2015. 29(23): 2449–62.

Svingen, T. and Koopman, P. Building the mammalian testis: origins, differentiation, and assembly of the component cell populations. *Genes Development*, 2013. 27(22): 2409–2426.

Swanson, D. A., Liu, M. L., Baker, P. J., Garrett, L., Stitzel, M., Wu, J., Harris, M., Banerjee, R., Shane, B., and Brody, L. C. Targeted disruption of the methionine synthase gene in mice. *Mol Cell Biol*, 2001. 21(4): 1058–65.

Swayne, B. G., Kawata, A., Behan, N. A., Williams, A., Wade, M. G., Macfarlane, A. J., and Yauk, C. L. Investigating the effects of dietary folic acid on sperm count, DNA damage and mutation in Balb/c mice. *Mutat Res*, 2012. 737(1-2): 1–7.

Szabo, R., Samson, A. L., Lawrence, D. A., Medcalf, R. L., and Bugge, T. H. Passenger mutations and aberrant gene expression in congenic tissue plasminogen activator-deficient mouse strains. *J Thromb Haemost*, 2016. 14(8): 1618–28.

Szenker, E., Ray-Gallet, D., and Almouzni, G. The double face of the histone variant H3.3. *Cell research*, 2011. 21(3): 421–434.

Taiwo, O., Wilson, G. A., Morris, T., Seisenberger, S., Reik, W., Pearce, D., Beck, S., and Butcher, L. M. Methylome analysis using MeDIP-seq with low DNA concentrations. *Nat Protoc*, 2012. 7(4): 617–36.

Tang, Y., Poustovoitov, M. V., Zhao, K., Garfinkel, M., Canutescu, A., Dunbrack, R., Adams, P. D., and Marmorstein, R. Structure of a human ASF1a/HIRA complex and insights into specificity of histone chaperone complex assembly. *Nature structural molecular biology*, 2006. 13(10): 921–929.

Tee, W. W. and Reinberg, D. Chromatin features and the epigenetic regulation of pluripotency states in ESCs. *Development*, 2014. 141(12): 2376–90.

Tirado-Magallanes, R., Rebbani, K., Lim, R., Pradhan, S., and Benoukraf, T. Whole genome DNA methylation: beyond genes silencing. *Oncotarget*, 2017. 8(3): 5629–5637.

Tobi, E. W., Slieker, R. C., Stein, A. D., Suchiman, H. E., Slagboom, P. E., van Zwet, E. W., Heijmans, B. T., and Lumey, L. H. Early gestation as the critical time-window for changes in the prenatal environment to affect the adult human blood methylome. *Int J Epidemiol*, 2015. 44(4): 1211–23.

Torres-Padilla, M. E., Bannister, A. J., Hurd, P. J., Kouzarides, T., and Zernicka-Goetz, M. Dynamic distribution of the replacement histone variant H3.3 in the mouse oocyte and preimplantation embryos. *Int J Dev Biol*, 2006. 50(5): 455–61.

Tost, J. and Gut, I. G. Analysis of gene-specific DNA methylation patterns by pyrosequencing technology. *Methods Mol Biol*, 2007. 373: 89–102.

Tracey, R., Manikkam, M., Guerrero-Bosagna, C., and Skinner, M. K. Hydrocarbons (jet fuel JP-8) induce epigenetic transgenerational inheritance of obesity, reproductive disease and sperm epimutations. *Reprod Toxicol*, 2013. 36: 104–16.

Trasler, J., Deng, L., Melnyk, S., Pogribny, I., Hiou-Tim, F., Sibani, S., Oakes, C., Li, E., James, S. J., and Rozen, R. Impact of Dnmt1 deficiency, with and without low folate diets, on tumor numbers and DNA methylation in Min mice. *Carcinogenesis*, 2003. 24(1): 39–45.

Tuorto, F., Liebers, R., Musch, T., Schaefer, M., Hofmann, S., Kellner, S., Frye, M., Helm, M., Stoecklin, G., and Lyko, F. RNA cytosine methylation by Dnmt2 and NSun2 promotes tRNA stability and protein synthesis. *Nat Struct Mol Biol*, 2012. 19(9): 900–5.

Uthus, E. O. and Brown-Borg, H. M. Methionine flux to transsulfuration is enhanced in the long living Ames dwarf mouse. *Mech Ageing Dev*, 2006. 127(5): 444–50.

Vallaster, M. P., Kukreja, S., Bing, X. Y., Ngolab, J., Zhao-Shea, R., Gardner, P. D., Tapper, A. R., and Rando, O. J. Paternal nicotine exposure alters hepatic xenobiotic metabolism in offspring. *eLife*, 2017. 6: e24771.

van der Heijden, G. W., Derijck, A. A., Ramos, L., Giele, M., van der Vlag, J., and de Boer, P. Transmission of modified nucleosomes from the mouse male germline to the zygote and subsequent remodeling of paternal chromatin. *Dev Biol*, 2006. 298(2): 458–69.

Vanden Berghe, T., Hulpiau, P., Martens, L., Vandenbroucke, R. E., Van Wonterghem, E., Perry, S. W., Bruggeman, I., Divert, T., Choi, S. M., Vuylsteke, M., Shestopalov, V. I., Libert, C., and Vandenabeele, P. Passenger Mutations Confound Interpretation of All Genetically Modified Congenic Mice. *Immunity*, 2015. 43(1): 200–9.

Vasta, V., Shimizu-Albergine, M., and Beavo, J. A. Modulation of Leydig cell function by cyclic nucleotide phosphodiesterase 8A. *Proceedings of the National Academy of Sciences*, 2006. 103(52): 19925.

Veenendaal, M. V., Painter, R. C., de Rooij, S. R., Bossuyt, P. M., van der Post, J. A., Gluckman, P. D., Hanson, M. A., and Roseboom, T. J. Transgenerational effects of prenatal exposure to the 1944-45 Dutch famine. *Bjog*, 2013. 120(5): 548–53.

Ventura-Junc, P., Irarrzaval, I., Rolle, A. J., Gutirrez, J. I., Moreno, R. D., and Santos, M. J. In vitro fertilization (IVF) in mammals: epigenetic and developmental alterations. Scientific and bioethical implications for IVF in humans. *Biological research*, 2015. 48: 68–68.

Veselovska, L., Smallwood, S. A., Saadeh, H., Stewart, K. R., Krueger, F., Maupetit-Mhouas, S., Arnaud, P., Tomizawa, S.-i., Andrews, S., and Kelsey, G. Deep sequencing and de novo assembly of the mouse oocyte transcriptome define the contribution of transcription to the DNA methylation landscape. *Genome Biology*, 2015. 16: 209.

Vojtech, L., Woo, S., Hughes, S., Levy, C., Ballweber, L., Sauteraud, R. P., Strobl, J., Westerberg, K., Gottardo, R., Tewari, M., and Hladik, F. Exosomes in

human semen carry a distinctive repertoire of small non-coding RNAs with potential regulatory functions. *Nucleic Acids Res*, 2014. 42(11): 7290–304.

Voutounou, M., Glen, C. D., and Dubrova, Y. E. The effects of methyl-donor deficiency on mutation induction and transgenerational instability in mice. *Mutat Res*, 2012. 734(1-2): 1–4.

Wallock, L. M., Tamura, T., Mayr, C. A., Johnston, K. E., Ames, B. N., and Jacob, R. A. Low seminal plasma folate concentrations are associated with low sperm density and count in male smokers and nonsmokers. *Fertility and Sterility*, 2001. 75(2): 252–259.

Wang, J., Raskin, L., Samuels, D. C., Shyr, Y., and Guo, Y. Genome measures used for quality control are dependent on gene function and ancestry. *Bioinformatics (Oxford, England)*, 2015. 31(3): 318–323.

Wang, Y. Epididymal sperm count. *Curr Protoc Toxicol*, 2003. Chapter 16: Unit16.6.

Wansleeben, C., van Gurp, L., de Graaf, P., Mousson, F., Marc Timmers, H. T., and Meijlink, F. An ENU-induced point mutation in the mouse Btaf1 gene causes post-gastrulation embryonic lethality and protein instability. *Mech Dev*, 2011. 128(5-6): 279–88.

Ward, W. S. and Zalensky, A. O. The unique, complex organization of the transcriptionally silent sperm chromatin. *Crit Rev Eukaryot Gene Expr*, 1996. 6(2-3): 139–47.

Wasson, G. R., McGlynn, A. P., McNulty, H., O'Reilly, S. L., McKelvey-Martin, V. J., McKerr, G., Strain, J. J., Scott, J., and Downes, C. S. Global DNA and p53 region-specific hypomethylation in human colonic cells is induced by folate depletion and reversed by folate supplementation. *J Nutr*, 2006. 136(11): 2748–53.

Waterland, R. A., Dolinoy, D. C., Lin, J. R., Smith, C. A., Shi, X., and Tahiliani, K. G. Maternal methyl supplements increase offspring DNA methylation at Axin Fused. *Genesis*, 2006. 44(9): 401–6.

Waterston, R. H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J. F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P., Antonarakis, S. E., Attwood, J., Baertsch, R., Bailey, J., Barlow, K., Beck, S., Berry, E., Birren, B., Bloom, T., Bork, P., Botcherby, M., Bray, N., Brent, M. R., Brown, D. G., Brown, S. D., Bult, C., Burton, J., Butler, J., Campbell, R. D., Carninci, P., Cawley, S., Chiaromonte, F., Chinwalla, A. T., Church, D. M., Clamp, M., Clee, C., Collins, F. S., Cook, L. L., Copley, R. R., Coulson, A., Couronne, O., Cuff, J., Curwen, V., Cutts, T., Daly, M., David, R., Davies, J., Delehaunty, K. D., Deri, J., Dermitzakis, E. T., Dewey, C., Dickens, N. J., Diekhans, M., Dodge, S., Dubchak, I., Dunn, D. M., Eddy, S. R., Elnitski, L., Emes, R. D., Eswara, P., Eyras, E., Felsenfeld, A., Fewell, G. A., Flicek, P., Foley, K., Frankel, W. N., Fulton, L. A., Fulton, R. S., Furey, T. S., Gage, D., Gibbs, R. A., Glusman, G., Gnerre, S., Goldman, N., Goodstadt, L., Grafham, D., Graves, T. A., Green, E. D., Gregory, S., Guigo, R., Guyer, M., Hardison, R. C., Haussler, D., Hayashizaki, Y., Hillier, L. W., Hinrichs, A., Hlavina, W., Holzer, T., Hsu, F., Hua, A., Hubbard, T., Hunt, A., Jackson, I., Jaffe, D. B., Johnson, L. S., Jones, M., Jones, T. A., Joy, A., Kamal, M., Karlsson, E. K., et al. Initial sequencing and comparative analysis of the mouse genome. *Nature*, 2002. 420(6915): 520–62.

Watkins, A. J., Dias, I., Tsuro, H., Allen, D., Emes, R. D., Moreton, J., Wilson, R., Ingram, R. J. M., and Sinclair, K. D. Paternal diet programs offspring health through sperm- and seminal plasma-specific pathways in mice. *Proc Natl Acad Sci U S A*, 2018.

Watkins-Chow, D. E. and Pavan, W. J. Genomic copy number and expression variation within the C57BL/6J inbred mouse strain. *Genome Res*, 2008. 18(1): 60–6.

Weber, M., Davies, J. J., Wittig, D., Oakeley, E. J., Haase, M., Lam, W. L., and Schubeler, D. Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nat Genet*, 2005. 37(8): 853–62.

Weber, M., Hellmann, I., Stadler, M. B., Ramos, L., Paabo, S., Rebhan, M., and Schubeler, D. Distribution, silencing potential and evolutionary impact of

promoter DNA methylation in the human genome. *Nat Genet*, 2007. 39(4): 457–66.

Wei, Y., Yang, C. R., Wei, Y. P., Zhao, Z. A., Hou, Y., Schatten, H., and Sun, Q. Y. Paternally induced transgenerational inheritance of susceptibility to diabetes in mammals. *Proc Natl Acad Sci U S A*, 2014. 111(5): 1873–8.

Wicker, T., Yu, Y., Haberer, G., Mayer, K. F. X., Marri, P. R., Rounsley, S., Chen, M., Zuccolo, A., Panaud, O., Wing, R. A., and Roffler, S. DNA transposon activity is associated with increased mutation rates in genes of rice and other grasses. *Nature Communications*, 2016. 7: 12790.

Wong, W. Y., Merkus, H. M., Thomas, C. M., Menkveld, R., Zielhuis, G. A., and Steegers-Theunissen, R. P. Effects of folic acid and zinc sulfate on male factor subfertility: a double-blind, randomized, placebo-controlled trial. *Fertil Steril*, 2002. 77(3): 491–8.

Wu, L., Yavas, G., Hong, H., Tong, W., and Xiao, W. Direct comparison of performance of single nucleotide variant calling in human genome with alignment-based and assembly-based approaches. *Scientific reports*, 2017. 7(1): 10963–10963.

Wyck, S., Herrera, C., Requena, C. E., Bittner, L., Hajkova, P., Bollwein, H., and Santoro, R. Oxidative stress in sperm affects the epigenetic reprogramming in early embryonic development. *Epigenetics Chromatin*, 2018. 11(1): 60.

Wyrobek, A. J., Gordon, L. A., Burkhart, J. G., Francis, M. W., Kapp, R. W., J., Letz, G., Malling, H. V., Topham, J. C., and Whorton, M. D. An evaluation of the mouse sperm morphology test and other sperm tests in nonhuman mammals. A report of the U.S. Environmental Protection Agency Gene-Tox Program. *Mutat Res*, 1983. 115(1): 1–72.

Xu, J. and Sinclair, K. D. One-carbon metabolism and epigenetic regulation of embryo development. *Reprod Fertil Dev*, 2015. 27(4): 667–76.

Yalcin, B., Wong, K., Agam, A., Goodson, M., Keane, T. M., Gan, X., Nellaker, C., Goodstadt, L., Nicod, J., Bhomra, A., Hernandez-Pliego, P., Whitley,

H., Cleak, J., Dutton, R., Janowitz, D., Mott, R., Adams, D. J., and Flint, J. Sequence-based characterization of structural variation in the mouse genome. *Nature*, 2011. 477(7364): 326–9.

Yamada, K., Gravel, R. A., Toraya, T., and Matthews, R. G. Human methionine synthase reductase is a molecular chaperone for human methionine synthase. *Proc Natl Acad Sci U S A*, 2006. 103(25): 9476–81.

Yehuda, R., Halligan, S. L., and Bierer, L. M. Relationship of parental trauma exposure and PTSD to PTSD, depressive and anxiety disorders in offspring. *J Psychiatr Res*, 2001. 35(5): 261–70.

Yoshikawa, M. and Fujii, Y. R. Human Ribosomal RNA-Derived Resident MicroRNAs as the Transmitter of Information upon the Cytoplasmic Cancer Stress. *BioMed Research International*, 2016. 2016: 14.

Yu, J. Y., Chung, K. H., Deo, M., Thompson, R. C., and Turner, D. L. MicroRNA miR-124 regulates neurite outgrowth during neuronal differentiation. *Exp Cell Res*, 2008. 314(14): 2618–33.

Yung, H.-w., Korolchuk, S., Tolkovsky, A. M., Charnock-Jones, D. S., and Burton, G. J. Endoplasmic reticulum stress exacerbates ischemia-reperfusion-induced apoptosis through attenuation of Akt protein synthesis in human choriocarcinoma cells. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*, 2007. 21(3): 872–884.

Zhang, P., Si, X., Skogerbo, G., Wang, J., Cui, D., Li, Y., Sun, X., Liu, L., Sun, B., Chen, R., He, S., and Huang, D. W. piRBase: a web resource assisting piRNA functional study. *Database (Oxford)*, 2014. 2014: bau110.

Zhang, X., Cozen, A. E., Liu, Y., Chen, Q., and Lowe, T. M. Small RNA Modifications: Integral to Function and Disease. *Trends Mol Med*, 2016. 22(12): 1025–1034.

Zhang, Y., Zhang, X., Shi, J., Tuorto, F., Li, X., Liu, Y., Liebers, R., Zhang, L., Qu, Y., Qian, J., Pahima, M., Liu, Y., Yan, M., Cao, Z., Lei, X., Cao, Y., Peng, H., Liu, S., Wang, Y., Zheng, H., Woolsey, R., Quilici, D., Zhai, Q., Li, L., Zhou, T., Yan,

W., Lyko, F., Zhang, Y., Zhou, Q., Duan, E., and Chen, Q. Dnmt2 mediates intergenerational transmission of paternally acquired metabolic disorders through sperm small non-coding RNAs. *Nature Cell Biology*, 2018. 20(5): 535–540.

Zhao, Q., Zhang, J., Chen, R., Wang, L., Li, B., Cheng, H., Duan, X., Zhu, H., Wei, W., Li, J., Wu, Q., Han, J.-D. J., Yu, W., Gao, S., Li, G., and Wong, J. Dissecting the precise role of H3K9 methylation in crosstalk with DNA maintenance methylation in mammals. *Nature Communications*, 2016. 7: 12464.

Zhu, B., Xiahou, Z., Zhao, H., Peng, B., Zhao, H., and Xu, X. MTHFR promotes heterochromatin maintenance. *Biochem Biophys Res Commun*, 2014. 447(4): 702–6.

# Chapter 9

# Appendix

## 9.1 MEDIPS script

This is a generalised version of the script used to perform differenital methylation analysis using the MEDIPS package in R (Lienhard et al., 2014).

```
# !/usr/local/bin/Rscript
# Analysis R-Script for MEDIPS Template
# Author: Gina Blake, based on script By Russell Hamilton

# source("http://bioconductor.org/biocLite.R")
# biocLite("MEDIPS")

library("MEDIPS")

# biocLite("BSgenome.Mmusculus.UCSC.mm10")

library("BSgenome.Mmusculus.UCSC.mm10")
seqlengths(BSgenome.Mmusculus.UCSC.mm10)

# Set the current working directory where the bam files are
wd < - "/Users/.."
setwd(wd)

# Some MeDIPs specific options
BSgenome = "BSgenome.Mmusculus.UCSC.mm10"
uniq = 1e-3
extend = 300
ws = 500
shift = 0
chr.select = c("chr1", "chr2", "chr3", "chr4", "chr5", "chr6", "chr7", "chr8",
"chr9", "chr10", "chr11", "chr12", "chr13", "chr14", "chr15", "chr16", "chr17",
"chr18", "chr19", "chrX", "chrY")
```

# Load in the bam files
# Group 1: C57 (sample 7 excluded)

Sample1 = ".bam"
Sample2 = ".bam"
Sample3 = ".bam"
Sample4 = ".bam"
Sample5 = ".bam"
Sample6 = ".bam"
Sample8 = ".bam"

# Group 2: *Mtrr* (*Mtrr*$^{+/+}$, *Mtrr*$^{+/gt}$or *Mtrr*$^{gt/gt}$samples)

Sample a = ".bam"
Sample b= ".bam"
Sample c= ".bam"
Sample d= ".bam"
Sample e= ".bam"
Sample f= ".bam"
Sample g= ".bam"
Sample h= ".bam"

C57 bam <- c( Sample1, Sample2, Sample3, Sample4, Sample5, Sample6, Sample8)

Mtrr bam <- c( Sample a, Sample b, Sample c, Sample d, Sample e, Sample f, Sample g, Sample h)

# Saturation QC Plots

# C57

sr1=MEDIPS.saturation(file= C57 bam , BSgenome = BSgenome, extend = extend, shift = shift, uniq = uniq, window_size = ws, nit=10, nrit=1, empty_ins=TRUE, rank=TRUE, paired = TRUE)

MEDIPS.plotSaturation(sr1)

# *Mtrr*

sr2=MEDIPS.saturation(file= Mtrr bam , BSgenome = BSgenome, extend = extend, shift = shift, uniq = uniq, window_size = ws, nit=10, nrit=1, empty_bins=TRUE, rank=TRUE, paired = TRUE)

MEDIPS.plotSaturation(sr2)

# Sequence pattern (”CG”) coverage analysis to plot Pie charts and histograms

# C57

for(i in 1:length(C57 bam))
{
cr = MEDIPS.seqCoverage(file = C57 bam[i], pattern = ”CG”, BSgenome = BSgenome, chr.select = chr.select, extend = extend, shift = shift, uniq =uniq)

pdf(paste(wd,as.character(C57 bam[i]),”_coveragePie.pdf”,sep=””),
width=7, height=7) MEDIPS.plotSeqCoverage(seqCoverageObj = cr, type = ”pie”, cov.level = c(0, 5, 10, 20, 25, 50, 100, 250))
dev.off()

pdf(paste(wd,as.character(C57 bam[i]),”_coverageHist.pdf”,sep=””), width=7, height=7)
MEDIPS.plotSeqCoverage(seqCoverageObj = cr, type = ”hist”, t = 100, main = ”Sequence pattern coverage, histogram”)
dev.off()

```
}

    # Mtrr

for(i in 1:length(Mtrr bam))
{
cr = MEDIPS.seqCoverage(file = Mtrr bam[i], pattern = ”CG”, BSgenome = BSgenome, chr.select = chr.select, extend = extend, shift = shift, uniq =uniq)

pdf(paste(wd,as.character(Mtrr bam[i]),”_coveragePie.pdf”,sep=””), width=7, height=7)
MEDIPS.plotSeqCoverage(seqCoverageObj = cr, type = ”pie”, cov.level = c(0, 5, 10, 20, 25, 50, 100, 250))
dev.off()

pdf(paste(wd,as.character(Mtrr bam[i]),”_coverageHist.pdf”,sep=””), width=7, height=7)
MEDIPS.plotSeqCoverage(seqCoverageObj = cr, type = ”hist”, t = 100, main = ”Sequence pattern coverage, histogram”)
dev.off()
}

# CpG Enrichment analysis

    # C57

er1 = MEDIPS.CpGenrich(file = Sample1, BSgenome = BSgenome,chr.select = chr.select, extend = extend, shift = shift, uniq = uniq)
er2 = MEDIPS.CpGenrich(file = Sample2, BSgenome = BSgenome,chr.select = chr.select, extend = extend, shift = shift, uniq = uniq)
er3 = MEDIPS.CpGenrich(file = Sample3, BSgenome = BSgenome,chr.select = chr.select, extend = extend, shift = shift, uniq = uniq)
er4 = MEDIPS.CpGenrich(file = Sample4, BSgenome = BSgenome,chr.select =
```

```
chr.select, extend = extend, shift = shift, uniq = uniq)
er5 = MEDIPS.CpGenrich(file = Sample5, BSgenome = BSgenome,chr.select =
chr.select, extend = extend, shift = shift, uniq = uniq)
er6 = MEDIPS.CpGenrich(file = Sample6, BSgenome = BSgenome,chr.select =
chr.select, extend = extend, shift = shift, uniq = uniq)
er7 = MEDIPS.CpGenrich(file = Sample8, BSgenome = BSgenome,chr.select =
chr.select, extend = extend, shift = shift, uniq = uniq)

    # Mtrr

er8 = MEDIPS.CpGenrich(file = Sample a, BSgenome = BSgenome,chr.select =
chr.select, extend = extend, shift = shift, uniq = uniq)
er9 = MEDIPS.CpGenrich(file = Sample b, BSgenome = BSgenome,chr.select =
chr.select, extend = extend, shift = shift, uniq = uniq)
er10 = MEDIPS.CpGenrich(file = Sample c, BSgenome = BSgenome,chr.select =
chr.select, extend = extend, shift = shift, uniq = uniq)
er11 = MEDIPS.CpGenrich(file = Sample d, BSgenome = BSgenome,chr.select
= chr.select, extend = extend, shift = shift, uniq = uniq)
er12 = MEDIPS.CpGenrich(file = Sample e, BSgenome = BSgenome,chr.select =
chr.select, extend = extend, shift = shift, uniq = uniq)
er13 = MEDIPS.CpGenrich(file = Sample f, BSgenome = BSgenome,chr.select =
chr.select, extend = extend, shift = shift, uniq = uniq)
er14 = MEDIPS.CpGenrich(file = Sample g, BSgenome = BSgenome,chr.select =
chr.select, extend = extend, shift = shift, uniq = uniq)
er15 = MEDIPS.CpGenrich(file = Sample h, BSgenome = BSgenome,chr.select =
chr.select, extend = extend, shift = shift, uniq = uniq)

# MSET Group 1 (C57)

    C57.MSET = MEDIPS.createSet(file = Sample1, BSgenome = BSgenome,
extend = extend, shift = shift, uniq = uniq, window_size = ws, chr.select =
chr.select, paired =TRUE)
```

C57.MSET = c(C57.MSET, MEDIPS.createSet(file = Sample2, BSgenome = BSgenome, extend = extend, shift = shift, uniq = uniq, window_size = ws, chr.select = chr.select, paired =TRUE))

C57.MSET = c(C57.MSET, MEDIPS.createSet(file = Sample3, BSgenome = BSgenome, extend = extend, shift = shift, uniq = uniq, window_size = ws, chr.select = chr.select, paired =TRUE))

C57.MSET = c(C57.MSET, MEDIPS.createSet(file = Sample4, BSgenome = BSgenome, extend = extend, shift = shift, uniq = uniq, window_size = ws, chr.select = chr.select, paired =TRUE))

C57.MSET = c(C57.MSET, MEDIPS.createSet(file = Sample5, BSgenome = BSgenome, extend = extend, shift = shift, uniq = uniq, window_size = ws, chr.select = chr.select, paired =TRUE))

C57.MSET = c(C57.MSET, MEDIPS.createSet(file = Sample6, BSgenome = BSgenome, extend = extend, shift = shift, uniq = uniq, window_size = ws, chr.select = chr.select, paired =TRUE))

C57.MSET = c(C57.MSET, MEDIPS.createSet(file = Sample8, BSgenome = BSgenome, extend = extend, shift = shift, uniq = uniq, window_size = ws, chr.select = chr.select, paired =TRUE))

# MSET Group 2 (Mtrr)

Mtrr.MSET = MEDIPS.createSet(file = Sample a, BSgenome = BSgenome,extend = extend, shift = shift, uniq = uniq, window_size = ws, chr.select = chr.select)

Mtrr.MSET = c(Mtrr.MSET, MEDIPS.createSet(file = Sample b, BSgenome = BSgenome,extend = extend, shift = shift, uniq = uniq, window_size = ws,chr.select = chr.select, paired =TRUE))

Mtrr.MSET = c(Mtrr.MSET, MEDIPS.createSet(file = Sample c, BSgenome = BSgenome,extend = extend, shift = shift, uniq = uniq, window_size = ws,chr.select = chr.select, paired =TRUE))

Mtrr.MSET = c(Mtrr.MSET, MEDIPS.createSet(file = Sample d, BSgenome = BSgenome,extend = extend, shift = shift, uniq = uniq, window_size = ws,chr.select = chr.select, paired =TRUE))

Mtrr.MSET = c(Mtrr.MSET, MEDIPS.createSet(file = Sample e, BSgenome = BSgenome,extend = extend, shift = shift, uniq = uniq, window_size = ws,chr.select = chr.select, paired =TRUE))

Mtrr.MSET = c(Mtrr.MSET, MEDIPS.createSet(file = Sample f, BSgenome = BSgenome,extend = extend, shift = shift, uniq = uniq, window_size = ws,chr.select = chr.select, paired =TRUE))

Mtrr.MSET = c(Mtrr.MSET, MEDIPS.createSet(file = Sample g, BSgenome = BSgenome,extend = extend, shift = shift, uniq = uniq, window_size = ws,chr.select = chr.select, paired =TRUE))

Mtrr.MSET = c(Mtrr.MSET, MEDIPS.createSet(file = Sample h, BSgenome = BSgenome,extend = extend, shift = shift, uniq = uniq, window_size = ws,chr.select = chr.select, paired =TRUE))

MSet_List = c(C57.MSET, Mtrr.MSET)

# Calculate coupling vector for CpG density dependent normalization of MeDIP-seq data

CS.C57 = MEDIPS.couplingVector(pattern = ”CG”, refObj = C57.MSET)

# Export wiggle file for the coupling set

```
MEDIPS.exportWIG(CSet = CS.C57, file = "CS.C57.wig", format = "pdensity",
descr = "Coupling Plot CS.C57")

# Calculate and plot a calibration curve

for(i in 1:length(Mtrr.MSET))
{
MEDIPS.plotCalibrationPlot(MSet = Mtrr.MSET[[i]], CSet = CS.C57, main=
"Calibrationplot.Mtrr", plot chr = "all", rpkm = TRUE, xrange = TRUE)
dev.off()
}

# Differential methylation between two conditions

mr.edgeR = MEDIPS.meth(MSet1 = C57.MSET, MSet2 = Mtrr.MSET, CSet =
CS.C57, p.adj = "bonferroni", diff.method = "edgeR", MeDIP = T, CNV = F,
minRowSum = 1)

# Select significant windows

mr.edgeR.s = MEDIPS.selectSig(results = mr.edgeR, p.value = 0.01, adj = T, ratio
= NULL, bg.counts = NULL, CNV = F)

dm1=mr.edgeR.s$edgeR.adj.p.value<0.01
dm2=mr.edgeR.s$edgeR.p.value<0.01
write.table(mr.edgeR.s[which(dm1),], paste(wd,"filename.tsv",sep=""), sep="t",
quote=F, row.names=F, col.names=T)

# Gain in Group.1 over Group.2

mr.edgeR.s.gain = mr.edgeR.s[which(mr.edgeR.s[, grep("logFC",
colnames(mr.edgeR.s))] > 0), ]
mr.edgeR.s.gain.m = MEDIPS.mergeFrames(frames=mr.edgeR.s.gain, distance=1)
```

```
dm1=mr.edgeR.s.gain.m$edgeR.adj.p.value<0.01
dm2=mr.edgeR.s.gain.m$edgeR.p.value<0.01
write.table(mr.edgeR.s.gain.m[which(dm1),], paste(wd,"filename.tsv",sep=""),
sep="t", quote=F, row.names=F, col.names=T)

# Loss in Group.1 over Group.2

mr.edgeR.s.loss = mr.edgeR.s[which(mr.edgeR.s[, grep("logFC",
colnames(mr.edgeR.s))] < 0), ]
mr.edgeR.s.loss.m = MEDIPS.mergeFrames(frames=mr.edgeR.s.loss, distance=1)
dm1=mr.edgeR.s.loss.m$edgeR.adj.p.value<0.01
dm2=mr.edgeR.s.loss.m$edgeR.p.value<0.01
write.table(mr.edgeR.s.loss.m[which(dm1),],paste(wd,"filename.tsv",sep=""),
sep="t", quote=F, row.names=F, col.names=T)

# END OF SCRIPT
```

CHAPTER

# EPIGENETICS OF TRANSGENERATIONAL INHERITANCE OF DISEASE

26

**Georgina E.T. Blake, Joanna Rakoczy, Erica D. Watson**
*University of Cambridge, Cambridge, United Kingdom*

## CHAPTER OUTLINE



**Epigenetics in Human Disease.** https://doi.org/10.1016/B978-0-12-812215-0.00026-1

## 26.1 INTRODUCTION

It is becoming clear that the environment in which we live influences our health and potentially the health of our descendants. Although environmental insults can lead to beneficial or disease-causing genetic mutations, inheritance is rare making it difficult for adaptation to keep pace with a rapidly fluctuating environment. Furthermore, there are many noncommunicable diseases in which a genetic cause is not present or not yet identified. An alternative mode of inheritance called "epigenetic inheritance" is quickly coming into the forefront as a mechanism for the inheritance of non-communicable diseases and predicting disease risk.

The epigenome is known to be responsive to environmental conditions (e.g., available nutrients, presence of toxicants, etc.) and, therefore, it is an excellent candidate to explain swift adaptation of populations to ever-changing surroundings. Epigenetic inheritance involves germline transmission of epigenetic "information" from one generation to the next, independent of the DNA base sequence. Currently, very little is understood regarding the mechanism of epigenetic inheritance because it was only recently discovered that epigenetic marks, such as DNA methylation, are not entirely wiped clean between each generation.

Although human populations have demonstrated increased disease risk on exposure of the parental or grandparental generations to environmental stressors, we do not understand the extent to which epigenetic information is inherited because very few epigenetic studies have been completed. Identifying a mechanism for epigenetic inheritance is challenging because it is difficult to separate genetic influences and parental physiology from inherited epigenetic factors. As a result, researchers have turned to plant and animal models to control for these confounding factors. There are an increasing number of exciting models of epigenetic inheritance that we, as researchers, can mechanistically explore. A greater mechanistic understanding of epigenetic inheritance will impact how disease prevention and prediction is approached.

## 26.2 DEFINING EPIGENETIC INHERITANCE

Epigenetic inheritance is the transmission of information between generations independent of the DNA base sequence [1,2]. DNA methylation, histone modifications, and noncoding RNA (ncRNA) are postulated as epigenetic vectors of information that can be carried in the germline to mediate epigenetic inheritance [1]. These factors are responsive to environmental stressors [1], allowing environmentally induced epigenetic changes to potentially be inherited via the germline. However, the mechanism remains unclear.

Exposure to an environmental stressor may induce epigenetic changes in the first generation (parental generation [F0]), leading to an increased susceptibility to certain phenotypes in subsequent generations (first filial generation [F1], second filial generation [F2], third filial generation [F3], etc.), even though they are not exposed to the environmental insult themselves [3]. We define this as transgenerational inheritance if phenotypes and/or epigenetic factors that cause the phenotypes persist into at least the F2 generation when inherited via the paternal lineage and at least the F3 generation when inherited via the maternal lineage [1,4] (Fig. 26.1). This is because exposure of a pregnant female (i.e., the F0 generation) to an environmental stressor results in the direct exposure of the fetus (i.e., the F1 generation) and its developing germ cells (i.e., the future F2 generation) to the insult. In this case, maternal physiology or direct exposure of the fetus or germ cells to the stressor cannot be excluded as a contributor to the phenotype or epimutation. As a result, inheritance up to the F2 generation in the maternal lineage is known as intergenerational inheritance (Fig. 26.1A). Transgenerational inheritance occurs when phenotypes persist to the F3 generation and beyond because these generations are not directly exposed to the insult (Fig. 26.1A). In the paternal lineage, exposure of an F0 male to an environmental insult also directly exposes his germ cells that will form the future F1 generation. Therefore, inheritance of the phenotype or epimutation to the F1 generation via the paternal lineage is termed intergenerational inheritance (Fig. 26.1B). Observation of the phenotype or altered epigenetic pattern in the F2 generation and beyond is considered to be transgenerational epigenetic inheritance because these individuals are not directly exposed to the environmental stressor (Fig. 26.1B).

To understand mechanisms of transgenerational epigenetic inheritance, it is vital to first understand the inheritance of epigenetic information from parent to child. Only then can we determine whether similar mechanisms are responsible for inheritance over subsequent generations. Many researchers using mammalian models focus on paternal inheritance because this allows the exclusion of confounding factors, such as the in utero environment and maternal care of the offspring, which are known to contribute to disease in the offspring (reviewed in Ref. [5]). Sperm passes on only genetic and epigenetic material to the offspring. However, additional subtle male influences also need to be considered, including seminal fluid composition and microbiome [6] and mate choice [7], which together may influence offspring health. Furthermore, genetic stability is intricately linked to epigenetic effects. It is affected by mutagenic chemicals (e.g., cigarette smoke) and represents an important potential cofounder in models of transgenerational epigenetic inheritance (see Section 26.13). Therefore, researchers must be mindful when designing experiments in animal models to study the mechanisms involved in epigenetic inheritance [8].

## 26.3 MULTIGENERATIONAL INHERITANCE OF HUMAN DISEASE

Although there is evidence that epigenetic inheritance may occur in human populations, multigenerational studies in humans are limited because of long generation times, widespread genetic diversity, and the availability of stringent information or tissue samples. Very few studies conclusively show epigenetic inheritance of disease let alone persistence of disease phenotypes over multiple generations. However, data showing parent-to-child transmission of disease risk is useful and might help us tease apart direct effects from epigenetic inheritance effects. Cohorts supporting the existence of epigenetic inheritance in humans focus on the effects of environmental stressors, such as poor nutrition (e.g.,

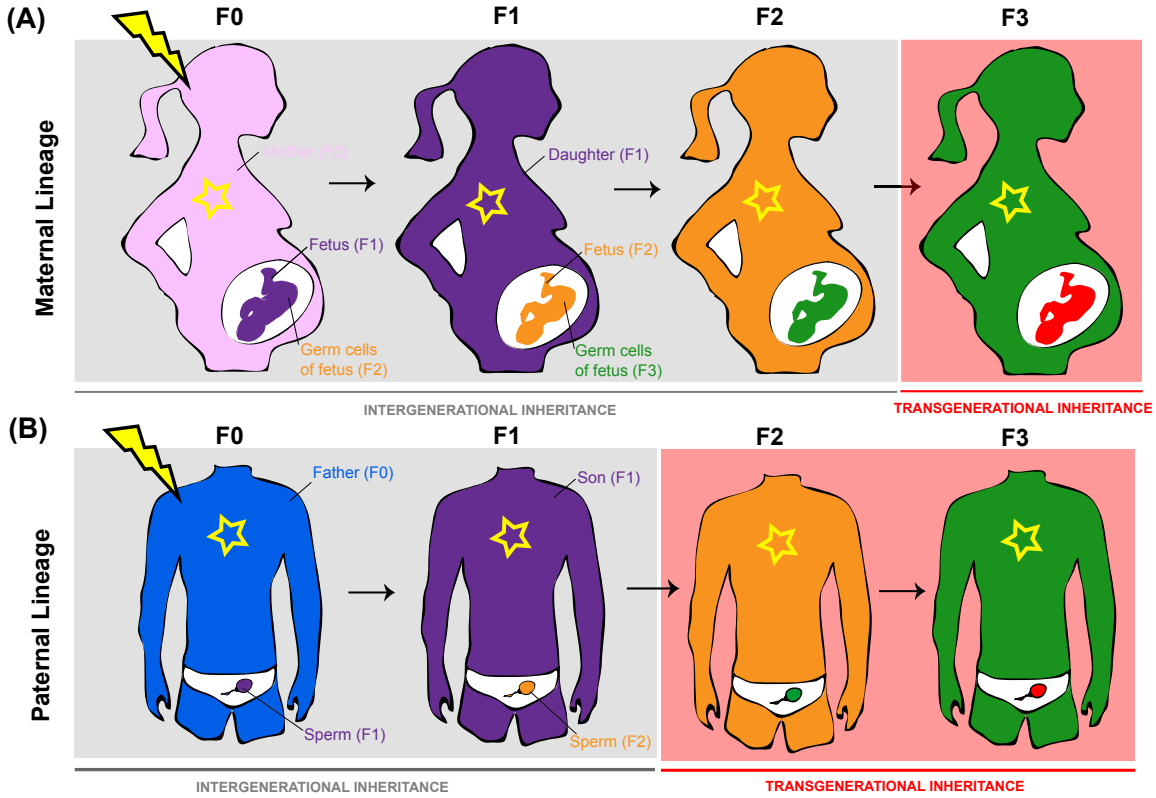



**FIGURE 26.1 Intergenerational and Transgenerational Epigenetic Inheritance.**

Epigenetic inheritance is induced by exposure of the parental (F0) generation to an environmental stressor (e.g., poor nutrition, toxicants, psychological stress, etc.; represented by *lightning bolt*) resulting in phenotypes and/or epimutations (represented by yellow star) that are inherited by subsequent generations (e.g., F1, F2, F3, etc.). Intergenerational epigenetic inheritance involves the appearance of phenotypes and/or epimutations in generations in which direct exposure of the stressor cannot be ruled out as a cause. Transgenerational epigenetic inheritance involves phenotypes and/or epimutations that persist in generations that are not directly exposed to the stressor. The number of generations that are required to pass before inheritance is considered to be transgenerational depends on whether inheritance is through the maternal or paternal lineage. (A) In the maternal lineage, exposure to an environmental stressor while an F0 female (pink) is pregnant results in direct exposure of the mother, fetus (the F1 generation; purple), and the fetus' developing primordial germ cells (the future F2 generation; orange) to the insult. If a phenotype or epimutation appears only until the F2 generation, it is termed intergenerational inheritance because a uterine factor or direct exposure to the stressor cannot be ruled out as a cause. The persistence of phenotypes and/or epimutations into the F3 generation (green) and beyond is considered transgenerational epigenetic inheritance because these individuals were not directly exposed to the insult, themselves. Instead, it is hypothesized that an abnormal epigenetic factor is inherited through the germline. (B) In the paternal lineage, F0 male (blue) exposure to an environmental stressor also directly exposes his germ cells (the future F1 generation; purple) to the insult. Therefore, observation of a phenotype or epimutation up to the F1 generation is termed intergenerational inheritance. If a phenotype and/or epimutation is observed in the F2 generation (orange) and beyond, this is this considered transgenerational inheritance because these individuals have not been directly exposed to the environmental stressor, themselves. *F0*, parental generation; *F1*, first filial generation; *F2*, second filial generation; *F3*, third filial generation.

famine or excess nutrients), environmental pollutants (e.g., smoking, toxicant exposure), and psychological insults (e.g., stress). Little is known about the molecular mechanism behind such findings, which is why studying animal and plant models is important. The most prominent findings of human multigenerational inheritance of disease are summarized below.

## 26.4 POTENTIAL STRESSORS LEADING TO EPIGENETIC INHERITANCE IN HUMANS

### 26.4.1 NUTRITION

David Barker and colleagues correlated fetal and placental weights with increased risk for cardiovascular disease and diabetes later in life, establishing the theory of the developmental origins of health and disease (Box 26.1) [9]. This theory suggests that the intrauterine environment in which the fetus is exposed (e.g., famine or a high-fat diet) will "developmentally program" the offspring leading to increased disease risk later in life. Initial investigations into the role of nutrition in fetal programming mainly focused on the Dutch Hunger Winter, which was a famine that occurred in Amsterdam from October 1944 to May 1945. Food rations with low nutritional value allocated to pregnant women resulted in sharp drop in infant birth weights [15,16], and remarkably, an increased risk for cardiovascular disease and metabolic syndrome later in adult life [12,13,17]. Increased disease risk can depend on the timing of in utero exposure (i.e., first trimester vs. third trimester). For example, the prevalence of type 2 diabetes was increased when babies were exposed to famine in the Ukraine during the first half of pregnancy compared with the second half of pregnancy [18]. In Austria, the number of people with type 2 diabetes peaked after times of famines (1918–19, 1938, and 1946–47) when they were exposed to undernutrition in utero during the first half of pregnancy [19]. Therefore, undernutrition during the first trimester of pregnancy appears to be the most sensitive time for fetal programming [12–14] likely because this is when most epigenetic patterns throughout the genome are being established as fetal cells differentiate [20]. Altogether, these studies indicate that the human epigenome may be sensitive to environmental stressors that may have lifelong consequences.

Extraordinarily, increased disease risk associated with in utero exposure to famine in human populations might be inherited between generations as well. For example, women that were exposed to

**BOX 26.1 DEVELOPMENTAL ORIGINS OF HEALTH AND DISEASE**

The phenomenon of developmental origins of health and disease (DOHaD), also known as the "Barker hypothesis," emerged from epidemiological studies comparing infant birth parameters with adult disease onset and mortality. The hypothesis was proposed by David Barker and colleagues in 1990 based on studies revealing that intrauterine growth restriction, low birth weight, and premature birth might lead to the origins of adult-onset hypertension, heart disease, and diabetes [9–11]. Therefore, undernutrition during gestation might "developmentally program" an individual such that the body's structure, function, and metabolism are permanently altered. Individuals exposed in utero to famine in early gestation had the highest cumulative incidence of coronary artery disease [12], predisposition to diabetes [12], increased female body weight, BMI, waist circumference [13], and breast cancer [14]. Altogether, these studies emphasize how the environment likely alters epigenetic status of an individual early in development with consequences later in life. The mechanism through which DOHaD occurs is likely epigenetic-based, whereby abnormal establishment of stable epigenetic marks during gestation alters gene expression, developmental pathways, and physiology. However, much more research is required to mechanistically understand developmental programming, particularly when these disease risks are inherited by subsequence generations.

famine during their own in utero development had first-born babies with increased birth weights [21]. In fact, the grandchildren of women, who were pregnant during famine, had increased neonatal adiposity [17]. In some cases, the multigenerational effects of undernutrition are sexually dimorphic. For instance, the granddaughters of women exposed to famine in utero had higher mean ponderal indices (relationship between weight and length) at birth compared with grandsons [17]. Although the mechanism for these multigenerational effects may be epigenetic in nature, it is possible that in utero exposure to poor nutrition leads to an atypical uterine environment in the F1 generation that ultimately programs the next generation for disease. Alternatively, when primordial germ cells develop and migrate to the fetal gonad between weeks 4–8 of gestation, they normally undergo extensive demethylation and erasure of imprints [22]. We speculate that environmental exposure during this period might alter the DNA methylation patterns of fetal germ cells, leading to the inheritance of an abnormal epigenome in the next generation. It is difficult to experimentally separate these two mechanistic possibilities.

Multigenerational effects of nutrition are also observed via the paternal lineage. In a human population in Overkalix, Sweden [23], detailed historical records of food supply of men during their slow-growth period in childhood, defined as the time before the onset of puberty, were linked to sex-specific causes of death of their offspring [24,25]. For instance, a good harvest year during a paternal grandfather's slow-growth period resulted in diabetes and reduced survival of his grandchildren [25]. However, paternal grandmothers in a similar situation had granddaughters with an increased risk of cardiovascular disease [23,26]. These studies identify another environmental exposure-sensitive period during development (i.e., slow-growth period) that may lead to the inheritance of disease risk in humans. The molecular/epigenetic mechanism through which this phenomenon occurs is not yet known. Fully understanding the mechanism is further complicated by differential effects of disease risk depending on which grandparent is exposed to the stressor and the sex of the grandchild being assessed.

Although it is impossible to collect tissue from retrospective studies focusing on historical famines to assess the inheritance of epigenetic factors between each generation, we might rely on another undernutrition-style insult in human populations caused by religious fasting to explore an epigenetic mechanism. During Ramadan both men and women abstain from food and water during the day and eat large high-fat/high-protein meals after sunset and before sunrise. Although pregnant women are exempt from practicing Ramadan, some pregnant women undergo fasting to be involved in the spiritual celebrations. A study of over 17,000 birth records in Saudi Arabia found that babies exposed to Ramadan fasting at any time during gestation had normal birth weights but low placenta weights and placenta to birth weight ratios [27], though this finding is inconsistent with another study [28]. Placental weight to birth weight ratios are used as an index of nutritional and placental efficiency, suggesting that Ramadan fasting may have a protective effect on placental function. This may be due to the fact that the caloric deficiency is not as severe as during famine. It will be important to establish prospective studies to understand the long-term and multigenerational effects of Ramadan fasting during pregnancy, with particular focus on epigenetic analysis to explore a possible mechanism involved in epigenetic inheritance of disease risk.

### 26.4.2 TOXICANT EXPOSURE: CIGARETTE SMOKING AND AIR POLLUTANTS

Recent focus has been placed on how toxicant exposure may alter the human epigenome leading to increased risk of asthma, cancer, and other diseases that may be epigenetically inherited. This is

particularly important to explore because of the wide range of toxicants in our environment. The Avon Longitudinal Study of Parents and Children concluded that smoking in teenage boys was associated with increased body mass index (BMI) in their daughters and sons at 9 years of age [23,29] and obesity in their granddaughters by age 12 [30]. Alternatively, nonsmoking women, who were exposed in utero to maternal smoking, had sons (but not daughters) with increased birth weights and BMI [31]. Further still, children with any in utero exposure to maternal smoking were at risk of asthma [32,33]. This risk increased if the child's maternal grandmother had also smoked during the fetal period of the child's mother [32]. However, if the paternal grandmother smoked, granddaughters had a greater risk of developing asthma [34]. These studies emphasize sexually dimorphic effects of cigarette smoke that result between parent−child transmission and intergenerational inheritance of phenotype. Whether these phenotypes are caused solely by epigenetic inheritance is unclear because, again, the effects of maternal physiology was not mechanistically excluded in these analyses. Furthermore, in each case, the germ cells that give rise to the F1 generation were also exposed to cigarette smoke suggesting the epigenetic contribution to disease risk would be a direct effect rather than inherited. Importantly, the potential mutagenic nature of some toxicants, such as cigarette smoke or air pollutants, means that underlying acquired mutations might arise to cause disease and/or to alter DNA methylation patterns (e.g., by changing a methylated CpG site to a base pair coupling that cannot be methylated) thus altering gene expression. Therefore, acquired genetic changes must also be considered when assessing environmental effects on the epigenome.

Recent studies into the effects of air pollution on offspring health have come to the forefront because air pollution has increased globally by 8% in the past 5 years, particularly in western countries (World Health Organization, 2016). So far, no studies have explored the intergenerational or transgenerational effects of air pollutants on health risk in humans. However, we might glean some understanding by assessing potential parental effects. When males (and their germ cells: the future F1 generation) were exposed to specific air pollutants, increased birth weight was evident in their daughters [35]. This study is important because it indicates air pollutants might alter the sperm epigenome, potentially affecting the development and health of the next generation. More experimentation is required to understand the effects of air pollutants (and other environmental toxicants) on the germ cell epigenome.

### 26.4.3 PSYCHOLOGICAL STRESS

It is becoming clear that the effects of psychological stress may be inherited. Studies focusing on the children of Holocaust survivors indicate that these children have increased anxiety and guilt [36]. In fact, there is a strong relationship between parental posttraumatic stress disorder (PTSD) and the occurrence of PTSD in their children, who were not involved in the Holocaust, themselves [37]. It was determined that plasma cortisol levels were lower in adult children of Holocaust survivors, particularly with maternal PTSD [38], indicating a sexually dimorphic effect. However, again, it is difficult to tease apart genetic and epigenetic influences from environmental causes (i.e., the mother's own mental state and the quality of maternal care) to fully understand how the inheritance of psychological stress risk affects offspring mental health. When designing animal based studies that address psychological stress, the type of maternal care associated with cross fostering in mice is a serious confounding factor that needs to be considered [8].

## 26.5 POTENTIAL MECHANISMS OF EPIGENETIC INHERITANCE IN HUMANS

Most information collected from studies of environmental stress leading to multigenerational inheritance of disease in human populations comes from detailed historical records, death certificates, questionnaires, and interviews. Limited molecular or epigenetic analysis has been conducted, though a few studies have investigated epigenetic biomarkers to provide insight into possible epigenetic mechanisms. This largely comprises assessment of DNA methylation patterns in offspring caused by exposure to undernutrition, toxinants, and psychological stress using directed locus-specific approaches. For example, in utero famine exposure of individuals correlates with decreased DNA methylation in the regulatory control region of the growth gene *IGF2* in blood, a change that was apparently maintained for over six decades [39]. This result was in comparison to *IGF2* methylation in same-sex siblings that had normal nutrition during gestation [39]. Others showed that increased DNA methylation at differentially methylated regions (DMRs) of *IGF2* genes in offspring of smoking mothers compared with nonsmoking mothers [40]. Interestingly, this smoking-related increase in methylation was most pronounced in male offspring [40].

Whether these alterations in DNA methylation have a functional relevance, such as by altering gene expression leading to a disease state, are still to be determined. Furthermore, assessment of locus-specific DNA methylation is not ideal as this type of analysis misses many prospective methylation changes that may exist throughout the entire genome, including repetitive regions. Greater focus should be placed on completing unbiased whole methylome analysis to completely understand how the effects of environmental stressors might be epigenetically inherited. Examples of studies that have attempted an unbiased search for epigenetic changes throughout the entire genome are as follows. One group performed a genome-wide DNA methylation analysis (using Illumina 450K array) of whole blood from individuals exposed to famine in utero found a small increase (0.7%−2.7%) in methylation at specific CpG sites associated with genes involved in growth, development, and metabolism (e.g., *FAM150B*, *SLC38A2*, *PPAP2C*, etc.) [41]. Only individuals exposed to famine during the first 10 weeks of gestation showed these specific methylation changes reinforcing that epigenetic programming during the first trimester is particularly sensitive to in utero undernutrition [41]. Once interesting loci are identified using whole-genome approaches, it will be important to confirm the methylation status of these specific loci in different populations around the world, ultimately to validate the findings on a larger scale. A major challenge will be to determine whether small changes in DNA methylation are sufficient to alter gene expression and cause disease.

## 26.6 DEVELOPING A MECHANISTIC UNDERSTANDING OF EPIGENETIC INHERITANCE IN ANIMAL MODELS

Although there is circumstantial and correlative evidence for the occurrence of epigenetic inheritance in human populations, true exploration of the mechanisms behind this phenomenon will only be achieved using experimental models, such as plants and animals. Currently, our understanding of the underlying epigenetic processes of transgenerational inheritance remains in its infancy. However, by meticulously controlling exposures to environmental stressors and genetic background effects, researchers will be able to tease apart the key mechanisms and gain a greater understanding of transgenerational epigenetic inheritance.

It is clear that exposure to environmental stressors (e.g., poor diet, toxicants, etc.) may alter the epigenome and this information may then be transmitted from the individual exposed to their offspring. Stable alterations of the epigenome or "epimutations" can affect gene expression or chromatin stability leading to increased risk of a phenotype or disease in the offspring [42]. There are two key paradigms of epigenetic inheritance: replication and reconstruction [42] (Fig. 26.2).

### 26.6.1 REPLICATING THE ABNORMAL EPIGENOME BETWEEN GENERATIONS

Replicative inheritance requires epigenetic marks, such as DNA itself, to be directly copied between cell divisions and generations [42] (Fig. 26.2A). Generally speaking, the majority of epigenetic marks are removed and replaced to reset the epigenome for transmission to the next generation in a process called epigenetic reprogramming (Box 26.2). These reprogramming events "wipe the epigenetic slate clean" to establish totipotency required to form the next generation. This process may also remove any epigenetic alterations encoded in the epigenome that represents the history of environmental exposures experienced by the parental generation [44]. Ultimately, epigenetic reprogramming limits the inheritance of epimutations between generations. However, recent studies in human primordial germ cells have shown that there are regions of the epigenome that are resistant to DNA demethylation [52]. Reprogramming "escapees" are largely retrotransposable elements, but other genomic regions have been identified, including over 7000 repeat-poor regions (with less than 10% sequence overlap with repeat sequences) and single-copy genes involved in human neural and metabolic disorders [53]. Furthermore, some repeat-poor escapees remain partially methylated in the early human preimplantation embryo suggesting these regions escape both waves of epigenetic reprogramming in the germ cell and during early embryogenesis [53]. Resistant regions are enriched for the histone methylation mark H3K9me3 and binding sites for Krüppel-associated box (KRAB) zinc finger proteins suggesting that they are targeted by proteins that induce heterochromatin formation [53]. Similarly, resistant loci have been identified in mouse primordial germ cells [54] and blastocysts [50]. Epigenetic marks that escape reprogramming have even been identified in germ cells and blastocysts that have not been environmentally stressed [55]. It is hypothesized that if exposure to an environmental stressor drastically alters the pattern of specific epigenetic marks, these changes may be more resistant to normal reprogramming and allow for the inheritance of environmental information. Therefore, the maintenance and inheritance of reprogramming-resistant loci into the next generation is the basis for the replicative model of epigenetic inheritance.

### 26.6.2 RECONSTRUCTING THE ABNORMAL EPIGENOME BETWEEN GENERATIONS

Reconstructive inheritance was postulated by Eva Jablonka as an alternative theoretical concept to replicative inheritance (Fig. 26.2B). The reconstruction model avoids the challenges presented by epigenetic reprogramming [43] by suggesting that epigenetic marks responsible for a phenotype undergo normal reprogramming in the germline and zygote but are then recreated in each successive generation [43]. Traces of the parental epigenetic landscape may be sufficient to reconstitute the parental epigenetic state in the offspring and thus bring about a phenotype [43]. For example, the reconstruction of abnormal epigenetic marks and phenotypes may be driven via altered cellular signaling or metabolism [42]. This mechanism may lead to serial reconstruction of an induced phenotype over multiple generations, such that an abnormal phenotype in the F0 generation programs the same defect in the F1 generation, and so on [56].

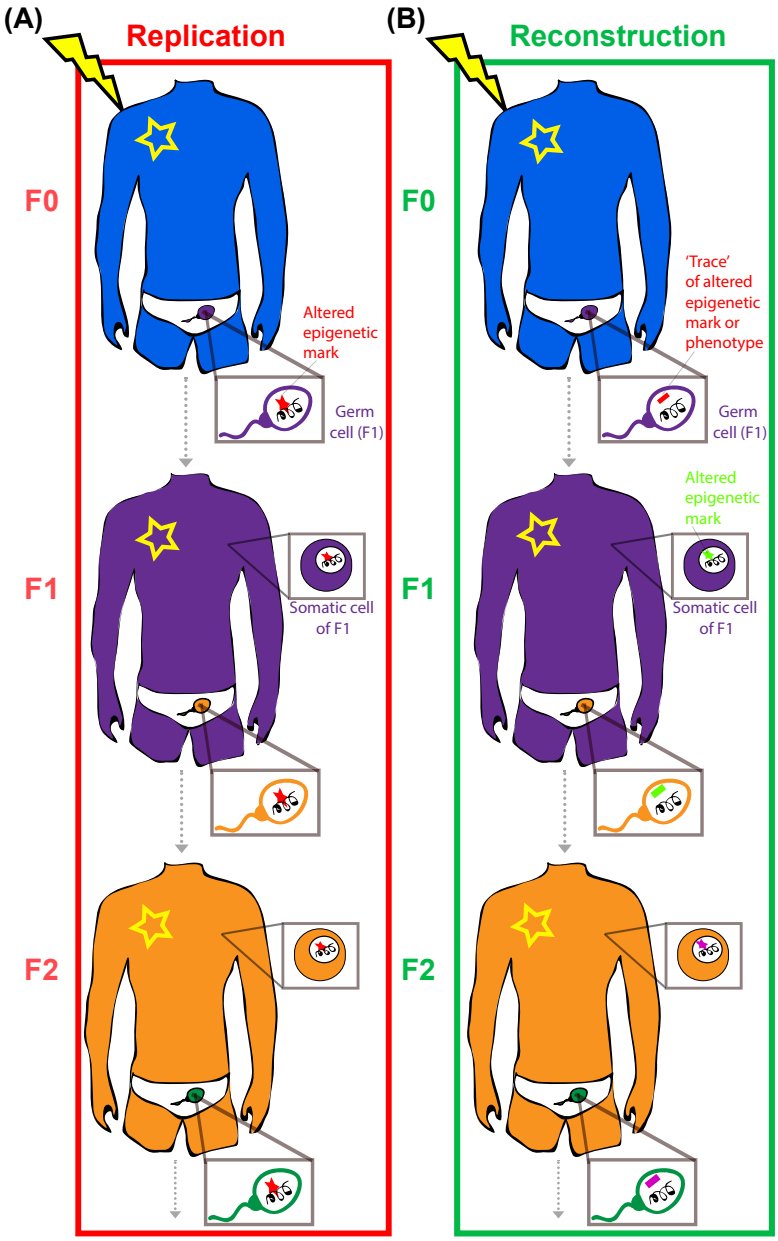



**FIGURE 26.2 Replication Versus Reconstruction Models of Epigenetic Inheritance.**

Two potential models of epigenetic inheritance include the replicative and reconstruction models of inheritance. (A) Replicative inheritance involves the exposure of the F0 generation to an environmental insult (represented by *lightning bolt*) resulting in the direct inheritance of an epimutation(s) (denoted by *red star*) via the germline. For this to happen, epimutations must be stable and escape two waves of epigenetic reprogramming that occur in each generation, allowing them to persist in subsequent generations (i.e., F1, F2, etc.). These epimutations have the potential to lead to a phenotype or disease state (represented by the *yellow star*). (B) Reconstructive inheritance, hypothesized by Eva Jablonka, is an alternative explanation for epigenetic inheritance [43]. This model suggests that epimutations caused by an environmental insult (*lightning bolt*) resulting in a phenotype or disease state (*yellow star*) in the F0 generation undergo epigenetic reprogramming. However, these epimutations are "reconstructed" (*green star*) in the F1 generation based on traces of the parental epigenetic landscape (represented by *colored rectangle* in germ cell). These "traces" are sufficient to direct the reestablishment of the abnormal epigenetic marks in the next generation with the potential to cause a similar phenotype or disease state. A similar event between the F1 and F2 generations results in successive programming of the phenotype over multiple generations.

### BOX 26.2 EPIGENETIC REPROGRAMMING

The dynamics of epigenetic reprogramming is best studied in mouse, though similar processes are thought to occur in humans as well. In humans and rodents, there are two waves of dynamic epigenetic reprogramming that occur between generations. The first wave is in the developing primordial germ cells, and second wave is in the postfertilization zygote [44]. As primordial germ cells migrate to the genital ridge of the developing embryo, DNA methylation is removed throughout the genome [45] and histone marks are remodeled [46]. Prior to birth, DNA methylation patterns are reestablished in the germ cells, including at imprinted genes [45]. In sperm, histones are replaced by protamines to increase DNA compaction within the nucleus [47]. Up to 85% and 99% of histones are replaced in human and mouse sperm, respectively [47]. Some histones are retained in sperm at two distinct sites: promoters of developmentally regulated genes [48] and gene-poor repeat regions [49]. Nucleosomes containing histones are retained genome wide in the mouse oocyte [48]. During the second wave of reprogramming in postfertilization zygotes, DNA methylation is again erased, although imprinted regions and some repetitive loci are protected and therefore, maintained [50]. Shortly after fertilization, protamines are replaced by histones on the DNA of the paternal pronucleus and appropriate histone modifications (e.g., methylation, acetylation, etc.) are acquired [51].

Despite their seemingly contradictory messages, the replication and reconstruction models need not be juxtaposed. In reality, there may be an element of direct inheritance of some epigenetic marks (replication) and reconstruction of additional marks linked to an ancestral phenotype.

## 26.7 MECHANISTIC CANDIDATES AND MACHINERY OF TRANSGENERATIONAL EPIGENETIC INHERITANCE

Mediators of epigenetic inheritance must be responsive to the environment. Key mechanistic candidates that are postulated to pass heritable information between generations via the germline include DNA methylation, histone modifications, and ncRNA [1]. Each of these mediators of epigenetic inheritance will be considered in key models in which they have been investigated.

### 26.7.1 DNA METHYLATION

In humans and other mammals, methylated cytosine in form of 5-methylcytosine (5mC) is the predominant form of DNA methylation. In other organisms, such as *Caenorhabditis elegans* and *Drosophila melanogaster*, 5mC is either absent or detected at extremely low levels. However, methylation of adenine to generate N6-methyladenine (6mA) was recently identified in these species [57]. 6mA is also present at very low levels in human cells [58]. DNA methylation in gene control regions is generally associated with gene repression, although its exact role in gene regulation is likely locus-dependent [59]. Beyond regulation of gene expression, a key role of DNA methylation is to silence repetitive DNA [60], a feature that becomes important for inheritance. However, we are only beginning to understand how DNA methylation and transcriptomic information overlap. As a result, in the context of epigenetic inheritance, exploring how DNA methylation status associates with gene expression and thus phenotype remains challenging. For example, intergenerational epigenetic inheritance of a metabolic phenotype caused by in utero undernutrition in mice results in DNA methylation patterns that are largely unchanged despite gene misexpression or altered patterns of DNA methylation associated with normal gene expression [55]. DNA methylation has many properties that

make it an attractive mechanistic candidate for epigenetic inheritance, including that (1) it can be environmentally modulated; (2) machinery exists to replicate methylation patterns onto newly synthesized DNA and thus it is mitotically heritable; and (3) loci resistant to epigenetic reprogramming in the zygote and germline have been identified. Examples of DNA methylation in models of epigenetic inheritance are mechanistically explored below (see Section 26.8).

### 26.7.2 HISTONE MODIFICATIONS

Chromatin packaging and nucleosome positioning on DNA are regulated, in part, by chemical modifications (e.g., methylation, acetylation, etc.) that are applied to histone proteins around which DNA is packaged. These modifications serve to regulate gene expression through the recruitment of protein complexes and by controlling the accessibility of the DNA to the transcription machinery [61]. Histone modifications can be environmentally modulated and propagated onto newly assembled chromatin by complex machinery [62]. However, most histones are removed during spermatogenesis and replaced with protamines [47]. Protamine modifications are similar to histone modifications and are present in sperm [63], though their functional importance remains uncertain, particularly during epigenetic inheritance. The retention of some histone modifications in the germline and zygote during epigenetic reprogramming makes them another possible vector of inheritance across generations [47]. The role of histones in epigenetic inheritance has largely been studied in *C. elegans*, though their role in a mammalian context has recently been explored (see Section 26.15).

### 26.7.3 NONCODING RNA

Germ cells have extensive RNA profiles making RNA another mechanistic candidate for epigenetic inheritance. These germ cell RNAs have diverse roles from direct regulation of gene expression to localizing other epigenetic pathways, such as piwi-interacting RNA (piRNA)-mediated transposon silencing [64]. Although once thought to lack undegraded RNA, sperm contains both long and short ncRNAs, including microRNAs (miRNAs), piRNAs, and transfer RNAs (tRNAs) (reviewed in Ref. [65]). The RNA profile of mature oocytes is equally, if not more diverse. Significantly more RNA is found in oocytes as they contain all the RNAs required for the first few days of embryonic development before zygotic genome activation [66]. The complexity of germ cell RNA profiles is increased by the chemical modification of RNAs, such as methylation [67], which provides an additional layer of epigenetic information that might be transmitted to the next generation. How epigenetic messages transmitted by RNA are perpetuated over multiple cell divisions and multiple generations currently remains unclear.

However, a unique aspect to RNA-mediated inheritance is that ncRNAs might be taken up by germ cells from the surrounding somatic tissues [68,69]. For example, maturing sperm might receive small ncRNAs (e.g., tRNA fragments) from small vesicles called exosomes released from the epididymis [68]. Thus, the theoretically impassable wall between somatic cells and the germline (the so-called Weismann barrier) may be circumvented and penetrated by RNAs. It is currently unclear how environmental stressors experienced by adult somatic cells influence the establishment or maintenance of DNA methylation and histone modifications in gametes. Evidence that somatic-to-germline transmission of epigenetic information by RNA is possible has allowed researchers to overcome this obstacle in our understanding of epigenetic inheritance (see Section 26.14).

### 26.7.4 EPIGENETIC "READERS" AND "WRITERS"

For epigenetic information to be created, understood, and enacted on, epigenetic regulatory machinery within the cell, the so-called epigenetic "readers" and "writers," is required [42]. To be heritable in normal and disease scenarios, DNA methylation must be retained though mitosis and meiosis [1,2]. A family of DNA methyltransferase (DNMT) enzymes enables establishment and maintenance of DNA methylation. DNMT1 ensures that methylation is recapitulated on the daughter strand of DNA after replication has occurred [70]. DNMT3A and DNMT3B establish de novo DNA methylation patterns in germ cells [70]. Alternatively, the ten-eleven translocases (TETs) remove methylation via a 5-hydroxymethylcytosine (5hmC) intermediate [71]. It is now becoming clear that 5hmC may have functional roles independent of 5mC demethylation [72]. Histone modifications must also be recreated on nascent histones on the daughter strand. An army of histone-modifying enzymes, including methyltransferases and acetyltransferases, work in large complexes with other proteins to add and remove histone marks [62]. How histone modifications are replicated still remains somewhat ambiguous [65]. Elaborate machinery is also associated with the activity of ncRNAs, such as the piRNA complexes that involve the argonaute superfamily cofactors [73].

Epigenetic machinery may play an important role in transgenerational epigenetic inheritance, particularly in the context of the reconstruction model. Individuals with deleterious mutations in this machinery may be more susceptible to epigenetic inheritance of phenotypes [74]. Evidence for this comes from the plant model *Arabidopsis thaliana* containing a mutation in the DNMT *MET1*. These plants are hypomethylated at a repetitive DNA element near the transcription start site of the FLOWERING WAGENINGEN A (*FWA*) locus [75], which causes a late flowering phenotype. Interestingly, this phenotype and *FWA* hypomethylation is inherited over multiple generations of wild-type offspring independent of the original *MET1* mutation [76] suggesting that reducing the expression of DNA methylation machinery can lead to epigenetic inheritance.

Similarly, histone methylation machinery is also vital for the process of epigenetic inheritance in *C. elegans* and to some extent in mice. Paternal deficiencies in the H3K4me3 chromatin modifiers ASH2, WDR5, or SET2 in *C. elegans* lead to increased longevity for up to three wild-type generations [77]. Haploinsufficiency for the H3K9 methyltransferase *Setdb1* gene in male mice can influence the coat color of his offspring when mated to a female with an agouti viable yellow ($A^{vy}$) allele that alters coat color (see Section 26.8) [74]. Overall, DNA methylation is altered at a specific class of retrotransposons of the *Setdb1*-deficient sperm. The influence on offspring coat color is an *in trans* effect and supports the suggestion that epigenetic inheritance may be more likely in individuals with altered levels of epigenetic modifiers [74].

## 26.8 THE CLASSIC MAMMALIAN EXAMPLE: AGOUTI VIABLE YELLOW

The classic mammalian model of transgenerational epigenetic inheritance, from which much of our mechanistic understanding has been derived, is the $A^{vy}$ mouse model [78] (Fig. 26.3A). The *agouti* gene (*a*) is involved in controlling mouse coat color. The $A^{vy}$ allele contains an intracisternal A-particle (IAP) element upstream of the *agouti* gene that regulates *agouti* gene expression in a DNA methylation-dependent manner [78,81]. The $A^{vy}$ allele shows variable DNA methylation at the IAP element leading to variable *agouti* gene expression and coat color in isogenic $A^{vy}/a$ mice [78]. DNA methylation silences ectopic *agouti* mRNA expression to generate brown (pseudoagouti) colored mice,

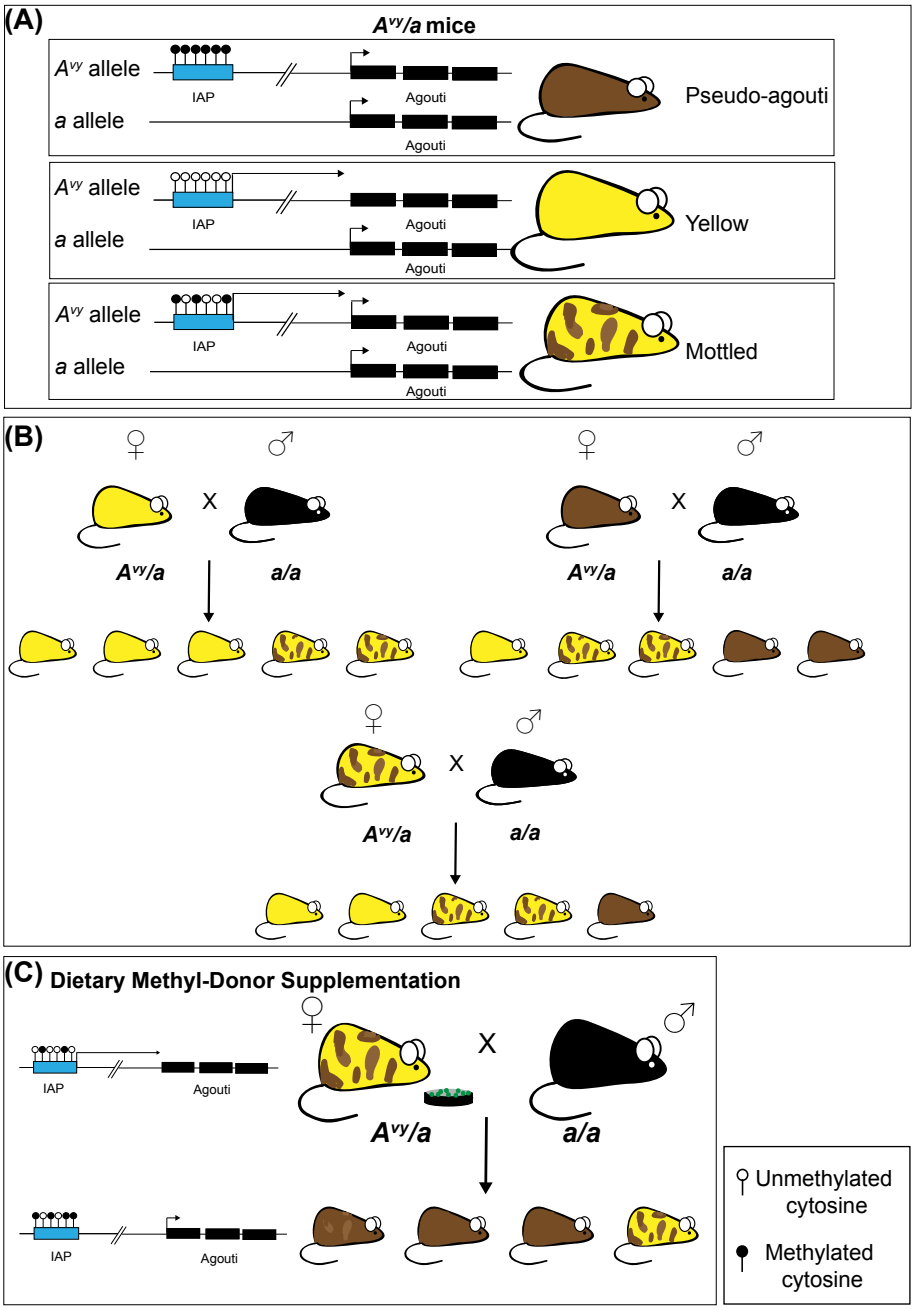



**FIGURE 26.3 Agouti Viable Yellow ($A^{vy}$): A Classic Example of Epigenetic Inheritance.**

(A) An intracisternal A-particle (IAP) upstream of the $A^{vy}$ allele regulates *agouti* gene expression in a DNA methylation-dependent manner [78]. In $A^{vy}/a$ heterozygous mice, different degrees of DNA methylation at the IAP element result in different coat colors. When fully methylated (upper panel), *agouti* expression is driven from the nascent promoter resulting in mice with a brown (pseudoagouti) coat color. When the IAP element is

whereas hypomethylation at the IAP element drives ectopic *agouti* expression ($A^{vy}$) leading to mice with a yellow coat color [78]. These mice also have obesity and a diabetic-like phenotype, owing to disruption of their satiety signaling pathways [81,82]. Remarkably, the methylation status of the $A^{vy}$ allele, and therefore, the coat and metabolic phenotypes, can be inherited over multiple generations via the maternal line [78]. For example, offspring of yellow mice have an increased propensity to have yellow offspring [78] (Fig. 26.3B). The methylation status of the agouti gene can also be influenced by environmental factors, such as nutrition [79,83]. Feeding $A^{vy}$ mothers a methyl-rich diet reduces the number of yellow offspring in her litters [79,83] (Fig. 26.3C). Whether DNA methylation pattern at the IAP element normalizes in response to increased methyl groups remains unclear [79,83]. The $A^{vy}$ locus undergoes epigenetic reprogramming during early embryogenesis, including the removal of IAP methylation following maternal transmission of the allele [84]. This suggests a mechanism of inheritance that might be independent of DNA methylation and reconstruction-based. For example, sperm ncRNA profiles, including tRNA fragments, are abnormal in the sperm of F1 males derived from mating an $A^{vy}$ obese male to wild-type lean female [82]. This obesity phenotype persists in the F1 and F2 generations, although it is attenuated by the F3 generation [82]. Although tRNA fragments have not yet been assessed in the F2 or F3 generations, RNA-based mechanisms may contribute to inheritance and reconstruction of abnormal DNA methylation profiles at the IAP element.

Recently, a large number of studies have tried to probe further the mechanisms of transgenerational inheritance of phenotypes in the $A^{vy}$ model (reviewed in Ref. [1]). These studies are often limited for the following reasons: (1) only parent-to-offspring transmission is considered meaning that true transgenerational inheritance is not studied; (2) a locus-specific approach rather than an unbiased genome-wide approach is taken, which limits the findings; and (3) the germline itself is not assessed. However, other models can be assessed to emphasize common paradigms and unanswered questions in the field. The following studies illustrate key aspects of epigenetic inheritance of disease, even though some studies are not in a disease context.

## 26.9 THE BANDWIDTH OF INHERITANCE

It is debated in the field of epigenetic inheritance whether specific information regarding an adverse environment is transmitted to the next generation or whether a more general signal of suboptimal

unmethylated (center panel), an ectopic promoter in the IAP element drives ectopic agouti expression leading to a yellow coat color and a metabolic phenotype. Partial methylation at the IAP element (lower panel) leads to a mottled coat color. (B) The methylation status of the agouti allele, and therefore the coat color and metabolic phenotypes, is heritable over multiple generations via the maternal line [78]. When mated to an *a/a* male mouse, a yellow $A^{vy}/a$ female mouse has an increased propensity for yellow offspring (top left) compared with a brown $A^{vy}/a$ female (top right) due to epigenetic inheritance at the $A^{vy}$ allele methylation patterns [78]. Mottled $A^{vy}/a$ females mated to *a/a* males (bottom) results in an intermediate propensity for yellow offspring compared with the other matings. In the schematic, only $A^{vy}/a$ offspring are depicted and *a/a* progeny have been excluded for simplification. (C) The methylation status of the $A^{vy}$ allele is influenced by dietary methyl donor supplementation. A methyl-rich diet fed to a mottled colored $A^{vy}/a$ female reduces the number of yellow offspring in her litters with respect to unsupplemented controls (in bottom of panel B) [79] due to increased methylation of the IAP element. Images based on [80] and [78].

conditions is inherited. This is described in terms of the extent or "bandwidth" of information that can be communicated [65]. Evidence supporting the transfer of specific information comes from a multigenerational study assessing olfactory fear conditioning in mice [85]. In this study, F0 males were conditioned to associate a specific odor with a foot shock [85]. The F1 and F2 generations derived from these conditioned males have increased behavioral sensitivity to the same odor and not other odors even though they had not been conditioned themselves [85]. Importantly, social transmission of odor sensitivity was excluded by undergoing techniques, such as in vitro fertilization and cross-fostering [85]. This odor sensitivity correlated with DNA hypomethylation of an odorant receptor gene, *Olfr151*, in sperm of the F0 and F1 generations [85]. Alternatively, recent evidence for inheritance of a general message comes from exposing mice to nicotine [86]. This exposure primes their male F1 offspring for increased survival when exposed to toxic levels of nicotine or cocaine suggesting a general enhanced xenobiotic resistance is inherited [86].

The DNA methylation status at any given cytosine residue is binary meaning that it can be methylated or unmethylated. However, a moderate change in CpG methylation ($<10\%$) in a given subpopulation of cells is often associated with complex disease phenotypes, particularly in the context of transgenerational epigenetic inheritance [87]. This is because even moderate methylation changes can profoundly alter gene expression patterns leading disease phenotypes. For example, exposure of pregnant rats to the toxic endocrine disruptor p,p′-dichlorodiphenyldichloroethylene leads to a transgenerational reduction in sperm quality and infertility up to the F3 generation males [88]. These phenotypes are associated with a 10%−50% reduction in methylation within DMR2 of the *Igf2/H19* locus in sperm collected from the F1−F3 generations and significant changes in *Igf2* and *H19* RNA expression [88]. Whether this dysregulation of gene expression leads to infertility is unclear. In general, the degree of DNA methylation change that can lead to dramatic alterations in gene expression and severe phenotypes remains poorly explored. Understanding the biological significance of small DNA methylation changes that result from environmental stressors is a challenge that should not be ignored.

## 26.10 THE IMPORTANCE OF THE REPETITIVE GENOME

A large part of most genomes are made up of repetitive elements, such as transposons and retrovirus-derived sequences that need to be silenced to prevent their transposition into other regions of the genome resulting in mutation [89]. Epigenetic mechanisms are vital to their suppression. Epigenetic marks at repetitive elements are more likely to be resistant to reprogramming than unique loci, thus providing scope of heritability [50,53,54]. The IAP element in the $A^{vy}$ mouse model has been classified as a metastable epiallele, and its epigenetic regulation highlights the importance of repetitive elements in the process of transgenerational epigenetic inheritance [78]. Numerous other repetitive loci have since been implicated in epigenetic inheritance.

The importance of the repetitive genome in transgenerational epigenetic inheritance has been extensively demonstrated in *C. elegans*. In this context, histone modifications and their regulatory enzymes are vital for the transmission of epigenetic information between generations. For example, *C. elegans* were genetically modified with a multicopy array of a transgene containing a fluorescent reporter gene under the control of a heat shock promoter to mimic endogenous repetitive loci [90]. The worms showed elevated reporter gene expression following exposure to high temperatures [90].

Remarkably, this elevated expression is heritable for up to 14 generations even in the absence of heat [90]. In contrast, when only a single copy of the transgene was present, heritability of elevated expression occurred for only seven generations [90]. It was proposed that the activity of the H3K9me3 methyltransferase, SET25, is reduced at high temperatures [90]. Therefore, heat treatment may allow transgene derepression by altering histone methylation. In fact, it takes many generations of worms grown in normal temperatures to reestablish proper histone methylation [90]. While providing an interesting insight into the mechanism, it will be interesting to resolve the extent to which histone epigenetic marks are involved in the transgenerational inheritance of phenotype in mammals.

## 26.11 TRACKING PHENOTYPES AND EPIMUTATIONS OVER MULTIPLE GENERATIONS

The aim of the field is to use animal and plant models to trace epigenetic changes and associated phenotypes between many generations to mechanistically understand epigenetic inheritance. Long generation times in mammalian models often limit the number of generations assessed, frequently only up to the F2 generation. This constrains our ability to rule out physiological and genetic effects and to determine how many generations a phenotype or epigenetic change might persist. Studies have reported that phenotypes caused by an environmental stressor diminish over time with a complete loss after several generations. For example, the metabolic phenotype observed in offspring derived from obese $A^{vy}$ males and lean wild-type females is attenuated by the F3 generation [82]. Because epigenetic inheritance potentially provides adaptive benefits to the environment, phenotypic attenuation seems logical. That is, if an adverse environmental stressor is not reexperienced by subsequent generations, a gradual dilution of the phenotype would be expected as an adaptive mechanism. Instead, the current environment should result in new epigenetic information being conveyed to the offspring.

Alternatively, another mouse study whereby cellular methylation levels are metabolically reduced displays transgenerational epigenetic inheritance of specific developmental phenotypes for at least up to the F4 generation [91]. In this model, a gene encoding an enzyme called methionine synthase reductase (*Mtrr*) is mutated leading to disruption of folate and methionine metabolisms [91]. Folate is vital for the transmission of one-carbon methyl groups required for cellular methylation of substrates, including DNA, histones, and RNA. *Mtrr* mutants display hyperhomocysteinemia, widespread epigenetic instability, including DNA hypo- and hypermethylation, and congenital malformations, such as neural tube, heart, and placental defects [91,92]. Remarkably, these congenital malformations persist transgenerationally in wild-type offspring at least up to the F4 generation when either maternal grandparent was a carrier for the *Mtrr* mutant allele. This effect was demonstrated to be independent of a maternal uterine effect [91]. Although the inherited epigenetic factor is not yet known in this model, widespread dysregulation of DNA methylation associated with gene misexpression specifically in the wild-type placentas of the F2 generation was apparent [91]. Although epigenetic instability beyond this generation is yet to be determined, the *Mtrr* model clearly demonstrates that reducing the availability of methyl groups causes transgenerational inheritance of phenotypes. This mouse line will be invaluable for studying the epigenetic mechanism via the germline.

To provide mechanistic evidence for the replicative model of transgenerational epigenetic inheritance, specific epimutations must be tracked over multiple generations. Much effort has gone into tracing epimutations in dietary models of multigenerational inheritance. A number of animal models

whereby diet is manipulated (i.e., high fat, low protein, "western diet," or undernutrition) display metabolic disease in the offspring. However, only a few studies have been successful in identifying epigenetic changes identified in the F0 generation that persist to the F1 generation and beyond. For example, a prediabetic mouse model displayed an impaired fasting glucose and insulin resistance phenotype that is passed on through the paternal lineage from the F0 to F2 generation [93]. An unbiased analysis of DNA methylation revealed DMRs in the sperm of the F0 males that were also identified in DNA of pancreatic islet cells of F2 males [93]. However, in an intergenerational model of in utero undernutrition, which leads to a robust metabolic phenotype and reduced birth weight in the F1 and F2 generations, altered DNA methylation patterns could not be tracked across generations [55,94]. Numerous DMRs were identified in the sperm from F1 males that were undernourished in utero versus control animals [55]. However, these epimutations did not persist in the somatic tissue of the F2 generation even though gene expression at associated loci was altered and the metabolic phenotype persisted [55]. This suggests that additional epigenetic mechanisms may be at work (e.g., histone marks, RNA regulation of transcription).

## 26.12 ALLOWING FOR PHENOTYPIC AND EPIGENETIC VARIABILITY

Under normal circumstances, considerable epigenetic variation exists between individuals [95]. Therefore, this must be considered when assessing epigenetic changes in models of epigenetic inheritance, particularly when pooling samples together for analysis. Remarkably, one study showed epigenetic variation as a result of diet manipulation is less than natural epigenetic variation between individuals [95]. Alternatively, the wide spectrum of phenotypes observed in some models, such as the *Mtrr* model [91], suggests a variable interindividual response to environmental stressors. It is currently unclear whether environmental insults affect the epigenome stochastically in each individual and/or within the germ cell population or at hotspots within the epigenome so that a population is similarly affected. Single-cell methylome and transcriptome technologies may permit us to better understand this heterogeneity [96] and epigenome sensitivity.

## 26.13 CONSIDERING GENETIC EFFECTS

We must take into account that epigenetic instability may promote genetic instability when examining the mechanisms of epigenetic inheritance. DNA methylation is a known mutagen because spontaneous deamination of methylcytosine forms thymine, and thus generates a C → T transition mutation [97,98]. Loss of DNA methylation may also lead to activation of transposable elements that could transpose into new genomic locations to generate genetic mutations [99]. Alternatively, changes in the number of tandem repeats can affect the level of DNA methylation detected. For example, there is a higher level of variation in DNA methylation at ribosomal DNA (rDNA) repeats that correspond to rDNA copy number rather than epigenetic changes induced by diet [95]. Therefore, in some cases, genetic variation may be the underlying cause of epigenetic variation.

Additionally, the environmental stressors that affect the epigenome may be mutagenic themselves. The literature is awash with examples of how toxicant exposure, from endocrine disruptors to jet fuel, can influence our epigenomes and the health of our offspring [100]. For example, in utero exposure of rats to the endocrine disruptor vinclozolin at a time critical for sex determination results in a range of

adult-onset diseases, such as infertility and kidney defects in the unexposed offspring for up to four generations [101]. These phenotypes were associated with widespread changes in DNA methylation in sperm of the F3 generation [102]. However, others have been unable to recapitulate this result [103], perhaps due to differences in methodology. Importantly, vinclozolin exposure results in an increase of repeat element copy number variations [104], highlighting the importance of examining the genome and epigenomes in parallel.

Genetic background effects dictated by the genetic strain of the animal might also influence whether phenotypes and/or epigenetic changes associated with specific environmental stressors appear [102]. This may be due to differences in interactions between the epigenome and underlying genome. Only a few transgenerational epigenetic inheritance studies have taken into account potential genetic background effects. For example, there is a lack of evidence for a genetic effect in $A^{vy}$ mice. This was demonstrated by sequencing the genomes of two $A^{vy}$ littermates, one pseudoagouti and one yellow, to assess genetic differences [105]. The germline mutation rates were no different than expected, and genetic changes were deemed unlikely to cause the coat color phenotype [105].

Other studies have shown that telomeres, which are stringently epigenetically regulated genomic regions [106], may provide a novel mechanistic interface between genome and epigenome during epigenetic inheritance. Researchers showed that a low-protein diet in mice caused intergenerational inheritance of premature reproductive tract aging in F2 females associated with reduced telomere length [107]. Altogether, how epigenetic instability affects genetic stability needs to be explored further in this and other models of epigenetic inheritance.

## 26.14 RNA-MEDIATED INHERITANCE OF PHENOTYPE

In recent studies, the mechanistic focus has shifted somewhat from the inheritance of DNA methylation and chromatin modifications to the inheritance of ncRNAs, such as miRNAs and tRNAs. Transgenerational inheritance mediated by RNA is demonstrated in *C. elegans*, *D. melanogaster*, and plants (reviewed by Ref. [108]). These organisms have RNA polymerases that replicate an RNA signal, which might allow for more robust RNA-mediated inheritance compared with mammals that lack an RNA replicative mechanism [65]. In *C. elegans*, RNA obtained from the environment can mediate heritable gene silencing. This phenomenon known as RNA interference (RNAi) is extensively used to experimentally generate gene knockdowns [109]. Although the mechanism is complex, it is thought that piRNAs, which are typically involved in transposon silencing in the germline, and RNAi converge on a common pathway, involving RNA and chromatin regulatory complexes [73]. It was demonstrated that a mechanism similar to those involved in exogenous RNAi plays a role in inheritance at endogenous loci. For example, starvation-induced developmental arrest in *C. elegans* causes the transgenerational inheritance of small RNAs and an increased life span over multiple generations [110].

In mice, our understanding of the importance of ncRNAs in epigenetic inheritance has come from injections of RNA into fertilized oocytes [111−113]. This technique has demonstrated causality between RNA molecules and phenotypes in the offspring. For instance, in a mouse model of traumatic stress, involving maternal separation in early postnatal life of the F1 generation, severe behavioral phenotypes were observed up to the F2 generation and were associated with altered miRNA profiles in the F1 sperm [111]. Microinjection of sperm miRNAs from stressed males into fertilized oocytes mimicked the behavioral phenotypes [111]. Furthermore, changes in DNA methylation at the

glucocorticoid receptor gene in sperm from F1 males and brains of F2 offspring correlated with gene misexpression [114]. Remarkably, environmental enhancement of the F1 stressed males when they were adults rescued the behavioral phenotype and DNA methylation patterns further demonstrating the plasticity of the epigenome [114].

Apart from miRNAs, it was recently demonstrated that diet manipulation alters fragmented tRNA profiles in mouse sperm leading to metabolic disease in the next generation. Fragmented tRNAs are 28−40 nucleotides in length and derived mainly from the 5′ end of tRNAs [112]. Poor diet in male mice can alter metabolic gene expression in early development of the F1 generation [112] and lead to adult metabolic disease phenotypes [113]. These effects have been linked to altered tRNA fragment profiles in the sperm [112,113]. When sperm tRNA fragments isolated from high-fat diet−fed mice were microinjected into fertilized eggs, it recapitulated metabolic disease in adulthood [112,113]. Additionally, oocyte microinjection of sperm tRNA fragments from mice fed a low-protein diet specifically repressed early embryonic genes associated with murine endogenous retrovirus-like (MERVL) elements in their control regions [112]. These genes are hypothesized to influence feto-placental development, the dysregulation of which might lead to adult-onset metabolic disease. Interestingly, microinjection of synthetic tRNA fragments was unable to recapitulate the phenotype [113], suggesting that RNA modifications (e.g., methylation) may be important in epigenetic inheritance of disease.

It is crucial to consider that RNA microinjection into oocytes adds to the existing pool of RNA, which leads to a supraphysiological RNA load in the blastocyst. The consequences of this procedure are not fully understood and may influence the phenotypes observed. Although RNA injections into fertilized oocytes suggest a causative role for RNA in epigenetic inheritance, it will be necessary to determine whether restoration of normal sperm RNA profiles can rescue the phenotype in question. An approach, such as blocking candidate RNA transmission by an RNAi-like mechanism, albeit challenging, might achieve this result.

RNA modifications might also be important in paramutation effects [115]. A paramutation involves a mutant allele (i.e., the paramutagenic allele) that alters the expression of the homologous wild-type allele (i.e., the paramutant allele) in trans via an RNA-mediated mechanism [116,117]. Remarkably, a meiotically stable paramutant allele can be transgenerationally inherited in absence of a paramutagenic allele [116,117]. For example, multigenerational transmission of a mouse paramutant "white-tail-tip" phenotype results after mutational transgene insertions into the *Kit* locus, such as $Kit^{lacZ}$ and $Kit^{copGFP}$ [116,118,119]. It was determined that the RNA methyltransferase DNMT2 is required for the transmission of the *kit* paramutation [120] suggesting that RNA modifications are broadly implicated in RNA-mediated inheritance.

RNA-mediated mechanisms of inheritance in rodents have not been studied in a true transgenerational context. Germline ncRNAs can clearly influence gene expression or disease phenotype in the next generation [111−113,121]. Yet, the mechanistic details of how this occurs remain unclear. Indeed, microinjection of miRNAs into oocytes can initiate phenotype inheritance over multiple generations, but an increase in expression of the same miRNA is not observed beyond the injected generation [118]. This suggests that RNA interacts with other epigenetic mechanisms to promote phenotype inheritance. It is possible that RNA-mediated inheritance fits into the reconstruction theory of transgenerational epigenetic inheritance because the reestablishment of altered DNA methylation or histone marks in each generation might be RNA-directed.

## 26.15 INTERACTIONS BETWEEN EPIGENETIC MECHANISMS

Within the individual cells, DNA methylation, chromatin modifications, and ncRNAs are interconnected and interdependent, and in the context of epigenetic inheritance, they likely act in a coordinated and collaborative manner to cause disease phenotypes. For example, DNA methylation or histone modification patterns might regulate small ncRNA expression. This, in turn, may direct DNA/histone methylation patterns to ultimately establish an interactive feedback system [42]. Collaborative, multifaceted, and often costly approaches to assess all epigenetic pathways in a single model are required to fully explore epigenetic mechanisms of inheritance.

Some studies have already demonstrated interdependence between epigenetic mechanisms. For instance, an initiating mutation in the *C. elegans spr-5* gene that encodes an H3K4 histone demethylase causes a transgenerational decline in fertility, even when all subsequent generations are wild type for *spr-5* [122,123]. This transgenerational effect correlates with an accumulation of H3K4me2 and an increase in 6mA levels in each generation [122,123]. Loss of a putative 6mA DNMT leads to a partial rescue of the *spr-5* infertility phenotype [123] demonstrating a mechanistic interaction between histone modifications and DNA methylation. In mice, ectopic expression of an *spr-5* ortholog, *Kdm1a*, during spermatogenesis leads to developmental abnormalities for up to three wild-type generations [124]. However, the sperm of the F1 generation, whose offspring displayed developmental abnormalities, had normal H3K4me2 and DNA methylation profiles [124]. This finding suggests that while the disruption of the histone methylation machinery can instigate the transgenerational inheritance of developmental abnormalities, additional epigenetic mechanisms are at work [124].

## 26.16 ADAPTIVE BENEFITS OF EPIGENETIC INHERITANCE

Epigenetic inheritance might be viewed as memory of an adverse environment so that subsequent generations are better adapted to cope with similar hostile conditions [125]. Disease phenotypes occur when the environment experienced by an individual does not match the one they are "primed" to experience. For example, an undernourished ancestor will prime offspring for famine conditions. However, if the offspring have abundant access to nutrition instead, the risk for metabolic disease and obesity is increased [126]. As a result, many studies have used a challenge paradigm to demonstrate the potential adaptive benefit of an ancestral stressor. To do this, progeny born following an ancestral exposure to a stressor are then challenged with that stressor to determine their ability to cope. The hypothesis is that inherited epigenetic factors will improve their ability to manage this challenge. For example, prenatal exposure to nicotine primes the offspring for increased survival when exposed postnatally to toxic levels of nicotine [86]. The mechanisms behind this remain to be explored, and this adaptive response has not been investigated over further generations.

Prolonged exposure to a stressor might be required for transgenerational inheritance effects to be obvious. As discussed previously (see Section 26.10), heat exposure of *C. elegans* containing a multicopy fluorescent reporter array under the control of a heat shock promoter leads to elevated transgene expression over many generations in the absence of heat [90]. Heat exposure for a single generation led to elevated transgene expression for only 7 generations, whereas exposure to high temperatures for five successive generations led to elevated transgene expression for 14 generations [90]. This type of transgenerational effect could be viewed as an adaptive response and might ensure

that inheritance of a phenotype only occurs when the environment experienced by the offspring is almost certain to be the same in subsequent generations. Ultimately, this would make it less likely that a potential disease-causing mismatch between the epigenotype and the environment is created.

## 26.17 CAN OUR MECHANISTIC UNDERSTANDING OF EPIGENETIC INHERITANCE BE EXTRAPOLATED TO HUMAN POPULATIONS?

Transgenerational epigenetic inheritance was first recognized and understood mechanistically in plants [1]. Although out of the disease context, plant models still provide a solid basis from which further studies have been based. As plants are sessile and unable to change their immediate surroundings or that of their offspring, it is easy to understand why providing a memory of environmental stress is beneficial [127]. For example, multigenerational drought exposure leads to DNA methylation changes at stress-responsive genes in rice resulting in increased drought tolerance over multiple generations [128]. Although some epigenetic mechanisms are conserved, the fundamental difference between plants and animals is that the plant germline is derived from somatic cells unlike in humans and animals [129]. Therefore, strong caveats are required when trying to apply what we have learnt in plants to human populations [129].

In recent years, focus has been placed on generating animal models of human diseases beyond those induced by dietary manipulations. These models may be particularly useful for investigating diseases where a so-called "missing heritability problem" has been identified (e.g., autism, asthma) [130]. Genome-wide association studies have identified a range of genetic loci that may increase susceptibility to these conditions but thus far, they do not account for the complete heritable risk of an individual [130]. This further implicates an epigenetic component. Two such examples include mouse models of autism and asthma. Valproic acid is used to generate an autism-like phenotype in mice and leads to the transgenerational inheritance of autism-like behaviors up to the F3 generation [131]. Similarly, maternal exposure to air pollutants showed an increased likelihood of allergic asthma in the F2 and F3 generation that may be related to DNA methylation changes in dendritic cells [132]. Epigenetic mechanisms have yet to be fully explored in these disease models but to do so might provide interesting insight into disease origin and inheritance.

Given the similarity in epigenetic regulation and the responses to environmental exposures between animal models and human populations, it is clear that the mechanistic understanding gleaned from animal models will hold some relevance to human populations [129]. Nonmammalian models, including *D. melanogaster* and *C. elegans*, provide us with highly tractable experimental systems to test fundamental principles of epigenetic inheritance over many generations. Despite sharing epigenetic similarities with mammalian systems, such as homologous histone-modifying machinery, there are major distinctions that need to be taken into account when extrapolating data to human populations. These differences include a lack of cytosine methylation and the presence of RNA polymerases in these invertebrate models. Fewer fundamental epigenomic differences exist between mammalian models and humans, and this might allow us to more easily apply mechanistic insights identified in rodents to epigenetic inheritance in human populations. Rather than focusing on the inheritance of shared disease phenotypes as discussed above (this section), identifying epigenetically labile sites in humans such as the mouse $A^{vy}$ allele will likely be more helpful to ascertain common mechanisms involved in epigenetic inheritance. Humans lack IAP elements, which are fundamental to

our best-understood mammalian model of transgenerational epigenetic inheritance. However, humans do have equivalent repetitive genomic sequences, such as LINE1 elements, which are regulated by similar epigenetic mechanisms and should be explored further in the context of epigenetic inheritance. It is difficult to separate genetic effects from epigenetic effects in humans because of genetic diversity within a single population. Therefore, animal models, which are usually inbred and isogenic, are experimentally valuable because potential confounding genetic effects can be avoided. Additionally, more histones are retained in human sperm than in rodent sperm (15% retention in human vs. 1% retention in mouse [47]) potentially offering more opportunities for histone-mediated inheritance in human populations than in rodent models. Although there is little overlap between the loci that escape reprogramming in the mouse versus in humans [54], this may not be important if replicative epigenetic inheritance paradigms are correct.

Using models that represent human disease, we will hopefully be able to mechanistically probe how epigenetic factors contribute to the disease inheritance. For these models to be useful, care should be taken to not only accurately mimic the disease but also the degree of environmental exposure to make it similar to what would be experienced by human populations (e.g., amount of toxicant exposure, type of diet, etc.).

## 26.18 CONCLUSIONS

Although our understanding of epigenetic inheritance of disease has come a long way in recent years, there are still many unanswered mechanistic questions. To fully understand the implications of how the environment influences the epigenome, unbiased next-generation sequencing techniques should be performed to correlate DNA and histone marks with gene expression changes in germ cells and somatic cells over multiple generations. These findings will allow us to determine whether even minor epigenetic changes can increase disease risk and their ability to be inherited. This will be a costly undertaking in each model of transgenerational inheritance and will likely require collaboration across research groups. The importance of epigenetic regulation of the repetitive regions of the genome has become apparent, disruption of which may affect widespread chromatin organization or genetic stability. As well, assessing the interplay between different epigenetic mechanisms and epigenetic remodeling machinery will be key. Although there may be commonalities in the mechanism of transgenerational epigenetic inheritance between key models, there is also a growing understanding that there may also be paradigm-specific effects.

As epigenome profiling tools develop further to allow for better mapping of the repetitive genome, our ability to explore these mechanisms will become easier. New technologies allowing for the manipulation of the epigenome by altering locus-specific DNA methylation or individual histone modifications using CRISPR-cas9−based acetyltransferases [133] and TALE-TET1 fusions [134] will enable us to explore whether the epigenetic changes we identify are capable of causing disease phenotypes over multiple generations.

Among other modes of adaptation, a flexible epigenome likely improves the chances of survival of our descendants by allowing them to prepare for and adapt to the environmental conditions into which they are born. By further exploring the mechanisms involved in epigenetic inheritance, we will begin to better understand the causes and inheritance of noncommunicable diseases in the human population. Only then successful predictive and preventative measures can be implemented, albeit the results may take many generations to resolve.

## LIST OF ABBREVIATIONS

| | |
|---|---|
| **5hmC** | 5-hydroxymethylcytosine |
| **5mC** | 5-methylcytosine |
| **6mA** | N6-methyladenosine |
| ***A*$^{vy}$** | Agouti viable yellow |
| **BMI** | Body mass index |
| **CRISPR** | Clustered regularly interspaced short palindromic repeats |
| **DMR** | Differentially methylated region |
| **DNA** | Deoxyribonucleic acid |
| **DNMT** | Deoxyribonucleic acid (DNA) methyltransferase |
| **DOHaD** | Developmental Origins of Health and Disease |
| **F** | Filial generation |
| **IAP** | Intracisternal A-particle |
| **LINE1** | Long interspersed nuclear element-1 |
| **miRNA** | Microribonucleic acid |
| **ncRNA** | Noncoding ribonucleic acid |
| **piRNA** | Piwi-interacting ribonucleic acid |
| **PTSD** | Posttraumatic stress disorder |
| **rDNA** | Ribosomal deoxyribonucleic acid |
| **RNA** | Ribonucleic acid |
| **RNAi** | Ribonucleic acid interference |
| **TET** | Ten-eleven translocation |
| **tRNA** | Transfer ribonucleic acid |
| **WHO** | World Health Organization |

## GLOSSARY

| | |
|---|---|
| **Body mass index (BMI)** | An approximate quantification of the amount of tissue mass in an individual to categorize that person as normal weight, underweight, overweight, or obese. It is calculated by dividing the body mass by the square of the body height (kg/m$^2$). |
| **Developmental origins of health and disease (DOHaD)** | A theory that examines how environmental factors during fetal development interact with genotypic variation to change the capacity of the organism to cope with its environment in later life. For example, intrauterine growth restriction is associated with increased risk of adult-onset cardiovascular disease and diabetes. |
| **Differentially methylated region (DMR)** | Genomic regions with different methylation status across different cells/tissues from the same or different individuals. Usually refers to methylation at CpG sites in DNA. |
| **Environmental stressor** | An external stimulus, such as a chemical or biological agent (e.g., tobacco, pollution, alcohol), environmental condition (e.g., temperature, poor diet), or an event (e.g., social stressor, psychological stress) that causes stress in an organism. The stressor may lead to physical, mental, or epigenetic responses in an individual. |

**Epigenetic inheritance** A type of nonconventional inheritance that is independent of the DNA base sequence and involves the inheritance of an epigenetic factor from one generation to the next.

**Epigenetic reprogramming** Erasure and reestablishment of epigenetic marks during development, first during gametogenesis, and again in early embryogenesis. This allows the formation of totipotent cells in each new generation.

**Epigenome** Chemical changes to DNA and histones or changes in ncRNA expression that can be inherited and are involved in regulating gene expression and suppression of transposable elements. The epigenome is dynamic and can be altered by environmental conditions.

**Epimutation** Alterations in the pattern of normal epigenetic mark that has the potential of being inherited from one generation to the next.

**Filial generation** Pertaining to the sequence of generations following the parental generation. Each generation is designated by an "F" followed by a number indicating the generation (e.g., F1 is the first filial generation).

**Germ cell** A totipotent, haploid cell that combines with a germ cell from the opposite sex to form a new individual. Also known as a gamete.

**Intergenerational epigenetic inheritance** The transmission of non-DNA base sequence (i.e., epigenetic) information between generations via the germline. It affects offspring phenotype and/or epigenetic marks, which is not inherited past the F1 generation through the paternal lineage and the F2 generation through the maternal lineage (see Fig. 26.1). In this type of epigenetic inheritance, direct exposure of the offspring to the causative environmental stressor or to parental physiological effects cannot be mechanistically ruled out.

**Paramutation** Interaction between two alleles at a single genetic locus whereby one mutant allele induces a heritable epigenetic change in the other wild-type allele.

**Reconstruction model** In epigenetic inheritance, this mechanistic model suggests that epigenetic marks responsible for a phenotype undergo normal reprogramming in the germline and zygote but are then "reconstructed" or recreated by each successive generation (see Fig. 26.2).

**Repetitive DNA** Patterns of nucleic acids that occur in multiple copies throughout the genome. A significant fraction of the genome is repetitive. Repeats that occur in tandem repeat sequences or dispersed throughout the genome as transposable elements (e.g., transposons, retrotransposons).

**Replication model** In epigenetic inheritance, this mechanistic model requires epigenetic marks (e.g., DNA methylation, histone modifications) to be directly copied between cell divisions and generations (see Fig. 26.2).

**Slow-growth period** The period before the onset of puberty, usually between ages 9 and 12 in humans.

**Transgenerational epigenetic inheritance** The transmission of non-DNA base sequence information between generations via the germline. It affects offspring phenotype and/or epigenetic marks, which must be inherited at least to the F2 generation through the paternal lineage and to the F3 generation through the maternal lineage (see Fig. 26.1) to exclude direct effects of these individuals to the initiating environmental stressor.

## ACKNOWLEDGMENTS

G.E.T.B. is supported by a Wellcome Trust 4-year doctoral training programme studentship in Developmental Mechanisms. J.R. is supported by a Newton International Fellowship. E.D.W. is a Lister Research Prize Fellow (Lister Institute for Preventative Medicine).

## REFERENCES

[1] Heard E, Martienssen RA. Transgenerational epigenetic inheritance: myths and mechanisms. Cell 2014; 157(1):95−109.
[2] Daxinger L, Whitelaw E. Understanding transgenerational epigenetic inheritance via the gametes in mammals. Nat Rev Genet 2012;13(3):153−62.
[3] Blake GE, Watson ED. Unravelling the complex mechanisms of transgenerational epigenetic inheritance. Curr Opin Chem Biol 2016;33:101−7.
[4] Daxinger L, Whitelaw E. Transgenerational epigenetic inheritance: more questions than answers. Genome Res 2010;20(12):1623−8.
[5] Bohacek J, Mansuy IM. Epigenetic inheritance of disease and disease risk. Neuropsychopharmacology 2013;38(1):220−36.
[6] Javurek AB, Spollen WG, Ali AM, Johnson SA, Lubahn DB, Bivens NJ, et al. Discovery of a novel seminal fluid microbiome and influence of estrogen receptor alpha genetic status. Sci Rep 2016;6:23027.
[7] Raveh S, Sutalo S, Thonhauser KE, Thoss M, Hettyey A, Winkelser F, et al. Female partner preferences enhance offspring ability to survive an infection. BMC Evol Biol 2014;14:14.
[8] Bohacek J, Mansuy IM. A guide to designing germline-dependent epigenetic inheritance experiments in mammals. Br J Pharmacol 2017;14(3):243−9.
[9] Barker DJ, Osmond C. Infant mortality, childhood nutrition, and ischaemic heart disease in England and Wales. Lancet 1986;1(8489):1077−81.
[10] Barker DJ, Winter PD, Osmond C, Margetts B, Simmonds SJ. Weight in infancy and death from ischaemic heart disease. Lancet 1989;2(8663):577−80.
[11] Barker DJ, Gluckman PD, Godfrey KM, Harding JE, Owens JA, Robinson JS. Fetal nutrition and cardiovascular disease in adult life. Lancet 1993;341(8850):938−41.
[12] de Rooij SR, Painter RC, Roseboom TJ, Phillips DI, Osmond C, Barker DJ, et al. Glucose tolerance at age 58 and the decline of glucose tolerance in comparison with age 50 in people prenatally exposed to the Dutch famine. Diabetologia 2006;49(4):637−43.
[13] Ravelli AC, van Der Meulen JH, Osmond C, Barker DJ, Bleker OP. Obesity at the age of 50 y in men and women exposed to famine prenatally. Am J Clin Nutr 1999;70(5):811−6.
[14] Painter RC, De Rooij SR, Bossuyt PM, Osmond C, Barker DJ, Bleker OP, et al. A possible link between prenatal exposure to famine and breast cancer: a preliminary study. Am J Hum Biol 2006;18(6):853−6.
[15] Lumey LH. Decreased birthweights in infants after maternal in utero exposure to the Dutch famine of 1944-1945. Paediatr Perinat Epidemiol 1992;6(2):240−53.
[16] Huang C, Li Z, Narayan KM, Williamson DF, Martorell R. Bigger babies born to women survivors of the 1959-1961 Chinese famine: a puzzle due to survival selection? J Dev Orig Health Dis 2010;1(6):412−8.
[17] Painter RC, Osmond C, Gluckman P, Hanson M, Phillips DI, Roseboom TJ. Transgenerational effects of prenatal exposure to the Dutch famine on neonatal adiposity and health in later life. BJOG 2008;115(10): 1243−9.
[18] Lumey LH, Khalangot MD, Vaiserman AM. Association between type 2 diabetes and prenatal exposure to the Ukraine famine of 1932-33: a retrospective cohort study. Lancet Diabetes Endocrinol 2015;3(10): 787−94.

[19] Thurner S, Klimek P, Szell M, Duftschmid G, Endel G, Kautzky-Willer A, et al. Quantification of excess risk for diabetes for those born in times of hunger, in an entire population of a nation, across a century. Proc Natl Acad Sci USA 2013;110(12):4703−7.
[20] Slieker RC, Roost MS, van Iperen L, Suchiman HE, Tobi EW, Carlotti F, et al. DNA methylation landscapes of human fetal development. PLoS Genet 2015;11(10):e1005583.
[21] Lumey LH, Stein AD, Ravelli AC. Timing of prenatal starvation in women and birth weight in their first and second born offspring: the Dutch famine birth cohort study. Eur J Obstet Gynecol Reprod Biol 1995;61(1): 23−30.
[22] Guo F, Yan L, Guo H, Li L, Hu B, Zhao Y, et al. The transcriptome and DNA methylome landscapes of human primordial germ cells. Cell 2015;161(6):1437−52.
[23] Pembrey ME, Bygren LO, Kaati G, Edvinsson S, Northstone K, Sjostrom M, et al. Sex-specific, male-line transgenerational responses in humans. Eur J Hum Genet 2006;14(2):159−66.
[24] Bygren LO, Kaati G, Edvinsson S. Longevity determined by paternal ancestors' nutrition during their slow growth period. Acta Biotheor 2001;49(1):53−9.
[25] Kaati G, Bygren LO, Edvinsson S. Cardiovascular and diabetes mortality determined by nutrition during parents' and grandparents' slow growth period. Eur J Hum Genet 2002;10(11):682−8.
[26] Bygren LO, Tinghog P, Carstensen J, Edvinsson S, Kaati G, Pembrey ME, et al. Change in paternal grandmothers' early food supply influenced cardiovascular mortality of the female grandchildren. BMC Genet 2014;15:12.
[27] Alwasel SH, Abotalib Z, Aljarallah JS, Osmond C, Alkharaz SM, Alhazza IM, et al. Secular increase in placental weight in Saudi Arabia. Placenta 2011;32(5):391−4.
[28] Alwasel SH, Abotalib Z, Aljarallah JS, Osmond C, Alkharaz SM, Alhazza IM, et al. Changes in placental size during Ramadan. Placenta 2010;31(7):607−10.
[29] Northstone K, Golding J, Davey Smith G, Miller LL, Pembrey M. Prepubertal start of father's smoking and increased body fat in his sons: further characterisation of paternal transgenerational responses. Eur J Hum Genet 2014;22(12):1382−6.
[30] Dougan MM, Field AE, Rich-Edwards JW, Hankinson SE, Glynn RJ, Willett WC, et al. Is grand-parental smoking associated with adolescent obesity? A three-generational study. Int J Obes 2016;40(3):531−7.
[31] Miller LL, Pembrey M, Davey Smith G, Northstone K, Golding J. Is the growth of the fetus of a non-smoking mother influenced by the smoking of either grandmother while pregnant? PLoS One 2014;9(2): e86781.
[32] Li YF, Langholz B, Salam MT, Gilliland FD. Maternal and grandmaternal smoking patterns are associated with early childhood asthma. Chest 2005;127(4):1232−41.
[33] Gilliland FD, Li YF, Peters JM. Effects of maternal smoking during pregnancy and environmental tobacco smoke on asthma and wheezing in children. Am J Respir Crit Care Med 2001;163(2):429−36.
[34] Miller LL, Henderson J, Northstone K, Pembrey M, Golding J. Do grandmaternal smoking patterns influence the etiology of childhood asthma? Chest 2014;145(6):1213−8.
[35] Robledo CA, Yeung E, Mendola P, Sundaram R, Maisog J, Sweeney AM, et al. Preconception maternal and paternal exposure to persistent organic pollutants and birth size: the LIFE study. Environ Health Perspect 2015;123(1):88−94.
[36] Nadler A, Kav-Venaki S, Gleitman B. Transgenerational effects of the holocaust: externalization of aggression in second generation of holocaust survivors. J Consult Clin Psychol 1985;53(3):365−9.
[37] Yehuda R, Halligan SL, Bierer LM. Relationship of parental trauma exposure and PTSD to PTSD, depressive and anxiety disorders in offspring. J Psychiatr Res 2001;35(5):261−70.
[38] Yehuda R, Teicher MH, Seckl JR, Grossman RA, Morris A, Bierer LM. Parental posttraumatic stress disorder as a vulnerability factor for low cortisol trait in offspring of holocaust survivors. Arch Gen Psychiatry 2007;64(9):1040−8.

[39] Heijmans BT, Tobi EW, Stein AD, Putter H, Blauw GJ, Susser ES, et al. Persistent epigenetic differences associated with prenatal exposure to famine in humans. Proc Natl Acad Sci USA 2008;105(44):17046–9.
[40] Murphy SK, Adigun A, Huang Z, Overcash F, Wang F, Jirtle RL, et al. Gender-specific methylation differences in relation to prenatal exposure to cigarette smoke. Gene 2012;494(1):36–43.
[41] Tobi EW, Slieker RC, Stein AD, Suchiman HE, Slagboom PE, van Zwet EW, et al. Early gestation as the critical time-window for changes in the prenatal environment to affect the adult human blood methylome. Int J Epidemiol 2015;44(4):1211–23.
[42] Miska EA, Ferguson-Smith AC. Transgenerational inheritance: models and mechanisms of non–DNA sequence–based inheritance. Science 2016;354(6308):59.
[43] Jablonka E. Epigenetic inheritance and plasticity: the responsive germline. Prog Biophys Mol Biol 2013; 111(2–3):99–107.
[44] Reik W, Surani MA. Germline and pluripotent stem cells. Cold Spring Harb Perspect Biol 2015;7(11).
[45] Seisenberger S, Andrews S, Krueger F, Arand J, Walter J, Santos F, et al. The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells. Mol Cell 2012;48(6):849–62.
[46] Seki Y, Hayashi K, Itoh K, Mizugaki M, Saitou M, Matsui Y. Extensive and orderly reprogramming of genome-wide chromatin modifications associated with specification and early development of germ cells in mice. Dev Biol 2005;278(2):440–58.
[47] Casas E, Vavouri T. Sperm epigenomics: challenges and opportunities. Front Genet 2014;5:330.
[48] Erkek S, Hisano M, Liang CY, Gill M, Murr R, Dieker J, et al. Molecular determinants of nucleosome retention at CpG-rich sequences in mouse spermatozoa. Nat Struct Mol Biol 2013;20(7):868–75.
[49] Carone BR, Hung JH, Hainer SJ, Chou MT, Carone DM, Weng Z, et al. High-resolution mapping of chromatin packaging in mouse embryonic stem cells and sperm. Dev Cell 2014;30(1):11–22.
[50] Smallwood SA, Tomizawa S, Krueger F, Ruf N, Carli N, Segonds-Pichon A, et al. Dynamic CpG island methylation landscape in oocytes and preimplantation embryos. Nat Genet 2011;43(8):811–4.
[51] van der Heijden GW, Derijck AA, Ramos L, Giele M, van der Vlag J, de Boer P. Transmission of modified nucleosomes from the mouse male germline to the zygote and subsequent remodeling of paternal chromatin. Dev Biol 2006;298(2):458–69.
[52] Tang WWC, Kobayashi T, Irie N, Dietmann S, Surani MA. Specification and epigenetic programming of the human germ line. Nat Rev Genet 2016;17(10):585–600.
[53] Tang Walfred WC, Dietmann S, Irie N, Leitch Harry G, Floros Vasileios I, Bradshaw Charles R, et al. A unique gene regulatory network resets the human germline epigenome for development. Cell 2015; 161(6):1453–67.
[54] Hackett JA, Sengupta R, Zylicz JJ, Murakami K, Lee C, Down TA, et al. Germline DNA demethylation dynamics and imprint erasure through 5-hydroxymethylcytosine. Science 2013;339(6118):448–52.
[55] Radford EJ, Ito M, Shi H, Corish JA, Yamazawa K, Isganaitis E, et al. In utero undernourishment perturbs the adult sperm methylome and intergenerational metabolism. Science 2014;345(6198):785–93.
[56] Aiken CE, Ozanne SE. Transgenerational developmental programming. Hum Reprod Update 2014;20(1): 63–75.
[57] Sun Q, Huang S, Wang X, Zhu Y, Chen Z, Chen D. N6-methyladenine functions as a potential epigenetic mark in eukaryotes. BioEssays 2015;37(11):1155–62.
[58] Huang S, Chen D. N6-methyladenine: a potential epigenetic mark in eukaryotes. Oncotarget 2015;6(18): 15744–5.
[59] Deaton AM, Bird A. CpG islands and the regulation of transcription. Genes Dev 2011;25(10):1010–22.
[60] Slotkin RK, Martienssen R. Transposable elements and the epigenetic regulation of the genome. Nat Rev Genet 2007;8(4):272–85.
[61] Lawrence M, Daujat S, Schneider R. Lateral thinking: how histone modifications regulate gene expression. Trends Genet 2016;32(1):42–56.

[62] Marmorstein R, Trievel RC. Histone modifying enzymes: structures, mechanisms, and specificities. Biochim Biophys Acta 2009;1789(1):58−68.
[63] Brunner AM, Nanni P, Mansuy IM. Epigenetic marking of sperm by post-translational modification of histones and protamines. Epigenetics Chromatin 2014;7(1):2.
[64] Holoch D, Moazed D. RNA-mediated epigenetic regulation of gene expression. Nat Rev Genet 2015;16(2): 71−84.
[65] Rando OJ. Intergenerational transfer of epigenetic information in sperm. Cold Spring Harb Perspect Med 2016;6(5).
[66] Veselovska L, Smallwood SA, Saadeh H, Stewart KR, Krueger F, Maupetit-Méhouas S, et al. Deep sequencing and de novo assembly of the mouse oocyte transcriptome define the contribution of transcription to the DNA methylation landscape. Genome Biol 2015;16:209.
[67] Zhang X, Cozen AE, Liu Y, Chen Q, Lowe TM. Small RNA modifications: integral to function and disease. Trends Mol Med 2016;22(12):1025−34.
[68] Vojtech L, Woo S, Hughes S, Levy C, Ballweber L, Sauteraud RP, et al. Exosomes in human semen carry a distinctive repertoire of small non-coding RNAs with potential regulatory functions. Nucleic Acids Res 2014;42(11):7290−304.
[69] Cossetti C, Lugini L, Astrologo L, Saggio I, Fais S, Spadafora C. Soma-to-germline transmission of RNA in mice xenografted with human tumour cells: possible transport by exosomes. PLoS One 2014;9(7): e101629.
[70] Almouzni G, Cedar H. Maintenance of epigenetic information. Cold Spring Harb Perspect Biol 2016;8(5).
[71] Rasmussen KD, Helin K. Role of TET enzymes in DNA methylation, development, and cancer. Genes Dev 2016;30(7):733−50.
[72] Choi I, Kim R, Lim HW, Kaestner KH, Won KJ. 5-hydroxymethylcytosine represses the activity of enhancers in embryonic stem cells: a new epigenetic signature for gene regulation. BMC Genom 2014;15:670.
[73] Ashe A, Sapetschnig A, Weick EM, Mitchell J, Bagijn MP, Cording AC, et al. piRNAs can trigger a multigenerational epigenetic memory in the germline of *C. elegans*. Cell 2012;150(1):88−99.
[74] Daxinger L, Oey H, Isbel L, Whitelaw NC, Youngson NA, Spurling A, et al. Hypomethylation of ERVs in the sperm of mice haploinsufficient for the histone methyltransferase Setdb1 correlates with a paternal effect on phenotype. Sci Rep 2016;6:25004.
[75] Kankel MW, Ramsey DE, Stokes TL, Flowers SK, Haag JR, Jeddeloh JA, et al. Arabidopsis MET1 cytosine methyltransferase mutants. Genetics 2003;163(3):1109−22.
[76] Soppe WJ, Jacobsen SE, Alonso-Blanco C, Jackson JP, Kakutani T, Koornneef M, et al. The late flowering phenotype of fwa mutants is caused by gain-of-function epigenetic alleles of a homeodomain gene. Mol Cell 2000;6(4):791−802.
[77] Greer EL, Maures TJ, Ucar D, Hauswirth AG, Mancini E, Lim JP, et al. Transgenerational epigenetic inheritance of longevity in *Caenorhabditis elegans*. Nature 2011;479(7373):365−71.
[78] Morgan HD, Sutherland HGE, Martin DIK, Whitelaw E. Epigenetic inheritance at the agouti locus in the mouse. Nat Genet 1999;23(3):314−8.
[79] Cropley JE, Suter CM, Beckman KB, Martin DI. CpG methylation of a silent controlling element in the murine Avy allele is incomplete and unresponsive to methyl donor supplementation. PLoS One 2010;5(2): e9055.
[80] Feil R, Fraga MF. Epigenetics and the environment: emerging patterns and implications. Nat Rev Genet 2012;13(2):97−109.
[81] Miltenberger RJ, Mynatt RL, Wilkinson JE, Woychik RP. The role of the agouti gene in the yellow obese syndrome. J Nutr 1997;127(9):1902S−7S.
[82] Cropley JE, Eaton SA, Aiken A, Young PE, Giannoulatou E, Ho JW, et al. Male-lineage transmission of an acquired metabolic phenotype induced by grand-paternal obesity. Mol Metab 2016;5(8):699−708.

[83] Dolinoy DC, Huang D, Jirtle RL. Maternal nutrient supplementation counteracts bisphenol A-induced DNA hypomethylation in early development. Proc Natl Acad Sci USA 2007;104(32):13056–61.
[84] Blewitt ME, Vickaryous NK, Paldi A, Koseki H, Whitelaw E. Dynamic reprogramming of DNA methylation at an epigenetically sensitive allele in mice. PLoS Genet 2006;2(4):e49.
[85] Dias BG, Ressler KJ. Parental olfactory experience influences behavior and neural structure in subsequent generations. Nat Neurosci 2014;17(1):89–96.
[86] Vallaster MP, Kukreja S, Bing XY, Ngolab J, Zhao-Shea R, Gardner PD, et al. Paternal nicotine exposure alters hepatic xenobiotic metabolism in offspring. eLife 2017;6:e24771.
[87] Leenen FAD, Muller CP, Turner JD. DNA methylation: conducting the orchestra from exposure to phenotype? Clin Epigenetics 2016;8(1):92.
[88] Song Y, Wu N, Wang S, Gao M, Song P, Lou J, et al. Transgenerational impaired male fertility with an Igf2 epigenetic defect in the rat are induced by the endocrine disruptor p,p′-DDE. Hum Reprod 2014;29(11):2512–21.
[89] Crichton JH, Dunican DS, Maclennan M, Meehan RR, Adams IR. Defending the genome from the enemy within: mechanisms of retrotransposon suppression in the mouse germline. Cell Mol Life Sci 2014;71(9):1581–605.
[90] Klosin A, Casas E, Hidalgo-Carcedo C, Vavouri T, Lehner B. Transgenerational transmission of environmental information in *C. elegans*. Science 2017;356(6335):320–3.
[91] Padmanabhan N, Jia D, Geary-Joo C, Wu X, Ferguson-Smith AC, Fung E, et al. Mutation in folate metabolism causes epigenetic instability and transgenerational effects on development. Cell 2013;155(1):81–93.
[92] Elmore CL, Wu X, Leclerc D, Watson ED, Bottiglieri T, Krupenko NI, et al. Metabolic derangement of methionine and folate metabolism in mice deficient in methionine synthase reductase. Mol Genet Metab 2007;91(1):85–97.
[93] Wei Y, Yang CR, Wei YP, Zhao ZA, Hou Y, Schatten H, et al. Paternally induced transgenerational inheritance of susceptibility to diabetes in mammals. Proc Natl Acad Sci USA 2014;111(5):1873–8.
[94] Jimenez-Chillaron JC, Isganaitis E, Charalambous M, Gesta S, Pentinat-Pelegrin T, Faucette RR, et al. Intergenerational transmission of glucose intolerance and obesity by in utero undernutrition in mice. Diabetes 2009;58(2):460–8.
[95] Shea JM, Serra RW, Carone BR, Shulha HP, Kucukural A, Ziller MJ, et al. Genetic and epigenetic variation, but not diet, shape the sperm methylome. Dev Cell 2015;35(6):750–8.
[96] Smallwood SA, Lee HJ, Angermueller C, Krueger F, Saadeh H, Peat J, et al. Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. Nat Methods 2014;11(8):817–20.
[97] Robertson KD, Jones PA. DNA methylation: past, present and future directions. Carcinogenesis 2000;21(3):461–7.
[98] Sved J, Bird A. The expected equilibrium of the CpG dinucleotide in vertebrate genomes under a mutation model. Proc Natl Acad Sci USA 1990;87(12):4692–6.
[99] Kim S, Gunesdogan U, Zylicz JJ, Hackett JA, Cougot D, Bao S, et al. PRMT5 protects genomic integrity during global DNA demethylation in primordial germ cells and preimplantation embryos. Mol Cell 2014;56(4):564–79.
[100] Skinner MK. Endocrine disruptors in 2015: epigenetic transgenerational inheritance. Nat Rev Endocrinol 2016;12(2):68–70.
[101] Anway MD, Leathers C, Skinner MK. Endocrine disruptor vinclozolin induced epigenetic transgenerational adult-onset disease. Endocrinology 2006;147(12):5515–23.
[102] Guerrero-Bosagna C, Covert TR, Haque MM, Settles M, Nilsson EE, Anway MD, et al. Epigenetic transgenerational inheritance of vinclozolin induced mouse adult onset disease and associated sperm epigenome biomarkers. Reprod Toxicol 2012;34(4):694–707.

[103] Iqbal K, Tran DA, Li AX, Warden C, Bai AY, Singh P, et al. Deleterious effects of endocrine disruptors are corrected in the mammalian germline by epigenome reprogramming. Genome Biol 2015;16:59.
[104] Skinner MK, Guerrero-Bosagna C, Haque MM. Environmentally induced epigenetic transgenerational inheritance of sperm epimutations promote genetic mutations. Epigenetics 2015;10(8):762−71.
[105] Oey H, Isbel L, Hickey P, Ebaid B, Whitelaw E. Genetic and epigenetic variation among inbred mouse littermates: identification of inter-individual differentially methylated regions. Epigenetics Chromatin 2015;8:54.
[106] Calado RT, Dumitriu B. Telomere dynamics in mice and humans. Semin Hematol 2013;50(2):165−74.
[107] Aiken CE, Tarry-Adkins JL, Ozanne SE. Transgenerational developmental programming of ovarian reserve. Sci Rep 2015;5:16175.
[108] Liebers R, Rassoulzadegan M, Lyko F. Epigenetic regulation by heritable RNA. PLoS Genet 2014;10(4): e1004296.
[109] Feng X, Guang S. Small RNAs, RNAi and the inheritance of gene silencing in *Caenorhabditis elegans*. J Genet Genom 2013;40(4):153−60.
[110] Rechavi O, Houri-Ze'evi L, Anava S, Goh WS, Kerk SY, Hannon GJ, et al. Starvation-induced transgenerational inheritance of small RNAs in *C. elegans*. Cell 2014;158(2):277−87.
[111] Gapp K, Jawaid A, Sarkies P, Bohacek J, Pelczar P, Prados J, et al. Implication of sperm RNAs in transgenerational inheritance of the effects of early trauma in mice. Nat Neurosci 2014;17(5):667−9.
[112] Sharma U, Conine CC, Shea JM, Boskovic A, Derr AG, Bing XY, et al. Biogenesis and function of tRNA fragments during sperm maturation and fertilization in mammals. Science 2016;351(6271):391−6.
[113] Chen Q, Yan M, Cao Z, Li X, Zhang Y, Shi J, et al. Sperm tsRNAs contribute to intergenerational inheritance of an acquired metabolic disorder. Science 2016;351(6271):397−400.
[114] Gapp K, Bohacek J, Grossmann J, Brunner AM, Manuella F, Nanni P, et al. Potential of environmental enrichment to prevent transgenerational effects of paternal trauma. Neuropsychopharmacology 2016;41.
[115] Kiani J, Grandjean V, Liebers R, Tuorto F, Ghanbarian H, Lyko F, et al. RNA-mediated epigenetic heredity requires the cytosine methyltransferase Dnmt2. PLoS Genet 2013;9(5):e1003498.
[116] Yuan S, Oliver D, Schuster A, Zheng H, Yan W. Breeding scheme and maternal small RNAs affect the efficiency of transgenerational inheritance of a paramutation in mice. Sci Rep 2015;5:9266.
[117] Chandler VL. Paramutation's properties and puzzles. Science 2010;330(6004):628−9.
[118] Rassoulzadegan M, Grandjean V, Gounon P, Vincent S, Gillot I, Cuzin F. RNA-mediated non-mendelian inheritance of an epigenetic change in the mouse. Nature 2006;441(7092):469−74.
[119] Wagner KD, Wagner N, Ghanbarian H, Grandjean V, Gounon P, Cuzin F, et al. RNA induction and inheritance of epigenetic cardiac hypertrophy in the mouse. Dev Cell 2008;14(6):962−9.
[120] Rassoulzadegan M, Cuzin F. Epigenetic heredity: RNA-mediated modes of phenotypic variation. Ann NY Acad Sci 2015;1341:172−5.
[121] Grandjean V, Fourre S, De Abreu DA, Derieppe MA, Remy JJ, Rassoulzadegan M. RNA-mediated paternal heredity of diet-induced obesity and metabolic disorders. Sci Rep 2015;5:18193.
[122] Greer EL, Beese-Sims SE, Brookes E, Spadafora R, Zhu Y, Rothbart SB, et al. A histone methylation network regulates transgenerational epigenetic memory in *C. elegans*. Cell Rep 2014;7(1):113−26.
[123] Greer EL, Blanco MA, Gu L, Sendinc E, Liu J, Aristizabal-Corrales D, et al. DNA methylation on N6-adenine in *C. elegans*. Cell 2015;161(4):868−78.
[124] Siklenka K, Erkek S, Godmann M, Lambrot R, McGraw S, Lafleur C, et al. Disruption of histone methylation in developing sperm impairs offspring health transgenerationally. Science 2015;350.
[125] Prokopuk L, Western PS, Stringer JM. Transgenerational epigenetic inheritance: adaptation through the germline epigenome? Epigenomics 2015;7(5):829−46.
[126] Ford SP, Long NM. Evidence for similar changes in offspring phenotype following either maternal undernutrition or overnutrition: potential impact on fetal epigenetic mechanisms. Reprod Fertil Dev 2011; 24(1):105−11.

[127] Kinoshita T, Seki M. Epigenetic memory for stress response and adaptation in plants. Plant Cell Physiol 2014;55(11):1859−63.
[128] Zheng X, Chen L, Xia H, Wei H, Lou Q, Li M, et al. Transgenerational epimutations induced by multi-generation drought imposition mediate rice plant's adaptation to drought condition. Sci Rep 2017;7:39843.
[129] Wang Y, Liu H, Sun Z. Lamarck rises from his grave: parental environment-induced epigenetic inheritance in model organisms and humans. Biol Rev Camb Philos Soc 2017;92.
[130] Trerotola M, Relli V, Simeone P, Alberti S. Epigenetic inheritance and the missing heritability. Hum Genom 2015;9:17.
[131] Choi CS, Gonzales EL, Kim KC, Yang SM, Kim JW, Mabunga DF, et al. The transgenerational inheritance of autism-like phenotypes in mice exposed to valproic acid during pregnancy. Sci Rep 2016;6:36250.
[132] Gregory DJ, Kobzik L, Yang Z, McGuire CC, Fedulov AV. Transgenerational transmission of asthma risk after exposure to environmental particles during pregnancy. Am J Physiol Lung Cell Mol Physiol 2017; 313(2):L395−405.
[133] Hilton IB, D'Ippolito AM, Vockley CM, Thakore PI, Crawford GE, Reddy TE, et al. Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. Nat Biotechnol 2015;33(5):510−7.
[134] Maeder ML, Angstman JF, Richardson ME, Linder SJ, Cascio VM, Tsai SQ, et al. Targeted DNA demethylation and activation of endogenous genes using programmable TALE-TET1 fusion proteins. Nat Biotechnol 2013;31(12):1137−42.

Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Chemical Biology

# Unravelling the complex mechanisms of transgenerational epigenetic inheritance

Georgina ET Blake[1] and Erica D Watson[1,2]

There are numerous benefits to elucidating how our environment affects our health: from a greater understanding of adaptation to disease prevention. Evidence shows that stressors we are exposed to during our lifetime might cause disease in our descendants. Transgenerational epigenetic inheritance involves the transmission of 'information' over multiple generations via the gametes independent of the DNA base sequence. Despite extensive research, the epigenetic mechanisms remain unclear. Analysis of model organisms exposed to environmental insults (e.g., diet manipulation, stress, toxin exposure) or carrying mutations in the epigenetic regulatory machinery indicates that inheritance of altered DNA methylation, histone modifications, or non-coding RNAs are key mechanisms. Tracking inherited epigenetic information and its effects for multiple generations is a significant challenge to overcome.

**Addresses**
[1] Department of Physiology, Development and Neuroscience, University of Cambridge, Cambridge CB2 3EG, UK
[2] Centre for Trophoblast Research, University of Cambridge, Cambridge CB2 3EG, UK

Corresponding author: Watson, Erica D (edw23@cam.ac.uk)

**Current Opinion in Chemical Biology** 2016, **33**:101–107

This review comes from a themed issue on **Chemical genetics and epigenetics**

Edited by **Danica G Fujimori** and **Stuart Conway**

http://dx.doi.org/10.1016/j.cbpa.2016.06.008



## Introduction

In recent years, the concept that epigenetic factors are inherited has rapidly developed. As more studies show that environmental stressors (e.g., poor diet, toxins, or psychological stress [1••,2••,3,4]) influence the epigenome, it is becoming clear that the environment experienced during our lifetime may impact the health of our descendants. How commonplace epigenetic inheritance is and the underlying mechanisms remain uncertain, though substantial research over the last few years have improved our understanding of this phenomenon.

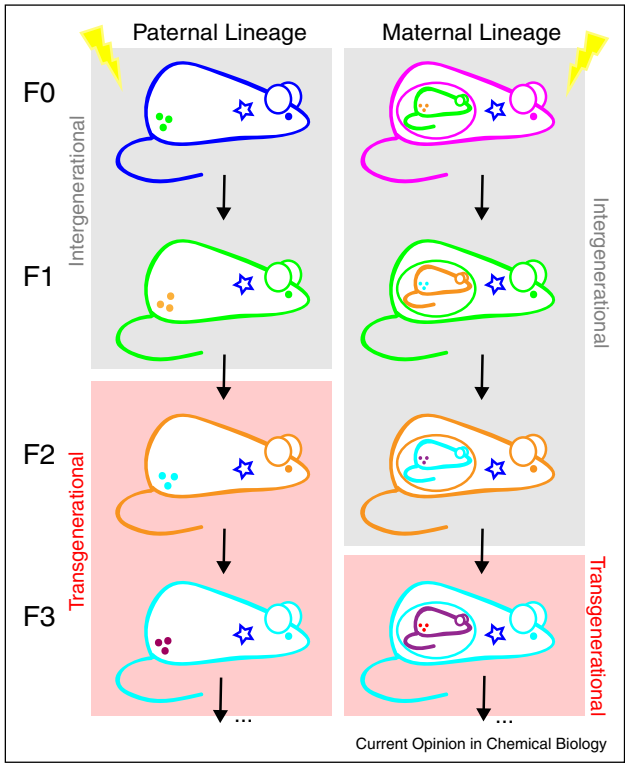
We define transgenerational epigenetic inheritance (TEI) as the transmission of non-DNA base sequence information between generations via the germline [5,6]. Epigenetic changes in the first generation (parental generation (F0)) occurring after exposure to an environmental insult increases risk for specific phenotypes in subsequent generations (first filial generation (F1), second filial generation (F2), third filial generation (F3), among others) even when they are not exposed to the insult themselves. To be transgenerationally inherited, the phenotype must persist beyond the F2 generation when inherited via the paternal lineage and F3 generation via the maternal lineage [5,6] (Figure 1). Both sperm and oocytes [7,8•,9••] transmit epigenetic information to the next generation, but paternal inheritance is typically studied for experimental tractability and lack of confounding influences (e.g., the uterine environment). A multigenerational search for inherited epigenetic factors, such as DNA and histone methylation, and non-coding (nc) RNAs, has ensued.

TEI in human populations is becoming evident [10], though it is difficult to study due to long generation times, genetic diversity and variable environmental conditions wherein we live. Plant and animal models of TEI, in which genetic and environmental conditions are meticulously controlled, are key for mechanistic exploration and for overcoming the challenges associated with tracking epigenetic information over multiple generations. A greater understanding of TEI will have important implications for disease risk prediction and prevention.

## DNA methylation: an important mechanistic candidate

Methylation of single DNA residues is well studied in the context of TEI. In mammals, 5-methylcytosine (5mC) is the predominant form of methylated DNA. In organisms (e.g., bacteria, fungi, *Caenorhabditis elegans* and *Drosophila melanogaster*) that lack or have low levels of 5mC, other forms of methylated DNA, such as the recently identified N6-methyladenine (6mA), are widespread [11]. 5mC is generally associated with gene repression [12] whereas 6mA is thought to promote activity [13]. The reality may be far more complex; linking methylation status to a specific gene expression profile and phenotype is challenging. For DNA methylation to be a heritable epigenetic mark, it should be mitotically and meiotically stable [5,6] and escape epigenetic reprogramming that normally occurs in primordial germ cells and post-fertilization embryos [14–16] (Figure 2). This epigenetic 'erasure' generates a totipotent state required to form the next generation [16]. Remarkably, 5mC within specific

**Figure 1**




Comparing transgenerational epigenetic inheritance (TEI) between the paternal and maternal lineages. Epigenetic alterations and phenotypes induced by environmental insults in the F0 generation may be inherited via the germline over several generations (F1, F2, F3, among others). In the paternal lineage: TEI occurs if direct exposure of an F0 male and his germ cells to an environmental insult causes a phenotype (star) and/or alters epigenetic patterns beyond the F1 generation. The F2 offspring is the first generation that was not directly exposed to the insult. In the maternal lineage: if environmental exposure occurs while a female is pregnant, the mother, the foetus (F1 generation) and its primordial germ cells (F2 generation) are all directly exposed. Thus, the persistence of phenotypes/epigenetic changes in the F3 generation and beyond is considered TEI [5,6]. Intergenerational inheritance is the term given to phenotypes/epigenetic effects that persist to only the F1 offspring via the paternal lineage and the F2 offspring via the maternal lineage. F0, parental generation; F1, first filial generation; F2, second filial generation; F3, third filial generation.

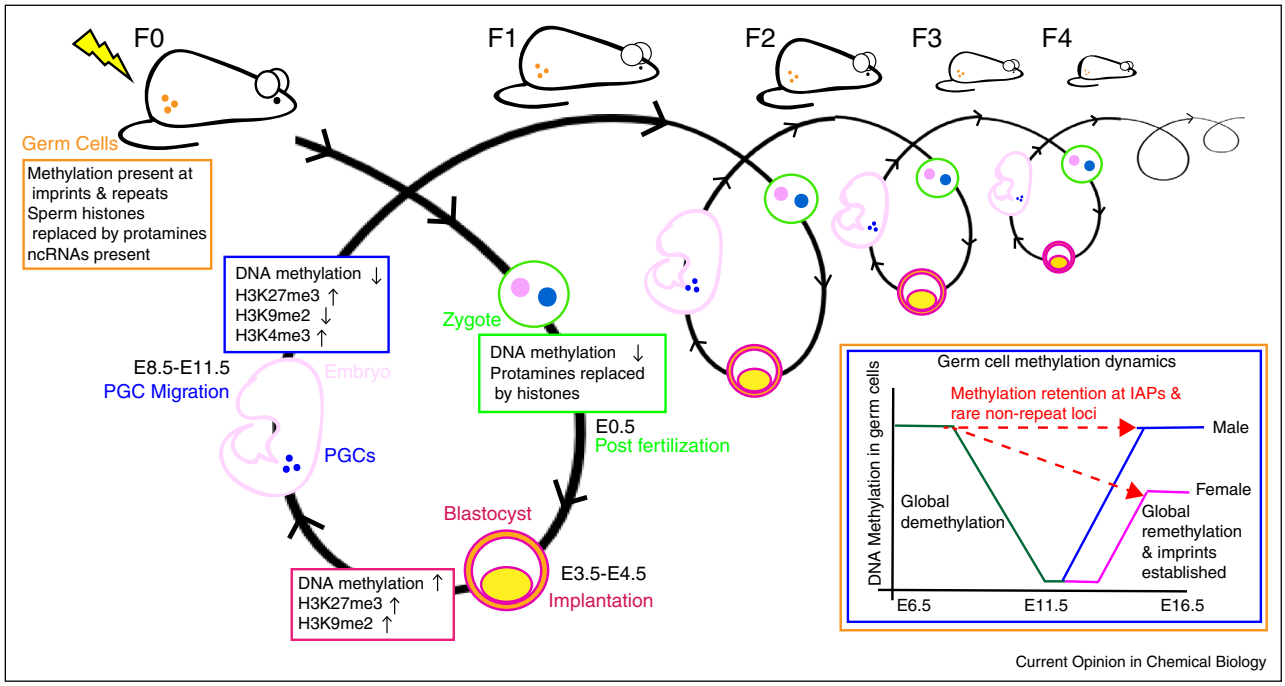
genomic regions including repeat sequences (e.g., intracisternal A particles [IAPs]) and rare regulatory elements (e.g., promoters next to IAPs) is resistant to reprogramming [14,15] (Figure 2). Presumably, this occurs to maintain genomic stability during widespread erasure [17]. Abnormal DNA methylation patterns caused by environmental stressors would have to generate resistance to reprogramming to appear and cause phenotypes in subsequent generations.

Owing to their resistance to reprogramming, the methylation status of repetitive elements is a mechanistic candidate of TEI [6,15]. A classic mouse model of TEI involving an IAP element is the agouti viable yellow ($A^{vy}$) epiallele [18]. Hypomethylation of a cryptic promoter in the IAP element upstream of the agouti gene drives its expression leading to a yellow coat colour, obesity and diabetes [19]. This hypomethylated status is inherited over several generations through the maternal line [18] and can be manipulated by environmental factors [20,21]. For example, providing a methyl-rich diet to $A^{vy}$ females decreases the frequency of yellow coats in their offspring [20,21]. It is unclear whether DNA methylation at the IAP element is normalised [21] or if an indirect effect is responsible [20].

Beyond the $A^{vy}$ model, it has been difficult to identify differentially methylated regions (DMRs) in the genome that are stable over multiple generations and that correlate with a phenotype. This is even when unbiased approaches to assess the germline methylome are implemented. One successful example is in a pre-diabetic mouse model characterized by insulin resistance and impaired fasting glucose [22]. F0 males transmit a similar pre-diabetic phenotype to the F1 and F2 generations [22]. Whole 5mC methylome analysis of sperm from F0 males revealed altered DNA methylation patterns compared to controls [22]. However, only a few of these abnormal patterns persisted in pancreatic islets of the male F1 and F2 offspring [22]. Conversely, unbiased methylome analysis of sperm from mice (F1) exposed to severe undernutrition while *in utero*, revealed altered DNA methylation that coincided with reduced birth weight and a robust metabolic phenotype [2••,23]. Over 100 DMRs concentrated in CpG islands and intergenic regions were identified [2••]. However, the subset of DMRs that were assessed in F2 somatic tissues were not maintained, even though neighbouring genes showed altered expression and the metabolic phenotype was observed [2••]. This suggests a parallel epigenetic mechanism may be involved. Future methylome-wide analysis of the F2 generation and beyond will more thoroughly determine whether DMRs are inherited.

Reproducibility of TEI data is another challenge. An example of this is the rodent vinclozolin model [24]. Males (F1) exposed *in utero* to the endocrine disruptor vinclozolin transmit several adult onset diseases up to the F4 generation [4]. Analysis of promoter regions revealed widespread alteration of 5mC in mature sperm of the F3 generation following ancestral vinclozolin exposure [25]. However, others showed that altered DNA methylation patterns in purified prospermatogonia of the F1 offspring were not apparent in the F2 generation [26]. The discrepancy between studies may come down to technical differences, including the sperm population assessed and method of methylation analysis used, or it may reflect the natural epigenetic variability that exists between individuals [27].

**Figure 2**




In mammals, inherited epigenetic information must escape multiple epigenetic reprogramming events in germ cells and the early embryo. Reprogramming involves dynamic changes in the epigenetic patterns within the DNA of the germ cells and pre-implantation embryo between each generation to re-establish pluripotency. This excludes some repetitive elements (e.g., IAPs) and rare non-repeat loci, which remain highly methylated [15]. The graph (bottom right) indicates DNA methylation dynamics of germ cells [6]. In cases of transgenerational inheritance, abnormal epigenetic marks caused by an environmental insult must escape multiple rounds of these reprogramming events. How these marks are stably transmitted between generations is the focus of much research. H3K27me3, histone 3 lysine 27 trimethylation; H3K4me3, histone 3 lysine 4 trimethylation; H3K9me2, histone 3 lysine 9 dimethylation; ncRNAs, non-coding RNAs; IAP, intracisternal A particle; E, embryonic day; F0, parental generation; F1, first filial generation; F2, second filial generation; F3, third filial generation; F4, fourth filial generation.

## Dysregulation of methylation machinery may initiate TEI

The machinery vital for the establishment and maintenance of DNA methylation may be an important initiator of TEI. In *Arabidopsis thaliana*, a mutation in the *DNA METHYLTRANSFERASE 1* (*MET1*) gene leads to heritable hypomethylation at a repetitive region near the transcriptional start site of the *FLOWERING WAGENINGEN* (*FWA*) gene [28]. This hypomethylation leads to ectopic *FWA* expression causing a late flowering phenotype for several wildtype generations [29]. Similarly, mutations in the mouse homolog of *MET1*, DNA methyltransferase 1 (*Dnmt1*), cause an analogous effect. Wildtype offspring derived from males mutant for *Dnmt1* showed a greater frequency than expected of DNA hypomethylation at the agouti locus and a yellow coat [30]. Importantly, whether DNA methyltransferases contribute to the mechanism of TEI beyond these epialleles requires further exploration.

Alternatively, in *Drosophila*, the DNA 6mA demethylase (DMAD) suppresses transposon expression in the ovary by ensuring low 6mA levels at these sites [13]. Although it is unclear whether dysregulation of DMAD and 6mA at transposable elements causes a transgenerational effect, it may play a yet-to-be determined mechanism in the *Drosophila* TEI model whereby females are fed a high calorie diet results in obesity in the F2 generation [31].

Remarkably, limiting the substrate for DNA methyltransferases leads to transgenerational effects on development [3]. A mutation in the mouse methionine synthase reductase *(Mtrr)* gene, which is necessary for the transmission of one-carbon methyl groups [32], results in epigenetic instability and the inheritance of congenital abnormalities at least up to four wildtype generations [3]. Even though these transgenerational effects occur through the maternal lineage, embryo transfer experiments demonstrated that the consequences were via the germline and independent of the uterine environment [3]. Specific germline-inherited epimutations have not yet been identified in the *Mtrr* model nor is it clear whether the regulation of DNA methylation machinery is affected.

## Is there a role for histone modifications in TEI?

The inheritance of histone modifications is not as well studied when considering TEI mechanisms. Most

histones in mouse (99%) and human (85%) sperm are removed and replaced by protamines to enable compact packaging of DNA during sperm maturation [33]. Recently, protamine modifications were identified [34], yet whether or not the 'protamine code' passes on epigenetic information between generations is uncertain. Histone retention in sperm tends to be at the promoters of housekeeping and developmentally-regulated genes [35] while histones are retained throughout the genome in the oocyte [36]. Whether abnormal histone modifications in either germ cell influence the phenotype of the offspring is under investigation. Recent evidence suggests that histone modifications and their regulatory enzymes convey epigenetic memory across generations. In *C. elegans*, histone 3 lysine 27 trimethylation (H3K27me3) regulated by the polycomb repressive complex 2 (PRC2) transmits memory of X-chromosome repression transgenerationally [37]. In another example, even though deficiencies in the H3K4me3 regulatory complex in *C. elegans* lead to increased longevity that persists transgenerationally, global H3K4me3 levels appear normal in the offspring [38]. Likewise, ectopic expression of *KDM1a*, a human H3K4 demethylase, during mouse spermatogenesis causes developmental abnormalities for three wildtype generations [39•]. Regardless, wildtype sperm of the F1 generation displayed normal epigenome-wide H3K4me2 profiles as well as normal DNA methylation patterns [39•]. Therefore, while disruption of the histone methylation machinery may initiate transgenerational inheritance of a phenotype, a second epigenetic factor may be involved.

Interconnection of epigenetic mechanisms are exemplified in worms with a mutation in a *KDM1a* ortholog (*spr-5*). The *spr-5* mutants have a progressive transgenerational decline in fertility and an accumulation of H3K4me2 [40]. Correspondingly, 6mA levels also increase transgenerationally in these mutants [41••] indicating another epigenetic mechanism is present. When a 6mA DNA methyltransferase was knocked down in *spr-5* mutant worms, the transgenerational loss of fertility phenotype was partially suppressed [41••]. Cross-talk between these two epigenetic pathways is evident [41••], but further experiments to determine the nature of these interactions are required.

## Non-coding RNAs: linking soma to germline

A mechanistic role of non-coding RNA (ncRNAs) is currently at the forefront of TEI research. Small ncRNAs act as sequence guides directing DNA or histone methylation, and by post-transcriptionally regulating mRNA [42]. RNA inheritance is best studied in *C. elegans* [43]. Starvation-induced expression of small RNAs or exogenous RNA interference (RNAi) results in heritable gene silencing that persists for several generations [44,45]. Although the mechanism is complex, it is hypothesized that piwi-interacting RNA (piRNA), which typically silences transposons in the germline, and exogenous RNAi may converge into a common pathway requiring secondary small RNAs and chromatin regulatory complexes to ultimately bring about stable TEI [45].

RNA inheritance also occurs in mammals. ncRNAs from mouse sperm exposed to an environmental stressor are sufficient to cause phenotypes [1••,9••,46]. For example, traumatic stress in mice (F1) due to maternal separation in early postnatal life is associated with behavioural phenotypes in the F2 male offspring [1••]. Deep sequencing of F1 sperm revealed upregulation of several microRNAs (miRNAs), which when microinjected into fertilized oocytes led to similar behavioural phenotypes in the resulting offspring [1••]. This technique demonstrates a causal relationship between germline RNA and phenotype. Similarly, mice fed either a high fat [9••] or low protein diet [8•] have increased levels of fragmented tRNA species in sperm and offspring with metabolic disease [8•,9••]. Fragmented tRNAs can repress genes associated with the endogenous retroelement, MERVL, and might influence feto-placental development [8•]. Synthetic versions of high fat diet-induced fragmented tRNAs in sperm were insufficient to cause metabolic disease [9••]. This might be because the synthetic tRNAs lacked necessary modifications. Indeed, RNA methylation mediated by the methyltransferase *Dnmt2* is required for the transmission of phenotype in the *Kit* paramutant model [47]. These studies indicate that ncRNA may be a mechanism for TEI, although whether this method of inheritance is sustained in subsequent generations is yet-to-be determined. Remarkably, sperm tRNA fragments may originate in the epididymis and may be transported extracellularly to sperm by exosomes [8•]. Thus, exosomes derived from the male genital tract may communicate the environmental conditions experienced by the paternal generation to his mature sperm [8•,48,49].

## Challenges

Identifying the heritable epigenetic information transmitted across multiple generations is difficult even in models with definitive phenotypic inheritance. The following reasons contribute to this challenge.

Firstly, only selected epigenetic loci are assessed in some studies attempting to show TEI. As a result, the full scope of epigenetic changes in each generation is not appreciated. In fact, a spectrum of epigenetic information (i.e., DNA methylation, histone modifications, and RNA expression) may act in concert to initiate and perpetuate the inheritance of phenotypes [39•,41••]. Ideally, we need to perform unbiased, large-scale studies incorporating epigenome-wide, genome-wide, and transcriptome-wide approaches over several generations in key models of TEI. This type of comprehensive analysis is costly, and likely will require collaboration between groups.

Secondly, an environmental insult may stochastically affect the epigenome in each germ cell, as evidenced

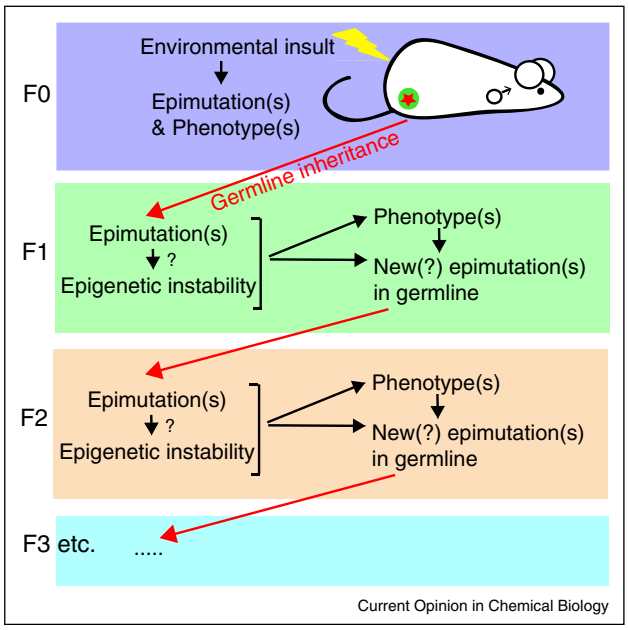
by phenotypic variability within a single model [3], in addition to naturally-occurring epivariation between individuals [27]. Consequently, resolving specific epimutations is difficult when germ cells are pooled for analysis. The emergence of single cell methylome and transcriptome technologies will permit us to better understand germ cell heterogeneity [50].

Thirdly, different 'epimutations' may be established in each generation caused by epigenetic instability in the previous generation (Figure 3). In this case, the search for stable epimutations transmitted over multiple generations may be fruitless. Support for this hypothesis comes from the observation that phenotypes frequently persist over more generations than identified epigenetic changes [1••,2••].

Fourthly, epigenetic instability might promote genetic instability. Indeed, genetic background (e.g., inbred vs outbred mice) can alter the susceptibility of an individual to transgenerational epigenetic effects [25]. Alternatively, the activation of transposable elements in the germline by DNA hypomethylation might lead to heritable genetic mutations [51]. Furthermore, analysis of the F3 generation following vinclozolin exposure in rats revealed changes in DNA methylation patterns associated with a significant increase in repeat element copy number variations [52•]. It is also possible that epigenetic and genetic mechanisms might interact in TEI through telomere regulation. Telomeres are heterochromatic tandem repeats rich in repressive histone marks [53] that normally protect chromosome ends from degradation [54]. Telomere shortening is associated with aging-related diseases [54] and can occur in response to diet manipulation. For example, feeding female rats a low protein diet results in an intergenerational reduction in telomere length associated with premature reproductive aging in the F2 female offspring [55], though the F3 generation was not assessed to confirm a transgenerational effect. Exploring the epigenetic status and stability of telomeres in transgenerational models may open up a new line of questioning.

Lastly, attributing phenotype to particular epigenetic changes can be problematic. Utilizing epigenetic strategies, such as TALE–TET1-fusions [56] and CRISPR-Cas9 based acetyltransferases [57], will enable us to target and alter epigenetic marks in vivo to determine the specific effects on gene expression and phenotype. These technologies are in their infancy and are currently limited by off-target effects. However, they provide many exciting possibilities for site-directed epimutagenesis.

**Figure 3**




Hypothesis: new epimutations may be generated in each generation. Some models of TEI reveal that phenotypes caused by an environmental insult persist over more generations than identified epigenetic abnormalities. This might be because epimutations inherited through the germline lead to more extensive epigenetic instability in the F1 offspring. This may result in an abnormal physiological or molecular milieu that causes new epimutations in the germ cells (i.e., F2 generation). For transgenerational inheritance to occur, epigenetically instability would be recreated in each subsequent generation. This hypothesis suggests that a different epigenetic profile would be expected in each individual of each generation rather than finding single stable epimutations that are consistently inherited. Red arrow, germline epigenetic inheritance; Red star, germ cell with one or more epimutation. F0, parental generation; F1, first filial generation; F2, second filial generation; F3, third filial generation; F4, fourth filial generation.

## Conclusions

Lamark's once discredited hypothesis that phenotypes acquired during a lifetime are passed on to offspring has been injected with new vitality, fuelling fresh perspectives on rapid adaption to a changing environment [58]. Human populations are likely affected by TEI as demonstrated by the Dutch Hunger Winter and Överkalix famine studies [10,59,60]. Fundamentally, a greater mechanistic understanding of TEI will impact our approach to disease prevention and prediction, the effects of which will hopefully have a lasting impact.

## Acknowledgements

GETB is supported by a studentship from the Wellcome Trust 4-year PhD programme in Developmental Mechanisms. EDW is a Lister Research Prize fellow and was supported by an Isaac Newton Trust/Wellcome Trust ISSF/University of Cambridge joint research grant.

## References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. •• Gapp K, Jawaid A, Sarkies P, Bohacek J, Pelczar P, Prados J, Farinelli L, Miska E, Mansuy IM: **Implication of sperm RNAs in**

**transgenerational inheritance of the effects of early trauma in mice**. *Nat Neurosci* 2014, **17**:667-669.
By microinjecting sperm RNA into fertilized oocytes, this is one of the first studies to demonstrate a direct causal link between miRNAs in sperm and phenotype in the next generation.

2. •• Radford EJ, Ito M, Shi H, Corish JA, Yamazawa K, Isganaitis E, Seisenberger S, Hore TA, Reik W, Erkek S *et al.*: ***In utero* undernourishment perturbs the adult sperm methylome and intergenerational metabolism**. *Science* 2014, **345**:785-793.
Important for using unbiased methods, this study shows over 100 loci with altered DNA methylation in sperm of mice that were undernourished *in utero*. However, these epimutations did not persist to the F2 generation even though metabolic disease and gene misexpression were observed.

3. Padmanabhan N, Jia D, Geary-Joo C, Wu X, Ferguson-Smith AC, Fung E, Bieda MC, Snyder FF, Gravel RA, Cross JC *et al.*: **Mutation in folate metabolism causes epigenetic instability and transgenerational effects on development**. *Cell* 2013, **155**:81-93.

4. Anway MD, Leathers C, Skinner MK: **Endocrine disruptor vinclozolin induced epigenetic transgenerational adult-onset disease**. *Endocrinology* 2006, **147**:5515-5523.

5. Daxinger L, Whitelaw E: **Understanding transgenerational epigenetic inheritance via the gametes in mammals**. *Nat Rev Genet* 2012, **13**:153-162.

6. Heard E, Martienssen RA: **Transgenerational epigenetic inheritance: myths and mechanisms**. *Cell* 2014, **157**:95-109.

7. Huypens P, Sass S, Wu M, Dyckhoff D, Tschop M, Theis F, Marschall S, de Angelis MH, Beckers J: **Epigenetic germline inheritance of diet-induced obesity and insulin resistance**. *Nat Genet* 2016 http://dx.doi.org/10.1038/ng.3527.

8. • Sharma U, Conine CC, Shea JM, Boskovic A, Derr AG, Bing XY, Belleannee C, Kucukural A, Serra RW, Sun F *et al.*: **Biogenesis and function of tRNA fragments during sperm maturation and fertilization in mammals**. *Science* 2015 http://dx.doi.org/10.1126/science.aad6780.
The authors suggest that sperm obtain tRNA fragments by fusion with exosomes extruded by epididymal cells, thus linking somatic and germ cells. These tRNA fragments may act by repressing endogenous retro-elements in the subsequent embryo.

9. •• Chen Q, Yan M, Cao Z, Li X, Zhang Y, Shi J, Feng GH, Peng H, Zhang X, Zhang Y *et al.*: **Sperm tsRNAs contribute to intergenerational inheritance of an acquired metabolic disorder**. *Science* 2015 http://dx.doi.org/10.1126/science.aad7977.
The authors identify the significance of sperm small tRNAs in transmitting metabolic disease to the next generation. Furthermore, synthetic tRNAs are unable to mediate this process suggesting that RNA modification is important in TEI models.

10. Pembrey M, Saffery R, Bygren LO: **Human transgenerational responses to early-life experience: potential impact on development, health and biomedical research**. *J Med Genet* 2014, **51**:563-572.

11. Sun Q, Huang S, Wang X, Zhu Y, Chen Z, Chen D: **N6-methyladenine functions as a potential epigenetic mark in eukaryotes**. *Bioessays* 2015, **37**:1155-1162.

12. Deaton AM, Bird A: **CpG islands and the regulation of transcription**. *Genes Dev* 2011, **25**:1010-1022.

13. Zhang G, Huang H, Liu D, Cheng Y, Liu X, Zhang W, Yin R, Zhang D, Zhang P, Liu J *et al.*: **N6-methyladenine DNA modification in Drosophila**. *Cell* 2015, **161**:893-906.

14. Smallwood SA, Tomizawa S, Krueger F, Ruf N, Carli N, Segonds-Pichon A, Sato S, Hata K, Andrews SR, Kelsey G: **Dynamic CpG island methylation landscape in oocytes and preimplantation embryos**. *Nat Genet* 2011, **43**:811-814.

15. Hackett JA, Sengupta R, Zylicz JJ, Murakami K, Lee C, Down TA, Surani MA: **Germline DNA demethylation dynamics and imprint erasure through 5-hydroxymethylcytosine**. *Science* 2013, **339**:448-452.

16. Reik W, Surani MA: **Germline and pluripotent stem cells**. *Cold Spring Harb Perspect Biol* 2015:7.

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

18. Morgan HD, Sutherland HGE, Martin DIK, Whitelaw E: **Epigenetic inheritance at the agouti locus in the mouse**. *Nat Genetics* 1999, **23**:314-318.

19. Miltenberger RJ, Mynatt RL, Wilkinson JE, Woychik RP: **The role of the agouti gene in the yellow obese syndrome**. *J Nutr* 1997, **127**:1902S-1907S.

20. Cropley JE, Suter CM, Beckman KB, Martin DI: **CpG methylation of a silent controlling element in the murine Avy allele is incomplete and unresponsive to methyl donor supplementation**. *PLoS One* 2010, **5**:e9055.

21. Dolinoy DC, Huang D, Jirtle RL: **Maternal nutrient supplementation counteracts bisphenol A-induced DNA hypomethylation in early development**. *Proc Natl Acad Sci U S A* 2007, **104**:13056-13061.

22. Wei Y, Yang CR, Wei YP, Zhao ZA, Hou Y, Schatten H, Sun QY: **Paternally induced transgenerational inheritance of susceptibility to diabetes in mammals**. *Proc Natl Acad Sci U S A* 2014, **111**:1873-1878.

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

24. Nadeau JH: **The nature of evidence for and against epigenetic inheritance**. *Genome Biol* 2015, **16**:137.

25. Guerrero-Bosagna C, Covert TR, Haque MM, Settles M, Nilsson EE, Anway MD, Skinner MK: **Epigenetic transgenerational inheritance of vinclozolin induced mouse adult onset disease and associated sperm epigenome biomarkers**. *Reprod Toxicol* 2012, **34**:694-707.

26. Iqbal K, Tran DA, Li AX, Warden C, Bai AY, Singh P, Wu X, Pfeifer GP, Szabo PE: **Deleterious effects of endocrine disruptors are corrected in the mammalian germline by epigenome reprogramming**. *Genome Biol* 2015, **16**:59.

27. Shea JM, Serra RW, Carone BR, Shulha HP, Kucukural A, Ziller MJ, Vallaster MP, Gu H, Tapper AR, Gardner PD *et al.*: **Genetic and epigenetic variation, but not diet, shape the sperm methylome**. *Dev Cell* 2015, **35**:750-758.

28. Kankel MW, Ramsey DE, Stokes TL, Flowers SK, Haag JR, Jeddeloh JA, Riddle NC, Verbsky ML, Richards EJ: **Arabidopsis MET1 cytosine methyltransferase mutants**. *Genetics* 2003, **163**:1109-1122.

29. Soppe WJJ, Jacobsen SE, Alonso-Blanco C, Jackson JP, Kakutani T, Koornneef M, Peeters AJM: **The late flowering phenotype of fwa mutants is caused by gain-of-function epigenetic alleles of a homeodomain gene**. *Mol Cell* 2000, **6**:791-802.

30. Chong S, Vickaryous N, Ashe A, Zamudio N, Youngson N, Hemley S, Stopka T, Skoultchi A, Matthews J, Scott HS *et al.*: **Modifiers of epigenetic reprogramming show paternal effects in the mouse**. *Nat Genetics* 2007, **39**:614-622.

31. Buescher JL, Musselman LP, Wilson CA, Lang T, Keleher M, Baranski TJ, Duncan JG: **Evidence for transgenerational metabolic programming in Drosophila**. *Dis Model Mech* 2013, **6**:1123-1132.

32. Elmore CL, Wu X, Leclerc D, Watson ED, Bottiglieri T, Krupenko NI, Krupenko SA, Cross JC, Rozen R, Gravel RA *et al.*: **Metabolic derangement of methionine and folate metabolism in mice deficient in methionine synthase reductase**. *Mol Genet Metab* 2007, **91**:85-97.

33. Casas E, Vavouri T: **Sperm epigenomics: challenges and opportunities**. *Front Genet* 2014, **5**:330.

34. Brunner AM, Nanni P, Mansuy IM: **Epigenetic marking of sperm by post-translational modification of histones and protamines**. *Epigenetics Chromatin* 2014, **7**:2.

35. Erkek S, Hisano M, Liang CY, Gill M, Murr R, Dieker J, Schubeler D, van der Vlag J, Stadler MB, Peters AH: **Molecular determinants of nucleosome retention at CpG-rich sequences in mouse spermatozoa**. *Nat Struct Mol Biol* 2013, **20**:868-875.

36. Gu L, Wang Q, Sun QY: **Histone modifications during mammalian oocyte maturation: dynamics, regulation and functions**. *Cell Cycle* 2010, **9**:1942-1950.

37. Gaydos LJ, Wang W, Strome S: **Gene repression. H3K27me and PRC2 transmit a memory of repression across generations and during development**. *Science* 2014, **345**:1515-1518.

38. Greer EL, Maures TJ, Ucar D, Hauswirth AG, Mancini E, Lim JP, Benayoun BA, Shi Y, Brunet A: **Transgenerational epigenetic inheritance of longevity in *Caenorhabditis elegans***. *Nature* 2011, **479**:365-371.

39. • Siklenka K, Erkek S, Godmann M, Lambrot R, McGraw S, Lafleur C, Cohen T, Xia J, Suderman M, Hallett M *et al.*: **Disruption of histone methylation in developing sperm impairs offspring health transgenerationally**. *Science* 2015 http://dx.doi.org/10.1126/science.aab2006.

This is the first study to show TEI downstream of altered histone methylation in mice. It also examines the broader epigenome including DNA and histone methylation, and the transcriptome.

40. Greer EL, Beese-Sims SE, Brookes E, Spadafora R, Zhu Y, Rothbart SB, Aristizabal-Corrales D, Chen S, Badeaux AI, Jin Q *et al.*: **A histone methylation network regulates transgenerational epigenetic memory in *C. elegans***. *Cell Rep* 2014, **7**:113-126.

41. •• Greer EL, Blanco MA, Gu L, Sendinc E, Liu J, Aristizabal-Corrales D, Hsu CH, Aravind L, He C, Shi Y: **DNA methylation on N6-adenine in *C. elegans***. *Cell* 2015, **161**:868-878.

This study indicates crosstalk between 6mA, the newly identified methylated DNA base, and histone modifications in a transgenerational *C. elegans* model displaying progressive decline in fertility.

42. Yan W: **Potential roles of noncoding RNAs in environmental epigenetic transgenerational inheritance**. *Mol Cell Endocrinol* 2014, **398**:24-30.

43. Feng X, Guang S: **Small RNAs, RNAi and the inheritance of gene silencing in *Caenorhabditis elegans***. *J Genet Genomics* 2013, **40**:153-160.

44. Rechavi O, Houri-Ze'evi L, Anava S, Goh WS, Kerk SY, Hannon GJ, Hobert O: **Starvation-induced transgenerational inheritance of small RNAs in *C. elegans***. *Cell* 2014, **158**:277-287.

45. Ashe A, Sapetschnig A, Weick EM, Mitchell J, Bagijn MP, Cording AC, Doebley AL, Goldstein LD, Lehrbach NJ, Le Pen J *et al.*: **piRNAs can trigger a multigenerational epigenetic memory in the germline of *C. elegans***. *Cell* 2012, **150**:88-99.

46. Grandjean V, Fourre S, De Abreu DA, Derieppe MA, Remy JJ, Rassoulzadegan M: **RNA-mediated paternal heredity of diet-induced obesity and metabolic disorders**. *Sci Rep* 2015, **5**:18193.

47. Kiani J, Grandjean V, Liebers R, Tuorto F, Ghanbarian H, Lyko F, Cuzin F, Rassoulzadegan M: **RNA-mediated epigenetic heredity requires the cytosine methyltransferase Dnmt2**. *PLoS Genet* 2013, **9**:e1003498.

48. Vojtech L, Woo S, Hughes S, Levy C, Ballweber L, Sauteraud RP, Strobl J, Westerberg K, Gottardo R, Tewari M *et al.*: **Exosomes in human semen carry a distinctive repertoire of small non-coding RNAs with potential regulatory functions**. *Nucleic Acids Res* 2014, **42**:7290-7304.

49. Cossetti C, Lugini L, Astrologo L, Saggio I, Fais S, Spadafora C: **Soma-to-germline transmission of RNA in mice xenografted with human tumour cells: possible transport by exosomes**. *PLoS One* 2014, **9**:e101629.

50. Smallwood SA, Lee HJ, Angermueller C, Krueger F, Saadeh H, Peat J, Andrews SR, Stegle O, Reik W, Kelsey G: **Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity**. *Nat Methods* 2014, **11**:817-820.

51. Kim S, Gunesdogan U, Zylicz JJ, Hackett JA, Cougot D, Bao S, Lee C, Dietmann S, Allen GE, Sengupta R *et al.*: **PRMT5 protects genomic integrity during global DNA demethylation in primordial germ cells and preimplantation embryos**. *Mol Cell* 2014, **56**:564-579.

52. • Skinner MK, Guerrero-Bosagna C, Haque MM: **Environmentally induced epigenetic transgenerational inheritance of sperm epimutations promote genetic mutations**. *Epigenetics* 2015, **10**:762-771.

This paper highlights the effect of epigenetic instability on genetic stability in the vinclozolin model of TEI. The authors demonstrate an increase of copy number variants associated with altered DNA methylation in the F3 generation.

53. Dan J, Yang J, Liu Y, Xiao A, Liu L: **Roles for histone acetylation in regulation of telomere elongation and two-cell state in mouse ES cells**. *J Cell Physiol* 2015, **230**:2337-2344.

54. Calado RT, Dumitriu B: **Telomere dynamics in mice and humans**. *Seminars Hematol* 2013, **50**:165-174.

55. Aiken CE, Tarry-Adkins JL, Ozanne SE: **Transgenerational developmental programming of ovarian reserve**. *Sci Rep* 2015, **5**:16175.

56. Maeder ML, Angstman JF, Richardson ME, Linder SJ, Cascio VM, Tsai SQ, Ho QH, Sander JD, Reyon D, Bernstein BE *et al.*: **Targeted DNA demethylation and activation of endogenous genes using programmable TALE-TET1 fusion proteins**. *Nat Biotechnol* 2013, **31**:1137-1142.

57. Hilton IB, D'Ippolito AM, Vockley CM, Thakore PI, Crawford GE, Reddy TE, Gersbach CA: **Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers**. *Nat Biotechnol* 2015, **33**:510-517.

58. Whitelaw E: **Disputing Lamarckian epigenetic inheritance in mammals**. *Genome Biol* 2015, **16**:60.

59. Bygren LO, Tinghog P, Carstensen J, Edvinsson S, Kaati G, Pembrey ME, Sjostrom M: **Change in paternal grandmothers' early food supply influenced cardiovascular mortality of the female grandchildren**. *BMC Genet* 2014, **15**:12.

60. Veenendaal MV, Painter RC, de Rooij SR, Bossuyt PM, van der Post JA, Gluckman PD, Hanson MA, Roseboom TJ: **Transgenerational effects of prenatal exposure to the 1944-45 Dutch famine**. *Bjog* 2013, **120**:548-553.