Defective folate metabolism causes germline epigenetic instability and distinguishes *Hira* as a phenotype inheritance biomarker

Georgina E.T. Blake[1,2,†], Xiaohui Zhao[1,2], Hong wa Yung[1,2], Graham J. Burton[1,2], Anne C. Ferguson-Smith[2,3], Russell S. Hamilton[2,3], Erica D. Watson[1,2,*]

[1]Department of Physiology, Development & Neuroscience, University of Cambridge, Cambridge UK

[2]Centre for Trophoblast Research, University of Cambridge, Cambridge UK

[3]Department of Genetics, University of Cambridge, Cambridge, UK

[†]Current address: College of Medicine and Health, University of Exeter Medical School, Exeter, UK

*Corresponding author: edw23@cam.ac.uk

ABSTRACT

The mechanism behind transgenerational epigenetic inheritance (TEI) is unclear, particularly through the maternal grandparental line. We previously showed that disruption of folate metabolism in mice by the *Mtrr*^{gt} hypomorphic mutation results in TEI of congenital malformations. Either maternal grandparent can initiate this phenomenon, which persists for at least four wildtype generations. In this work, we use a genome-wide approach to reveal genetic stability in the *Mtrr*^{gt} model and epigenome-wide differential DNA methylation in the germline of *Mtrr*$^{+/gt}$ maternal grandfathers. While epigenetic reprogramming occurs, wildtype grandprogeny and great grandprogeny exhibit transcriptional memory of germline methylation defects. One region encompasses the *Hira* gene, which is misexpressed in embryos at least until the F3 generation in a manner that distinguishes *Hira* transcript expression as a biomarker of maternal phenotypic inheritance.

INTRODUCTION

Environmental stressors can impact an individual's health and that of their progeny[1-5]. Phenotypic risk that persists for several generations in the absence of the stressor is termed transgenerational epigenetic inheritance (TEI)[6]. While the mechanism is unclear, this non-conventional inheritance likely occurs independent of DNA base-sequence mutations and involves the inheritance of an epigenetic factor(s) via the germline[6,7]. Candidate factors in mammals include: DNA methylation, histone modifications, and/or non-coding RNA[1,2,4,8-10]. How an epigenetic message resists reprogramming and is transmitted between one or even multiple generations to cause disease remains elusive. Few mammalian models of TEI exist and most focus on paternal inheritance[1,2,8-11]. We reported the *Mtrr*^{gt} mouse line, a rare model of maternal grandparental TEI in which congenital malformations are transgenerationally inherited for at least four wildtype generations[3] (Supplementary Fig. 1b-c). Phenotype inheritance persists even after blastocyst transfer of grandprogeny into control uteri (Supplementary Fig. 1d) indicating that TEI occurred independent of the maternal environment and instead via the germline[3].

MTRR (methionine synthase reductase) is a key enzyme required for one-carbon metabolism (i.e., folate and methionine metabolism; Supplementary Fig. 1a)[12-14]. Folate is a well-known vitamin important for neural tube closure, yet its function in development is complex and poorly understood. One-carbon metabolism is required for thymidine synthesis[15] and cellular methylation. Indeed, it transmits methyl groups for the methylation of homocysteine by methionine synthase (MTR) to form methionine and tetrahydrofolate[16]. Methionine acts as a precursor for S-adenosylmethionine (SAM), which serves as the sole methyl-donor for substrates involved in epigenetic regulation (e.g., DNA, RNA, histones)[17-19]. MTRR activates MTR through the reductive methylation of its vitamin B_{12} cofactor[14] (Supplementary Fig. 1a). Consequently, progression of one-carbon metabolism requires MTRR to maintain genetic and epigenetic stability.

The hypomorphic *Mtrr*^{gt} mutation reduces wildtype *Mtrr* transcript expression in mice[3,12] and consequently disrupts one-carbon metabolism by diminishing MTR activity[12]. Similar to humans with an *MTRR* mutation[13,20-22] or dietary folate deficiency[23], *Mtrr*$^{gt/gt}$ mice display hyperhomocysteinemia[3,12] and macrocytic anaemia[24], as well as altered DNA methylation patterns associated with gene misexpression[3] and a range of developmental phenotypes at midgestation

(e.g., growth defects and/or congenital malformations including heart, placenta, and neural tube closure defects)[3]. Therefore, *Mtrr*^{gt} mice are suitable for studying defective folate metabolism.

Crucially, the *Mtrr*^{gt} mouse line is a model of maternal grandparental TEI[3]. Through highly-controlled genetic pedigrees (Supplementary Fig. 1b-c), we demonstrated that F0 generation *Mtrr*$^{+/gt}$ male or female mice can initiate TEI of developmental phenotypes in the wildtype (*Mtrr*$^{+/+}$) descendants until the F4 generation[3]. The types of phenotypes at embryonic day (E) 10.5 are similar to those in *Mtrr*$^{gt/gt}$ conceptuses (see above)[3]. Regardless of whether an F0 *Mtrr*$^{+/gt}$ male or female initiates TEI, the spectrum and frequency of developmental phenotypes in the F2-F4 wildtype generations are largely comparable between pedigrees. An exception is the F1 generation where phenotypic risk at E10.5 occurs only when individuals are derived from an F0 *Mtrr*$^{+/gt}$ female (Supplementary Fig. 1b)[3]. Despite an absence of gross phenotypes at E10.5, F1 wildtype mice derived from F0 *Mtrr*$^{+/gt}$ males display indicators of direct epigenetic inheritance including locus-specific epigenetic dysregulation in placentas associated with gene misexpression at E10.5[3], a hematopoietic phenotype later in life[24], and the ability to perpetuate TEI in their offspring similar to F1 wildtype females derived from an *Mtrr*$^{+/gt}$ female[3]. Altogether, TEI originates in gametes from F0 *Mtrr*$^{+/gt}$ males or females leading to phenotypic inheritance via the F1 wildtype daughters until the F4 wildtype generation[3]. The mechanism is currently unclear.

In this work, we investigate potential mechanism(s) of TEI in the *Mtrr*^{gt} model using a genome-wide approach. First, we demonstrate that *Mtrr*$^{gt/gt}$ mice are genetically stable and hence reassert focus on an epigenetic mechanism. Second, we show that germline DNA methylation is altered in F0 *Mtrr*$^{+/gt}$ males. F0 sperm were chosen for analysis because: i) F0 *Mtrr*$^{+/gt}$ males initiate TEI in a similar manner to F0 *Mtrr*$^{+/gt}$ females, ii) sperm are more experimental tractable than oocytes, and iii) when assessing heritable effects, the uterine environment does not need to be controlled for in F0 *Mtrr*$^{+/gt}$ males. Even though differentially methylated regions (DMRs) in sperm of F0 *Mtrr*$^{+/gt}$ males are reprogrammed in somatic tissue of wildtype F1 and F2 progeny, our data shows evidence of transcriptional memory of germline epigenetic disruption that persists at least until the F3 generation. Transcriptional memory of sperm DMRs includes misexpression of *Hira*, a gene important for chromatin stability[25,26] and ribosome assembly[27], which we propose as a biomarker and potential mediator of maternal phenotypic inheritance in the *Mtrr*^{gt} model.

RESULTS

Genetic stability in *Mtrr*^{gt} mice

Since one-carbon metabolism is directly linked to DNA synthesis[15], we first addressed whether the *Mtrr*^{gt} allele influences genetic stability. Whole genome sequencing (WGS) was performed on phenotypically normal C57Bl/6J control embryos (N=2) and *Mtrr*$^{gt/gt}$ embryos with congenital malformations (N=6) (Supplementary Fig. 1c,f). DNA libraries were sequenced separately resulting in ~30x coverage per embryo (~3.5 x 10^8 paired-end reads/genome). The sequenced genomes were compared to the C57Bl/6J reference genome to identify structural variants (SVs) and single nucleotide polymorphisms (SNPs).

The *Mtrr*^{gt} mutation was generated by a gene-trap (gt) insertion into intron 9 of the *Mtrr* gene (Chr13) in the 129P2Ola/Hsd mouse strain before eight generations of backcrossing into the C57Bl/6J strain[3]. As a result, the majority of variants identified in *Mtrr*$^{gt/gt}$ embryos were located on Chr13 in the genomic region surrounding the *Mtrr* locus (Supplementary Fig. 2a-b). These variants included the gene-trap and several SNPs that showed sequence similarity to the 129P2Ola/Hsd genome and likely persisted due to *Mtrr*^{gt} genotype selection and regional crossover frequency. Variant identification in this region acted as an internal positive control of our bioinformatic method, demonstrating that it is capable of distinguishing genetic differences between experimental groups. Using these SNPs, we defined a 20 Mb region of 129P2Ola/Hsd sequence surrounding the *Mtrr*^{gt} allele (Fig. 1a). When this region was bioinformatically masked, C57Bl/6J and *Mtrr*$^{gt/gt}$ embryos contained a similar mean (± sd) frequency of SNPs (C57Bl/6J: 4,871 ± 791 SNPs/embryo; *Mtrr*$^{gt/gt}$: 5,138 ± 398 SNPs/embryo; p=0.781) and SVs (C57Bl/6J: 342 SVs/embryo; *Mtrr*$^{gt/gt}$: 301 SVs/embryo; p=0.6886; Fig. 1b,c) implying that the de novo mutation rate was unchanged by the *Mtrr*$^{gt/gt}$ mutation. These values were in line with expected de novo mutation rates[28]. Only 25 (21 SNPs and 4 SVs) variants were present in all six *Mtrr*$^{gt/gt}$ embryos and absent in C57Bl/6J embryos (Supplementary Fig. 2e,f, Supplementary Tables 1-2). When all SNPs and SVs were considered, the majority represented non-coding variants or were located in non-coding regions (Supplementary Fig. 2c,d). Moreover, genetic variation within the masked region had minimal functional effect (beyond the gene-trap insertion) since no variant overlapped with a known enhancer and expression of individual genes was similar among C57Bl/6J, 129P2Ola/Hsd and

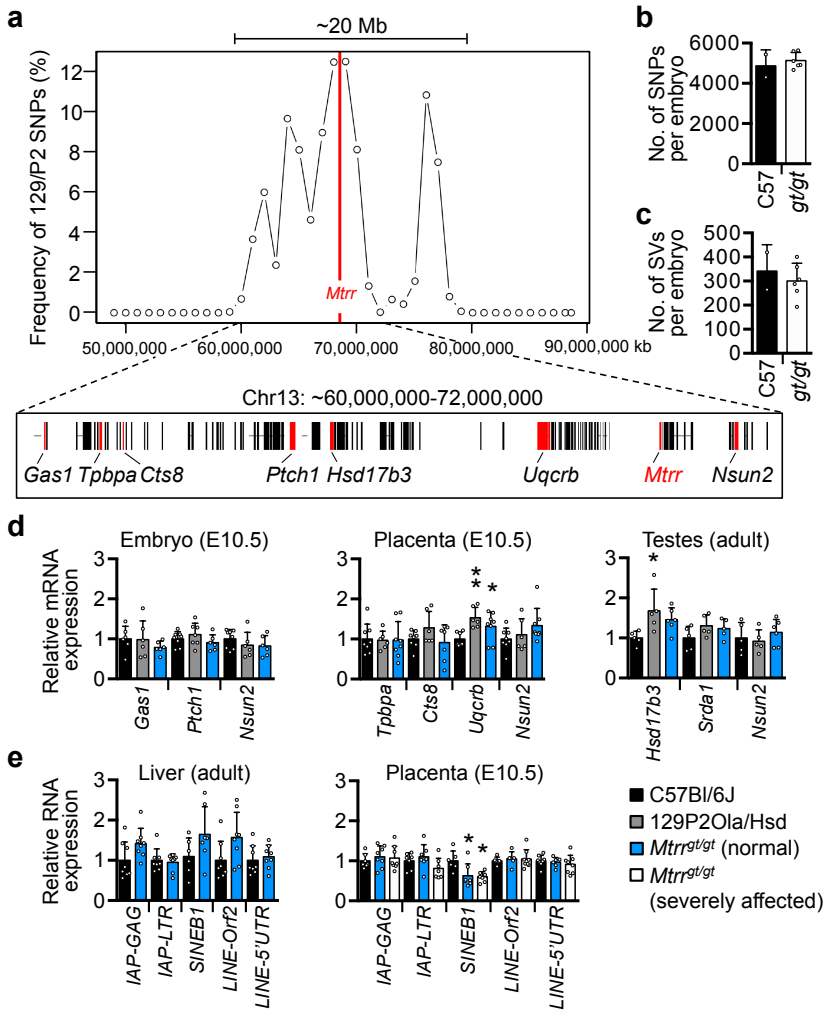


Figure 1. **The *Mtrr*^{gt} mouse line is genetically stable. a-c** Whole genome sequencing (WGS) of normal C57Bl/6 embryos (N=2 embryos) and severely affected *Mtrr*$^{gt/gt}$ embryos (N=6 embryos) at E10.5 to determine the frequency of genetic variants compared to the C57Bl/6J reference genome. **a** The frequency of 129P2Ola/Hsd single nucleotide polymorphisms (SNPs) in the region surrounding the gene-trap insertion site in the *Mtrr* gene (red line). The majority of genes within the 20 Mb region surrounding the *Mtrr* gene are shown below the graph. **b, c** The average number of (**b**) SNPs and (**c**) structural variants (SVs) per embryo in C57Bl/6 embryos (C57, black bars) and *Mtrr*$^{gt/gt}$ embryos (*gt/gt*, white bars). The 20 Mb region shown in (**a**) was masked when calculating the average number of genetic variants in (**b, c**). Data is plotted as mean ± standard deviation (sd). Independent t test. **d** Graphs showing RT-qPCR analysis of selected genes (highlighted red in **a**) in embryos (E10.5), placentas (E10.5) and/or adult testes from C57Bl/6J (black bars), 129P2Ola/Hsd (grey bars) and phenotypically-normal *Mtrr*$^{gt/gt}$ (blue bars) mice. **e** Data indicating RNA expression of specific groups of transposable elements as determined by RT-qPCR in C57Bl/6J tissue (black bars) and *Mtrr*$^{gt/gt}$ tissue (blue bars: phenotypically normal; white bars: severely affected (SA)). Adult liver and placenta at E10.5 were assessed. Data from RT-qPCR analyses in **d**, **e** are shown as mean ± sd and relative to C57Bl/6J tissue levels (normalized to 1). N=5-6 embryos, placentas or livers/ experimental group. One-way ANOVA with Dunnett's multiple comparison test, *p<0.05; **p<0.01.

Mtrr$^{gt/gt}$ mice (Fig. 1a,d). Genomic stability was further supported by the preserved repression of transposable elements[29,30] in *Mtrr*$^{gt/gt}$ tissue (Fig. 1e) despite global DNA hypomethylation caused by the *Mtrr*$^{gt/gt}$ mutation[3]. Overall, these data support genetic integrity within the *Mtrr*^{gt} model, and that phenotypic inheritance was unlikely caused by an increased frequency of de novo mutation. Therefore, focus shifted to an epigenetic mechanism.

Germline DNA methylation is altered in the *Mtrr*^{gt} model

MTRR plays a direct role in the transmission of one-carbon methyl groups for DNA methylation[3,12,14]. Therefore, germline DNA methylation was considered as a potential mediator of phenotype inheritance. Since an *Mtrr*$^{+/gt}$ female or male can initiate TEI (Supplementary Fig. 1b)[3]

and due to the experimental tractability of male gametes, we focussed our analysis on sperm. Spermatogenesis and male fertility are normal in *Mtrr*$^{+/+}$, *Mtrr*$^{+/gt}$, and *Mtrr*$^{gt/gt}$ males[31]. Mature spermatozoa were collected from caudal epididymides of C57Bl/6J, *Mtrr*$^{+/+}$, *Mtrr*$^{+/gt}$ and *Mtrr*$^{gt/gt}$ mice (Supplementary Fig. 1c,e,f) and sperm purity was confirmed by assessing imprinted regions of known methylation status via bisulfite pyrosequencing (Supplementary Fig. 3a). Global 5-methylcytosine (5mC) and 5-hydroxymethylcytosine (5hmC) levels were consistent across all *Mtrr* genotypes relative to C57Bl/6J controls as determined by mass spectrometry (Fig. 2a).

To analyse genome-wide distribution of sperm DNA methylation, methylated DNA immunoprecipitation followed by sequencing (MeDIP-seq) was performed. This approach allowed the unbiased detection of locus-specific changes in DNA methylation by identifying clusters of differentially methylated cytosines, thus reducing the potential impact of single-nucleotide variants[4,32]. MeDIP libraries of sperm DNA were prepared using eight males each from C57Bl/6J, *Mtrr*$^{+/+}$, *Mtrr*$^{+/gt}$ and *Mtrr*$^{gt/gt}$ genotypes (Supplementary Fig. 3b). Sequencing generated 179 million paired-end mappable reads on average per group (C57Bl/6J: 164 million reads; *Mtrr*$^{+/+}$: 172 million reads; *Mtrr*$^{+/gt}$: 203 million reads; *Mtrr*$^{gt/gt}$: 179 million reads). Using MEDIPS package[33], each *Mtrr* genotype was independently compared to C57Bl/6J controls. Loci of >500 bp with a methylation change of >1.5-fold and $p<0.01$ were defined as DMRs. The number of sperm DMRs identified increased with the severity of *Mtrr* genotype: 91 DMRs in *Mtrr*$^{+/+}$ males, 203 DMRs in *Mtrr*$^{+/gt}$ males and 599 DMRs in *Mtrr*$^{gt/gt}$ males (Fig. 2b,c). The presence of DMRs in sperm from *Mtrr*$^{+/+}$ males indicated a parental effect of the *Mtrr*^{gt} allele on offspring germline methylome since *Mtrr*$^{+/+}$ males derive from *Mtrr*$^{+/gt}$ intercrosses (Supplementary Fig. 1**e**). Hypo- and hypermethylated regions were identified in each *Mtrr* genotype when compared to C57Bl/6J controls (Fig. 2c), consistent with earlier findings in placentas[3]. These data suggested that the *Mtrr*^{gt} allele was sufficient to dysregulate sperm DNA methylation.

To ensure the robustness and reliability of the MeDIP-seq data, we randomly selected hyper- and hypomethylated DMRs to validate using bisulfite pyrosequencing. Sperm DNA from C57Bl/6J, *Mtrr*$^{+/+}$, *Mtrr*$^{+/gt}$ and *Mtrr*$^{gt/gt}$ males was assessed (N=8 males/group: four sperm samples from MeDIP-seq experiment plus four independent samples). DMRs were validated in the *Mtrr* genotype in which they were identified (Figs. 2d, 3, Supplementary Fig. 4). The overall validation

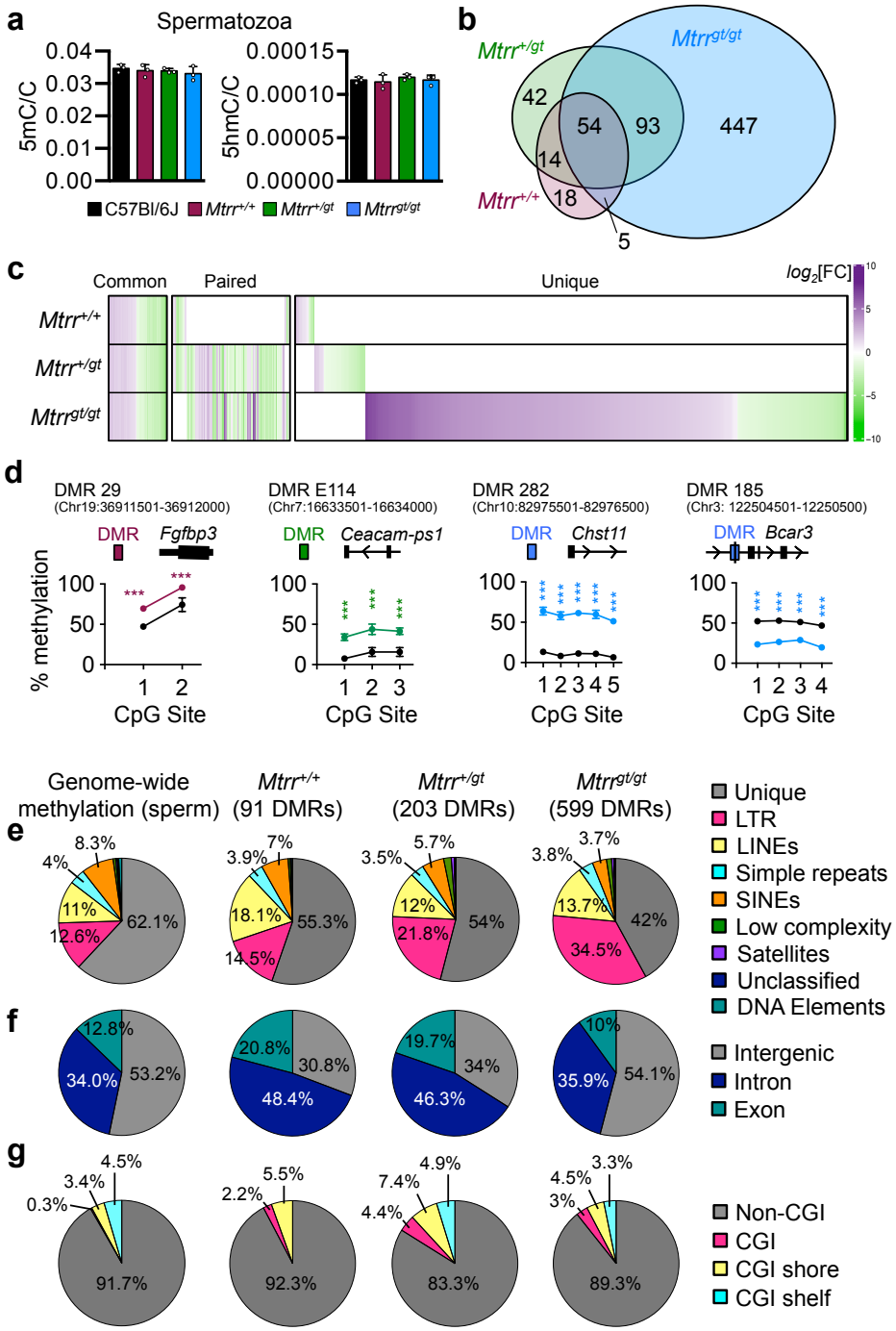


Figure 2. **Characterization of differential DNA methylation in spermatozoa from *Mtrr*^{gt} mouse line. a** Global 5-methylcytosine (5mC) and 5-hydroxymethylcytosine (5hmC) in spermatozoa from C57Bl/6J (black bars), wildtype (*Mtrr*$^{+/+}$; purple bars), *Mtrr*$^{+/gt}$ (green bars) and *Mtrr*$^{gt/gt}$ (blue bars) adult males (N=9 males/genotype analysed in 3 pools/genotype) as assessed by mass spectrometry. Data is presented as ratio of methylated cytosines per genomic cytosines (mean ± standard deviation (sd)). One-way ANOVA. **b-c** Methylated DNA immunoprecipitation followed by sequencing (MeDIP-seq) of spermatozoa DNA from *Mtrr*$^{+/+}$, *Mtrr*$^{+/gt}$ and *Mtrr*$^{gt/gt}$ males relative to C57Bl/6J controls was performed to determine differentially methylated regions (DMRs). N=8 males/group. **b** An intersectional analysis of DMRs by *Mtrr* genotype. **c** A heat map plotting log_2FoldChange of DNA methylation in spermatozoa from *Mtrr*$^{+/+}$, *Mtrr*$^{+/gt}$ and *Mtrr*$^{gt/gt}$ males compared to C57Bl/6J males ($p<0.05$). DMRs that were common between all *Mtrr* genotypes, paired between two *Mtrr* genotypes, or unique to a single *Mtrr* genotype when compared to C57Bl/6J controls are shown. **d** Examples of sperm DMRs identified via MeDIP-seq and validated by bisulfite pyrosequencing from wildtype (*Mtrr*$^{+/+}$) (purple circles), *Mtrr*$^{+/gt}$ (green circles), or *Mtrr*$^{gt/gt}$ (blue circles) males compared to C57Bl/6J sperm (black circles) N=8 males/group including four males from the MeDIP-seq analysis and four independent males. Data is shown as percentage methylation at each CpG site assessed (mean ± sd). Schematic of each DMR is

(Fig. 2 continued) is indicated in relation to the closest gene. Two-way ANOVA with Sidak's multiple comparisons test performed. **p<0.01, ***p<0.0001. See also Fig. 3 and Supplementary Fig. 4. **e-g** Relative distribution of methylated regions identified via MeDIP-seq in C57Bl/6J sperm (background methylome) and sperm DMRs from *Mtrr*$^{+/+}$, *Mtrr*$^{+/gt}$ and *Mtrr*$^{gt/gt}$ males among **e** unique sequences and repetitive elements, **f** coding and non-coding regions, and **g** CpG islands (CGIs), shores and shelves. **e** and **g**, Chi-squared test; **f**, Two-way ANOVA with Dunnett's multiple comparison test.

---

rate was 94.1% in hypomethylated DMRs and 58.3% in hypermethylated DMRs (Supplementary Table 3) and indicated a high degree of corroboration between techniques. The majority of DMRs that did not validate showed extensive methylation (>80% CpG methylation) in C57Bl/6J sperm and were identified as hypermethylated in the MeDIP-seq experiment (Supplementary Fig. 4). This might reflect some false positives in line with another study[4].

For most DMRs assessed, methylation change was consistent across all CpG sites and the absolute change in CpG methylation ranged from 10 to 80% of control levels (Figs. 2d, 3, Supplementary Fig. 4). Within each genotypic group, a high degree of inter-individual consistency of methylation change was also observed. Therefore, we conclude that the *Mtrr*^{gt} mutation, or parental exposure to it as in *Mtrr*$^{+/+}$ males, is sufficient to lead to distinct DNA methylation changes in sperm.

Most DMRs associate with metabolic dysregulation not genetic effects

A proportion of the DMRs were located within the region around the gene-trap insertion site in *Mtrr*$^{+/gt}$ and *Mtrr*$^{gt/gt}$ males (Fig. 1a, Supplementary Fig. 5b,c), consistent with *Mtrr*$^{gt/gt}$ liver[34] and suggesting that the gene-trap or underlying 129P2Ola/Hsd sequence might epigenetically dysregulate the surrounding region. However, comparison of the MeDIP-seq and WGS data sets revealed that genetic variation did not influence DMR calling to a great extent since only a small proportion (2.8-5.5%) of these DMRs contained one or more SNP. Eight DMRs overlapped with known enhancers (Supplementary Table 4), none of which associated with promoters[35] containing a genetic variant. Outside of the *Mtrr* genomic region, 54 DMRs were common to *Mtrr*$^{+/+}$, *Mtrr*$^{+/gt}$ and *Mtrr*$^{gt/gt}$ males (Fig. 2b,c, Supplementary Table 5) and were primarily located in distinct chromosomal clusters (Supplementary Fig. 5a-d). These data implicate epigenetic hotspots or

underlying genetic effects. However, beyond a polymorphic duplication on Chr19 in the C57Bl/6J strain[36] that accounted for a minor number of DMRs (2.5-15.8% of DMRs), no DMRs overlapped with an SV or were located <1 kb of an SV. Once potential genetic effects were accounted for, the majority of sperm DMRs in *Mtrr*$^{+/+}$ and *Mtrr*$^{+/gt}$ males (76/91 DMRs and 142/203 DMRs, respectively), and a proportion of sperm DMRs in *Mtrr*$^{gt/gt}$ males (174/599 DMRs) were attributed to the long-term metabolic consequences of the *Mtrr*^{gt} mutation.

Sperm DMR genomic distribution and potential regulatory function

DMR distribution was determined to explore regional susceptibility of the sperm methylome to the effects of the *Mtrr*^{gt} allele. First, the sperm 'background methylome' was established to resolve the expected genome-wide distribution of CpG methylation (see Methods). By comparing the regional distribution of sperm DMRs to the background methylome, we revealed that DMRs in all *Mtrr* genotypes were not significantly enriched in repetitive regions (Fig. 2e). However, sperm DMRs in *Mtrr*$^{+/+}$ and *Mtrr*$^{+/gt}$ males were over-represented in introns and exons, and under-represented in intergenic regions ($p<0.0003$, Chi-squared test; Fig. 2f). This was not the case for *Mtrr*$^{gt/gt}$ males since DMRs were proportionately distributed among most genomic regions (Fig. 2f). While the majority of sperm DMRs were located within CpG deserts, a proportion of DMRs from *Mtrr*$^{+/gt}$ and *Mtrr*$^{gt/gt}$ males were enriched in CpG islands ($p<0.0014$, Chi-squared; Fig. 2g), which has implications for gene regulatory control. Lastly, when considering only the subset of common DMRs shared by all *Mtrr* genotypes, a similar genomic distribution to *Mtrr*$^{+/+}$ and *Mtrr*$^{+/gt}$ males was observed (Supplementary Fig. 5d,e).

During sperm maturation, histones are replaced by protamines[37]. However, ~1% of histone-containing nucleosomes are retained[38] providing scope for epigenetic inheritance[39]. Nucleosome retention occurs primarily at promoters of developmentally-regulated genes and gene-poor repeat regions, though regional distribution and frequency differs between reports[40,41]. To determine whether sperm DMRs in *Mtrr* males were enriched in known sites of nucleosome retention, we utilized the MNase-seq dataset from C57Bl/6J spermatozoa in Erkek *et al*.[41]. First, by randomly selecting 10,000 regions of 500 bp as a proxy for DMRs, we determined that the expected frequency of DMR overlap with sites of nucleosome retention was 1.94%. Crucially, we observed

that 14.5-34.1% of DMRs identified in sperm from *Mtrr*$^{+/+}$, *Mtrr*$^{+/gt}$ and *Mtrr*$^{gt/gt}$ males were located in nucleosome retention regions (Table 1), indicating a significant enrichment ($p<0.0001$, binomial test). Therefore, these DMRs represented candidate regions for epigenetic inheritance.

Table 1 Sperm DMRs per male genotype in nucleosome retention or reprogramming resistant regions.

Number of DMRs (% of total):	*Mtrr*$^{+/+}$	*Mtrr*$^{+/gt}$	*Mtrr*$^{gt/gt}$
Total	91 (100%)	203 (100%)	599 (100%)
Regions of nucleosome retention[41]	31 (34.1%)	57 (28.1%)	87 (14.5%)
Reprogramming resistant regions in pre-implantation embryo[47]	37 (40.7%)	96 (47.3%)	325 (54.3%)
Reprogramming resistant regions in germline[46]	2 (2.2%)	5 (2.5%)	23 (3.8%)
Reprogramming resistant regions in pre-implantation embryo[47] & germline[46]	2 (2.2%)	4 (2.0%)	16 (2.7%)
Regions of nucleosome retention[41] & resistant to reprogramming in pre-implantation embryo[47]	19 (20.9%)	36 (17.7%)	49 (8.9%)
Regions of nucleosome retention[41] & resistant to reprogramming in germline[46]	1 (1.1%)	2 (1.0%)	4 (0.67%)

To better understand the normal epigenetic signatures within regions identified as sperm DMRs and to predict a potential gene regulatory role, mean enrichment for histone modifications and/or chromatin accessibility in mouse spermatozoa[42], epiblast and extraembryonic ectoderm at E6.5[43] was determined using published ChIP-seq and ATAC-seq data sets. All DMRs, except those surrounding the *Mtrr* gene-trapped site, were analysed (N=379 DMRs from all *Mtrr* genotypes combined) alongside 379 randomly selected regions representing the 'baseline genome' (see Methods). Compared to the sperm baseline genome, the majority of our DMRs were likely to associate with a closed chromatin state due to collective enrichment for protamine 1 (PRM1) and repressive histone mark H3K9me3, but not active histone marks (e.g., H3K4me1, H3K27ac) or Tn5 transposase sensitive sites (THSS)[42] (Supplementary Fig. 6a,b). This finding reinforces heterogeneity of DMR association with retained nucleosomes[41] or protamines (Table 1). In contrast, the DMRs were more likely in an open chromatin conformation state in epiblast and extraembryonic ectoderm at E6.5 based on collective enrichment for THSS when compared to

tissue-specific baseline genomes[43] (Supplementary Fig. 6c). Therefore, it is possible that the genomic regions identified as DMRs in *Mtrr* sperm have a regulatory role during development.

Some DMRs were located in regions of reprogramming resistance

DNA methylation is largely reprogrammed during pre-implantation development and in the developing germline[44,45] to 'reset' the methylome between each generation. Recently, several loci were identified as 'reprogramming resistant'[46-48] and thus, are implicated in epigenetic inheritance. Using published methylome data sets[46,47], we determined that 40.7-54.3% of sperm DMRs across all *Mtrr* genotypes fell within loci resistant to pre-implantation reprogramming (Table 1). Sixteen of these DMRs were common among $Mtrr^{+/+}$, $Mtrr^{+/gt}$ and $Mtrr^{gt/gt}$ males. Fewer DMRs correlated with regions resistant to germline reprogramming (2.2-3.8% of DMRs/*Mtrr* genotype; Table 1) or both pre-implantation and germline reprogramming (2.0-2.7% of DMRs/*Mtrr* genotype; Table 1). Only one DMR located in a region resistant to germline reprogramming was common to all *Mtrr* genotypes. Interestingly, several DMRs were characterized as regions of reprogramming resistance[46,47] and nucleosome retention[41] (Table 1). Overall, differential methylation of these key regions in sperm of $Mtrr^{gt}$ males might have important implications for epigenetic inheritance.

Sperm DMRs are reprogrammed in wildtype F1 and F2 generations

TEI in the $Mtrr^{gt}$ model occurs via the maternal grandparental lineage[3] (Supplementary Fig. 1b). To determine the heritability of germline DMRs, bisulfite pyrosequencing was used to validate ten sperm DMRs from F0 $Mtrr^{+/gt}$ males in tissue of wildtype F1 and F2 progeny. The advantage of assessing inheritance of DNA methylation in F0 sperm rather than in F0 oocytes was that potential confounding effects of the F0 uterine environment could be avoided. Candidate DMRs were hyper- or hypomethylated, and localized to regions of reprogramming resistance and/or to intra- or intergenic regions (Supplementary Table 6). In general, all DMRs tested lost their differential methylation in wildtype F1 and F2 embryos and placentas at E10.5, and showed DNA methylation patterns similar to C57Bl/6J tissue (Fig. 3). This result occurred even when wildtype F2 conceptuses displayed congenital malformations (Fig. 3). DMRs were also assessed in $Mtrr^{gt/gt}$

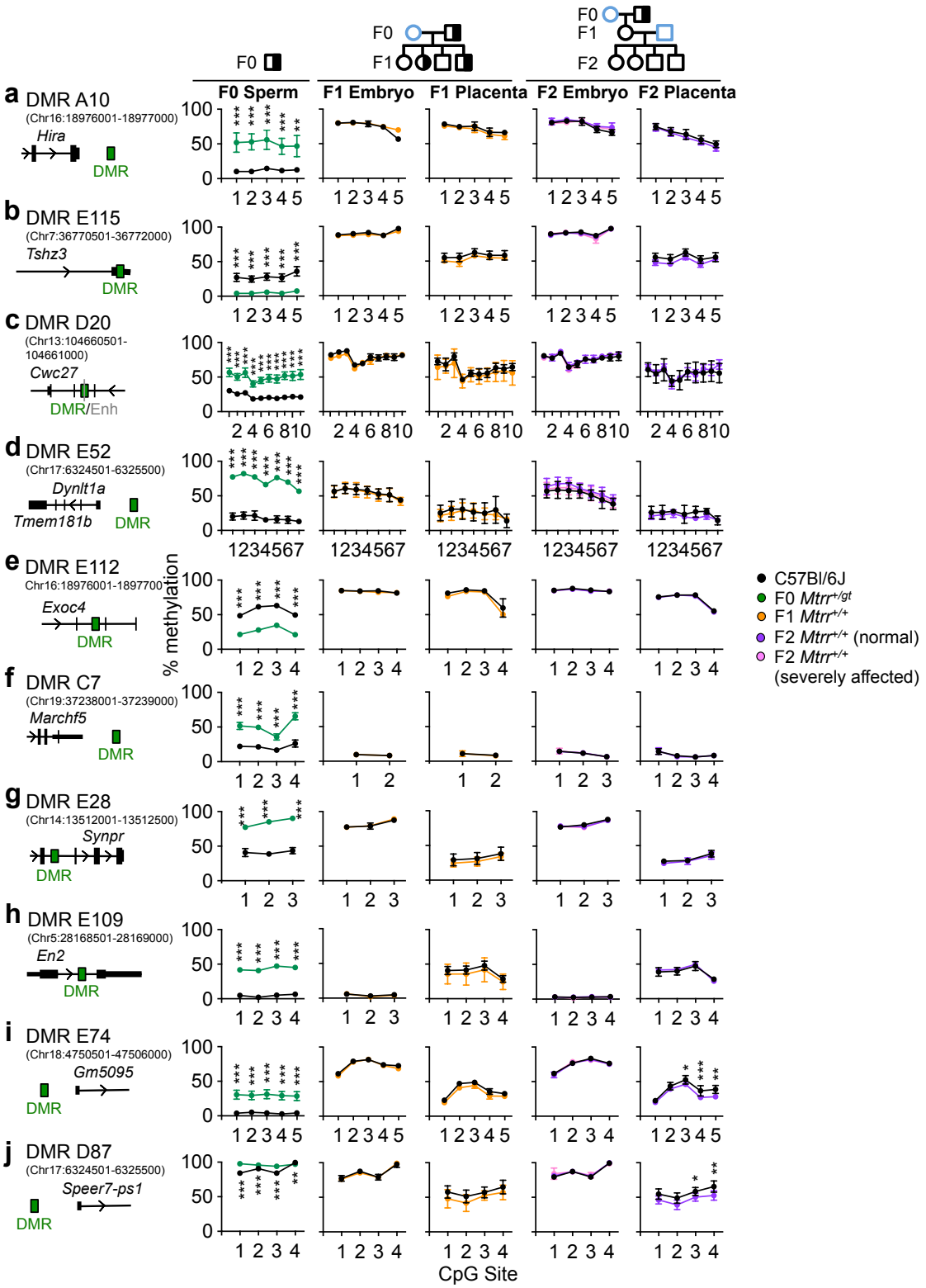


Figure 3. **Sperm DMRs are reprogrammed in somatic tissue of F1-F2 wildtype generations.** CpG methylation at specific sperm differentially methylated regions (DMRs) identified in F0 *Mtrr*$^{+/gt}$ males was assessed in the F1 and F2 wildtype embryos and placentas at E10.5. Pedigrees indicate specific mating scheme. **a-j** Schematic drawings of each DMR (green rectangles) assessed indicate its relationship to the closest gene (black) and enhancer (grey line) are followed by graphs showing the average percentage of methylation at individual CpGs for the corresponding DMR as determined by bisulfite pyrosequencing. In each graph, methylation was assessed in sperm from F0 *Mtrr*$^{+/gt}$ males (green circles), phenotypically normal F1 wildtype (*Mtrr*$^{+/+}$) embryos and placentas at E10.5 (orange circles), and phenotypically normal (purple circles) or severely affected (pink circles) F2 wildtype embryos and placentas at E10.5. C57Bl/6J (black circles) are shown as controls. Sperm: N=6-8 males/experimental group; Embryos: N=4-5 embryos/experimental group; Placentas: N=4-8 C57Bl/6J

(Fig. 3 continued) placentas, N=7-8 F1 or F2 wildtype placentas. Data is shown as mean ± sd for each CpG site. 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. Pedigree legend: circle, female; square, male; blue outline, C57Bl/6J control; black outline, *Mtrr*^{gt} mouse line; white filled, *Mtrr*$^{+/+}$; half-white/half-black filled, *Mtrr*$^{+/gt}$.

conceptuses at E10.5 to determine whether these regions were capable of differential methylation outside of the germline. In a manner similar to sperm from *Mtrr*$^{+/gt}$ males (Fig. 3), seven out of 10 DMRs were hypermethylated in *Mtrr*$^{gt/gt}$ embryos and/or placentas compared to control conceptuses (Supplementary Fig. 7a-j). In this case, it was unclear whether DNA methylation in these regions resisted epigenetic reprogramming or was erased and abnormally re-established/ maintained due to intrinsic *Mtrr*$^{gt/gt}$ homozygosity. Overall, altered patterns of DNA methylation in sperm of *Mtrr*$^{+/gt}$ males were not evident in somatic tissue of wildtype progeny and grandprogeny. This result was reminiscent of mouse models of parental exposure to environmental stressors (e.g., maternal under-nutrition[4], paternal folate deficiency[49], or paternal cigarette smoking[50]), which induced sperm DMRs associated with phenotypes in the direct offspring even though the DMRs were resolved in offspring somatic tissue. As a result, other epigenetic mechanisms (e.g., germline RNA content and/or histone modifications) with 5mC are implicated in phenotypic inheritance.

Transcriptional memory of sperm DMRs

A previous study in mice suggests that sperm DMRs can associate with perturbed transcription in offspring even when DNA methylation is re-established to normal levels[4]. To assess whether transcriptional memory occurred in the *Mtrr*$^{+/gt}$ maternal grandfather pedigree, expression of six genes located in or near sperm DMRs from F0 *Mtrr*$^{+/gt}$ males (Supplementary Table 6) was assessed in F1 and F2 wildtype individuals. While these genes displayed normal expression in F1 tissues (Fig. 4a-c), *Hira* (histone chaperone), *Cwc27* (spliceosome-associated protein), and *Tshz3* (transcription factor) were misexpressed in F2 wildtype embryos or adult livers compared to C57Bl/6J controls (Fig. 4d-f). This result might reflect transcriptional memory of the associated sperm DMR or wider epigenetic dysregulation in sperm of the F0 *Mtrr*$^{+/gt}$ males.

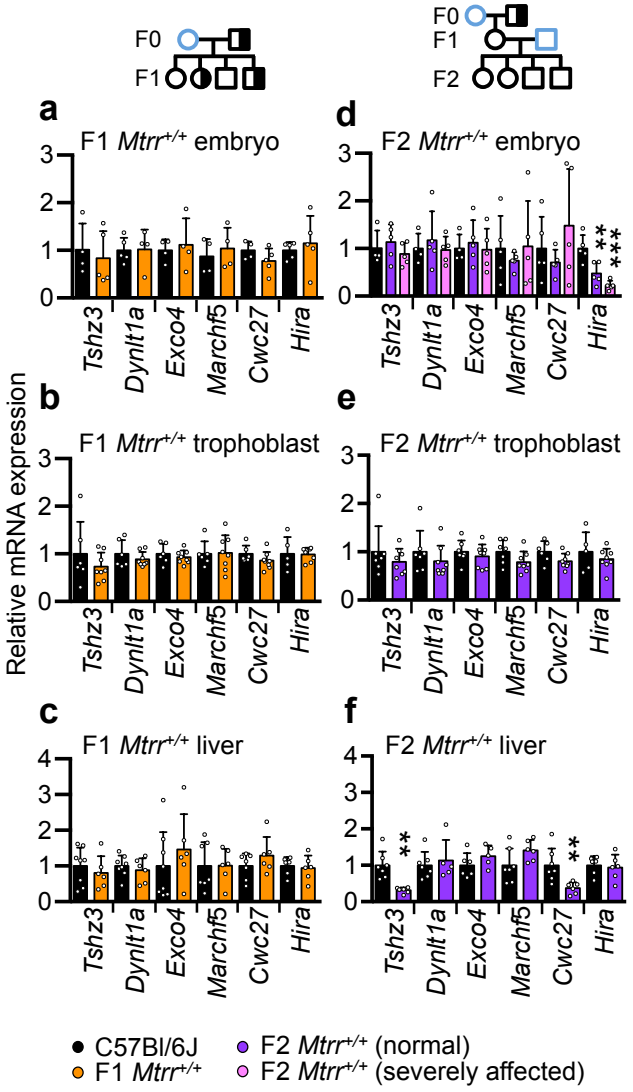


Figure 4. **Transcriptional memory of some DMR-associated genes in F1 and F2 wildtype somatic tissue. a-f** RT-qPCR analysis of mRNA expression of genes located in close proximity to sperm DMRs in **a, d** embryos at E10.5, **b, e** placental trophoblast at E10.5, and **c, f** adult livers of wildtype (*Mtrr*$^{+/+}$) F1 and F2 progeny. Tissue from phenotypically normal F1 wildtype individuals (orange bars), and phenotypically normal (purple bars) or severely affected (pink bars) F2 wildtype individuals derived from an F0 *Mtrr*$^{+/gt}$ male was assessed. C57Bl/6J control tissues (black bars) were assessed as controls. Pedigrees indicate specific mating scheme assessed. Data is plotted as mean ± standard deviation, and is presented as relative expression to C57Bl/6J levels (normalized to 1). N=4-7 individuals per group. Independent t tests or one-way ANOVA with Dunnett's multiple comparisons tests were performed. **p<0.01, ***p<0.001. Each gene was associated with the following sperm DMR shown in Fig. 3: *Tshz3*, DMR E115; *Dynlt1a*, DMR E52; *Exoc4,* DMR E112; *March5*, DMR C7; *Cwc27*, DMR D20; *Hira*, DMR A10. Pedigree legend: circle, female; square, male; blue outline, C57Bl/6J control; black outline, *Mtrr*^{gt} mouse line; white filled, *Mtrr*$^{+/+}$; half-white/half-black filled, *Mtrr*$^{+/gt}$.

To further predict whether the *Hira*, *Cwc27*, and *Tshz3* DMRs demarcate gene regulatory regions, their specific genetic and epigenetic characteristics beyond CpG methylation were considered. Genomically, the DMRs were located intragenically (*Cwc27* and *Tshz3* DMRs) or within 6 kb downstream of the gene (*Hira* DMR; Fig. 5a, Supplementary Figs. 8-9). Furthermore, *Cwc27* DMR overlapped with a known enhancer (Supplementary Fig. 8) while *Hira* and *Tshz3* DMRs overlapped with CpG islands[51] (Fig. 5a, Supplementary Fig. 9). Next, we assessed the three DMRs for enrichment of specific histone modifications during normal development using published ChIP-seq data sets[35] in wildtype embryonic stem cells (ESCs) and trophoblast stem cells (TSCs). While histone marks were largely absent at all three DMRs in TSCs, enrichment for one or more repressive histone mark (e.g., H3K4me3 and/or H3K9me3) at these DMRs was apparent in ESCs (Fig. 5a, Supplementary Fig. 8-9). Altogether, DMRs identified in sperm of *Mtrr*$^{+/gt}$ males highlight

regions that might be important for transcriptional regulation in the early conceptus by other epigenetic mechanisms. Therefore, future analyses of broader epigenetic marks at these sites are required in the *Mtrr*^{gt} mouse line.

HIRA as a potential biomarker of maternal phenotypic inheritance

The importance of the *Hira* DMR in phenotypic inheritance was further considered based on its known resistance to germline reprogramming[46], and its potential function in gene regulation (Fig. 5a) and transcriptional memory (Fig. 4d). HIRA is a histone H3.3 chaperone, which lends itself well to a role in epigenetic inheritance given its broad functionality in transcriptional regulation[26], maintenance of chromatin structure in the developing oocyte[25] and the male pronucleus after fertilization[27], and in rRNA transcription and ribosome function[27]. *Hira*$^{-/-}$ mice[52] and the *Mtrr*^{gt} mouse line[3] display similar phenotypes including growth defects, congenital malformations, and embryonic lethality by E10.5. Furthermore, *Mtrr*^{gt} genotypic severity correlated with the degree of hypermethylation in the *Hira* DMR in sperm (Fig. 5b,c) suggesting that the *Hira* DMR is responsive to alterations in folate metabolism.

The potential regulatory legacy of the *Hira* DMR in sperm was further explored in the *Mtrr*$^{+/gt}$ maternal grandfather pedigree. The *Hira* DMR, which was 6 kb downstream of the *Hira* gene and overlaps with a CpG island and CTCF binding site in ESCs and TSCs (Fig. 5a), was substantially hypermethylated in sperm of F0 *Mtrr*$^{+/gt}$ males compared to controls (39.0 ± 4.1% more methylated CpGs per CpG site assessed; Fig. 5b,c). As with the other sperm DMRs assessed (Fig. 3), the *Hira* DMR was reprogrammed in F1-F3 wildtype embryos and placentas at E10.5 (Fig. 3a, Supplementary Fig. 10a). While we originally assessed *Hira* mRNA in the F1-F2 generations in the maternal grandfather pedigree (Fig. 4), further analysis revealed that *Hira* isoforms (mRNA and long non-coding RNA (lncRNA 209); Fig. 5a) were differentially regulated at E10.5 (Fig. 6a-c). *Hira* lncRNA function is unknown, though lncRNA-based mechanisms often control cell fates during development by influencing nuclear organization and transcriptional regulation[53]. Notably, this pattern of RNA expression was associated with generational patterns of phenotypic inheritance. For instance, we observed down-regulation of *Hira* mRNA in F2-F3 wildtype embryos and not F1 wildtype embryos at E10.5 (Fig. 6a-c). Conversely, significant up-regulation of *Hira* lncRNA

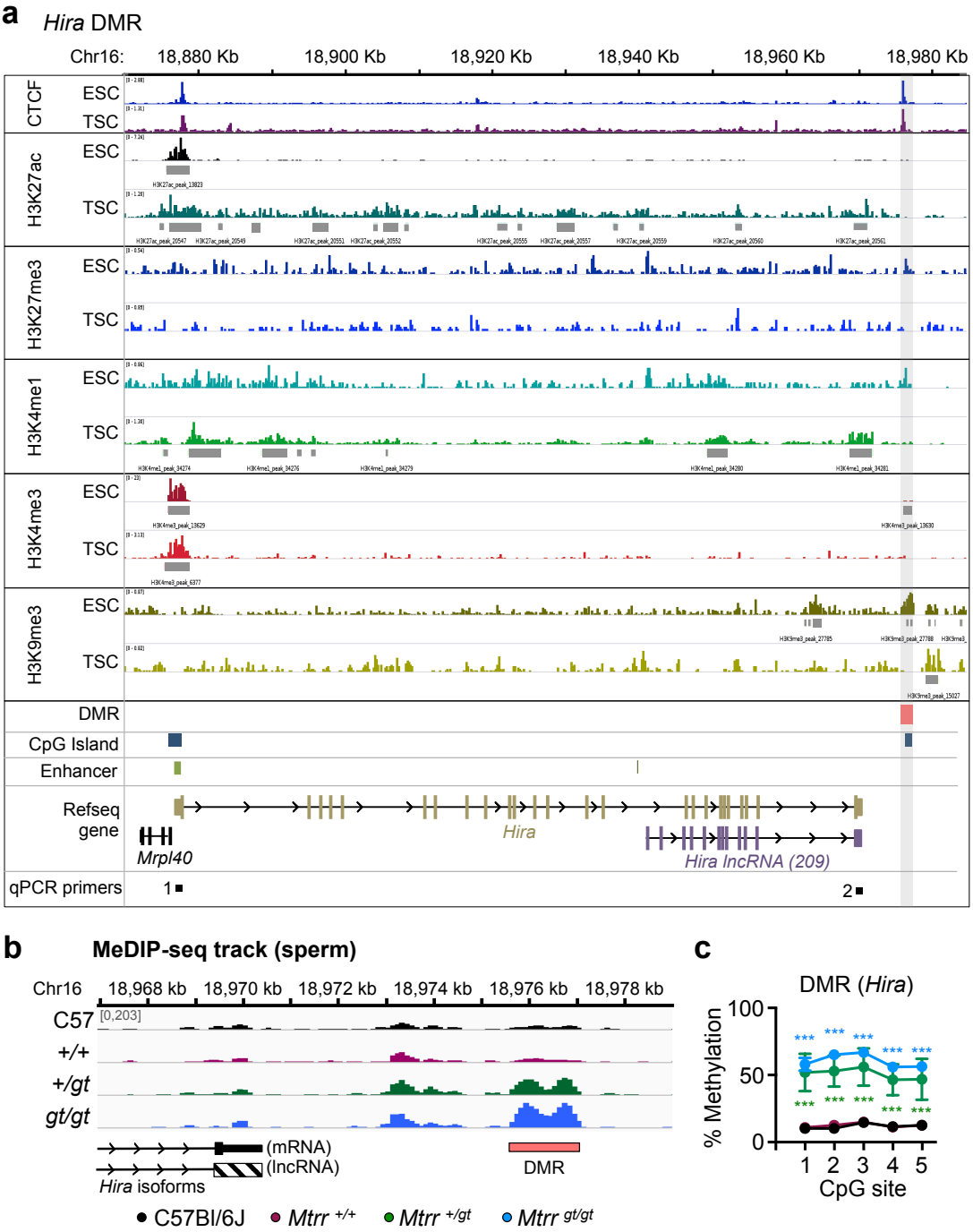


Figure 5. **Epigenetic characteristics of *Hira* DMR in mice. a** Enrichment of DNA binding proteins (CTCF) and histone modifications (H3K27ac, H3K27me3, H3K4me1, H3K4me3, H3K9me3) in the *Hira* locus on mouse chromosome (Chr) 16 using published data sets[35] of chromatin immunoprecipitation followed by sequencing (ChIP-seq) analyses in cultured wildtype embryonic stem cells (ESCs) and trophoblast stem cells (TSCs). Grey rectangles indicate enrichment peak calling for each histone modification. Red rectangle and grey shading indicated the *Hira* differentially methylated region (DMR) identified in sperm of *Mtrr*$^{+/gt}$ and *Mtrr*$^{gt/gt}$ males. Blue rectangles indicate CpG islands. Green rectangles indicate enhancers. Schematics of protein encoding (brown) and long non-coding RNA (lncRNA) encoding (purple) *Hira* isoforms are shown. Region of RT-qPCR primer sets 1 and 2 are also indicated. **b** Partial schematic drawing of *Hira* transcripts and *Hira* DMR (red rectangle) in relation to MeDIP-seq reads in sperm from C57Bl/6J (black), wildtype (*Mtrr*$^{+/+}$; purple), *Mtrr*$^{+/gt}$ (green) and *Mtrr*$^{gt/gt}$ (blue) males. N=8 males per group. **c** The average percentage methylation at individual CpG sites (mean ± standard deviation) in the *Hira* DMR in sperm from C57Bl/6J (black circles; N=4 males), *Mtrr*$^{+/+}$ (purple circles; N=4 males), *Mtrr*$^{+/gt}$ (green circles; N=8 males) and *Mtrr*$^{gt/gt}$ (blue circles; N=8 males) males. Two-way ANOVA with Sidak's multiple comparisons test performed on mean methylation per CpG site. ***p<0.001.

expression was apparent only in F1 wildtype embryos at E10.5 (Fig. 6a-c). This expression pattern was embryo-specific since the corresponding placentas showed normal *Hira* transcript levels in each generation assessed compared to controls (Fig. 6a-b). Since phenotypes at E10.5 were apparent in F2 generation onwards, yet were absent in the F1 generation of the *Mtrr*$^{+/gt}$ maternal grandfather pedigree[3], embryo-specific *Hira* RNA misexpression reflected the pattern of phenotypic inheritance.

To further investigate a potential link between *Hira* expression and phenotypic inheritance, we analysed wildtype F1-F3 conceptuses at E10.5 derived from F0 *Mtrr*$^{+/gt}$ females (Supplementary Fig. 1b), of which all generations, including the F1 generation, display a broad spectrum of developmental phenotypes[3]. Supporting our hypothesis, *Hira* mRNA expression was down-regulated in F1-F3 wildtype embryos derived from an F0 *Mtrr*$^{+/gt}$ female compared to C57Bl/6J controls (Fig. 6d), thus correlating with maternal phenotypic inheritance. As expected, *Hira* lncRNA transcripts were unchanged in F1 and F3 wildtype embryos (Fig. 6d). However, *Hira* lncRNA was down-regulated in F2 wildtype embryos (Fig. 6d), which display the highest frequency of phenotypes among the three generations[3]. Regardless, these data suggested that altered *Hira* mRNA transcripts are a potential biomarker of maternal phenotypic inheritance. This was because dysregulation of *Hira* mRNA occurred only in wildtype embryos with high phenotypic risk due to their derivation from an oocyte with *Mtrr*^{gt} ancestry rather than sperm (Fig. 7).

How the *Hira* gene is regulated is unknown. *Mtrr*$^{gt/gt}$ embryos at E10.5, which demonstrate a greater phenotypic risk than the *Mtrr*$^{+/gt}$ maternal grandparental pedigrees, also displayed dyregulation of *Hira* mRNA and lncRNA expression (Supplementary Figs. 7k-m, 10d). This finding was in association with normal DNA methylation at the *Hira* DMR in *Mtrr*$^{gt/gt}$ embryos at E10.5 (Supplementary Fig. 7b), implicating additional mechanisms of epigenetic regulation. Histone methylation profiles at the *Hira* locus in normal ESCs and TSCs indicate a potential role for the *Hira* promoter and *Hira* DMR (Fig. 5a) in gene regulation that will require future investigation.

HIRA protein levels were also dysregulated in *Mtrr*$^{gt/gt}$ embryos and F2 wildtype embryos and placentas (Fig. 6e,f, Supplementary Fig. 10e,f) and not always in a manner predicted by the direction of *Hira* mRNA expression. For example, there was an increase in HIRA protein levels in F2 wildtype embryos when *Hira* mRNA was down-regulated (Fig. 6b,d-f). This discrepancy might

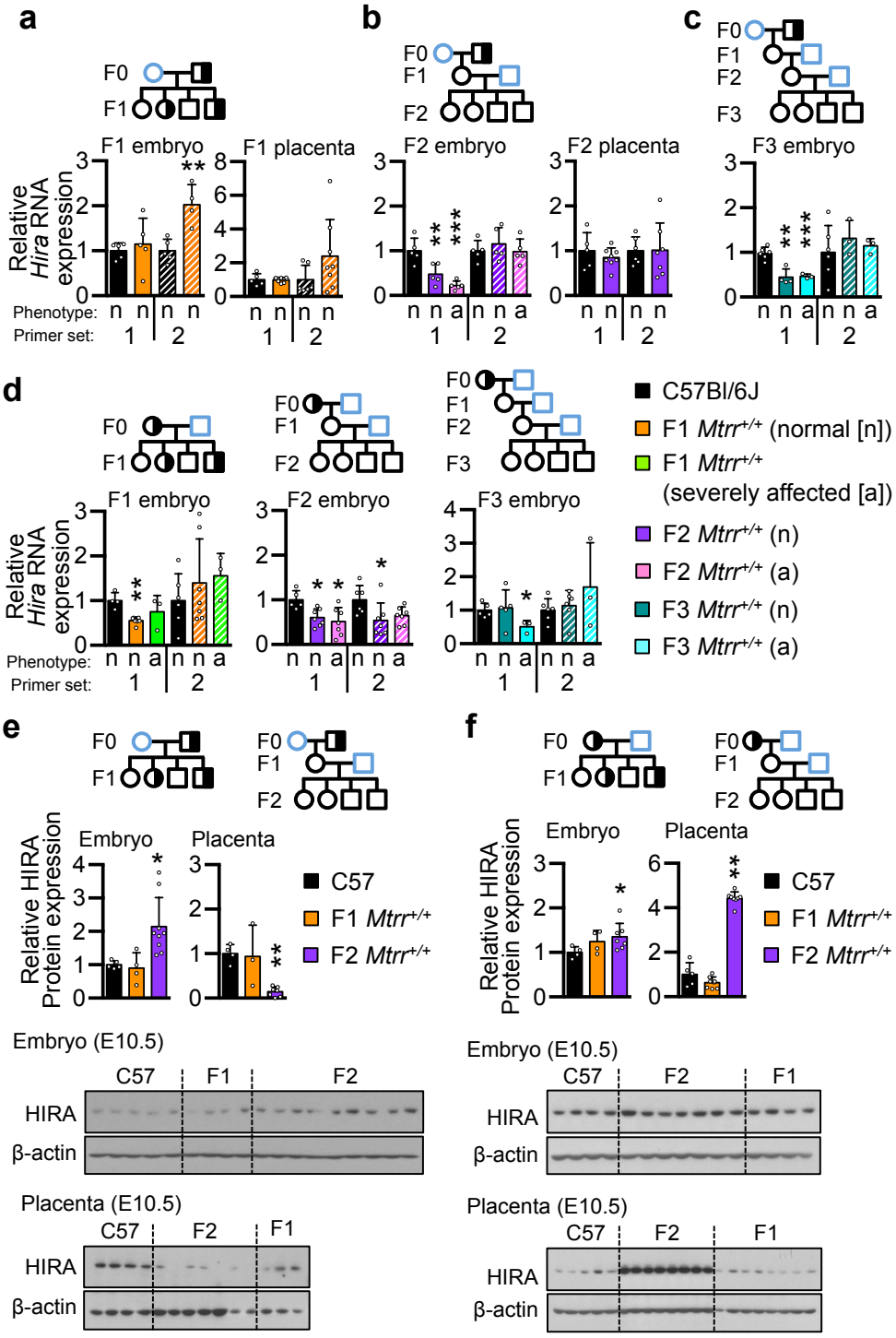


Figure 6. **Dysregulation of *Hira* RNA in F1-F3 wildtype conceptuses at E10.5 aligns with a pattern of maternal phenotypic inheritance. a-d** RT-qPCR data showing *Hira* mRNA (solid bars, primer set 1) and *Hira* lncRNA (striped bars, primer set 2) in embryos and placentas at E10.5. The following conceptuses were assessed from the indicated pedigree: **a** F1 wildtype ($Mtrr^{+/+}$) conceptuses derived from F0 $Mtrr^{+/gt}$ males (orange bars; N=4-8 normal embryos and placentas), **b** F2 wildtype conceptuses derived from F0 $Mtrr^{+/gt}$ males (purple and pink bars; N=5-8 embryos and placentas/phenotype), **c** F3 wildtype conceptuses derived from F0 $Mtrr^{+/gt}$ males (teal and turquoise bars; N=3-4 embryos/phenotype) and **d** F1 wildtype (orange bars, N=3-7 embryos/phenotype), F2 wildtype (pink and purple bars; N=5-7 embryos/phenotype, and F3 wildtype conceptuses (teal and turquoise bars; N=3-5 embryos/phenotype) derived from F0 $Mtrr^{+/gt}$ females. C57Bl/6J conceptuses (black bars; N=4-8 embryos and placenta/experiment) were controls. Phenotypically-normal (n) and severely affected (a) conceptuses were assessed. **e, f** Western blot analysis showing HIRA protein expression in embryos (N=4-9) and placentas (N=3-8) of F1 wildtype (orange bars) and F2 wildtype (purple bars) conceptuses derived from either **e** F0 $Mtrr^{+/gt}$ males or **f** F0 $Mtrr^{+/gt}$ females. C57Bl/6J embryos and placentas (black bars) were assessed as controls. Western blots

(Fig. 6 continued) showing HIRA protein with β-actin as a loading control are shown, and quantified using the background subtraction method. HIRA protein levels were first normalised to β-actin. All data was plotted as mean ± standard deviation and relative to C57Bl/6J (normalised to 1). Independent t tests or Kruskal-Wallis test with Dunn's multiple comparison test. *p<0.05, **p<0.01, ***p<0.001. Pedigree legend, circle, female; square, male; blue outline, C57Bl/6J line; black outline, *Mtrr*^{gt} mouse line; white fill, *Mtrr*$^{+/+}$; half black-half white fill, *Mtrr*$^{+/gt}$; black fill, *Mtrr*$^{gt/gt}$.

result from defective HIRA protein degradation, a drastic translational up-regulation of HIRA protein to compensate for low mRNA levels, or alternatively, negative feedback to down-regulate *Hira* mRNA due to high levels of HIRA protein in the embryo. Further work will be required to delineate whether the HIRA chaperone mediates maternal phenotypic inheritance in the *Mtrr*^{gt} model and other models of TEI.

DISCUSSION

We investigated potential mechanisms contributing to epigenetic inheritance in *Mtrr*^{gt} mice, a unique model of mammalian TEI[3]. In the *Mtrr*^{gt} model, inheritance of developmental phenotypes and epigenetic instability occurs via the maternal grandparental lineage with an F0 *Mtrr*$^{+/gt}$ male or female initiating the TEI effect[3]. Here, we assessed DNA methylation in spermatozoa to understand how the germline epigenome was affected by the *Mtrr*^{gt} allele. We chose sperm due to its experimental tractability and to avoid the confounding factors of the F0 uterine environment when assessing epigenetic inheritance via the germline. We identified several distinct DMRs in regions of predicted importance in transcriptional regulation and epigenetic inheritance including regions of nucleosome retention and reprogramming resistance. This result illustrates widespread epigenetic instability in the male germline of the *Mtrr*^{gt} model, particularly in the F0 *Mtrr*$^{+/gt}$ males of the maternal grandfather pedigree. While largely resolved in somatic tissue of subsequent wildtype generations, some germline DMRs were associated with transcriptional changes at associated loci in the F1-F3 progeny. This transcriptional memory of a germline DMR persisted for at least three generations, longer than previously reported in another model[4]. This observation indicates additional epigenetic factors beyond DNA methylation in the mechanism of TEI. Furthermore, the histone chaperone gene *Hira* emerged as a transcriptional biomarker and potential mediator of maternal phenotypic inheritance.

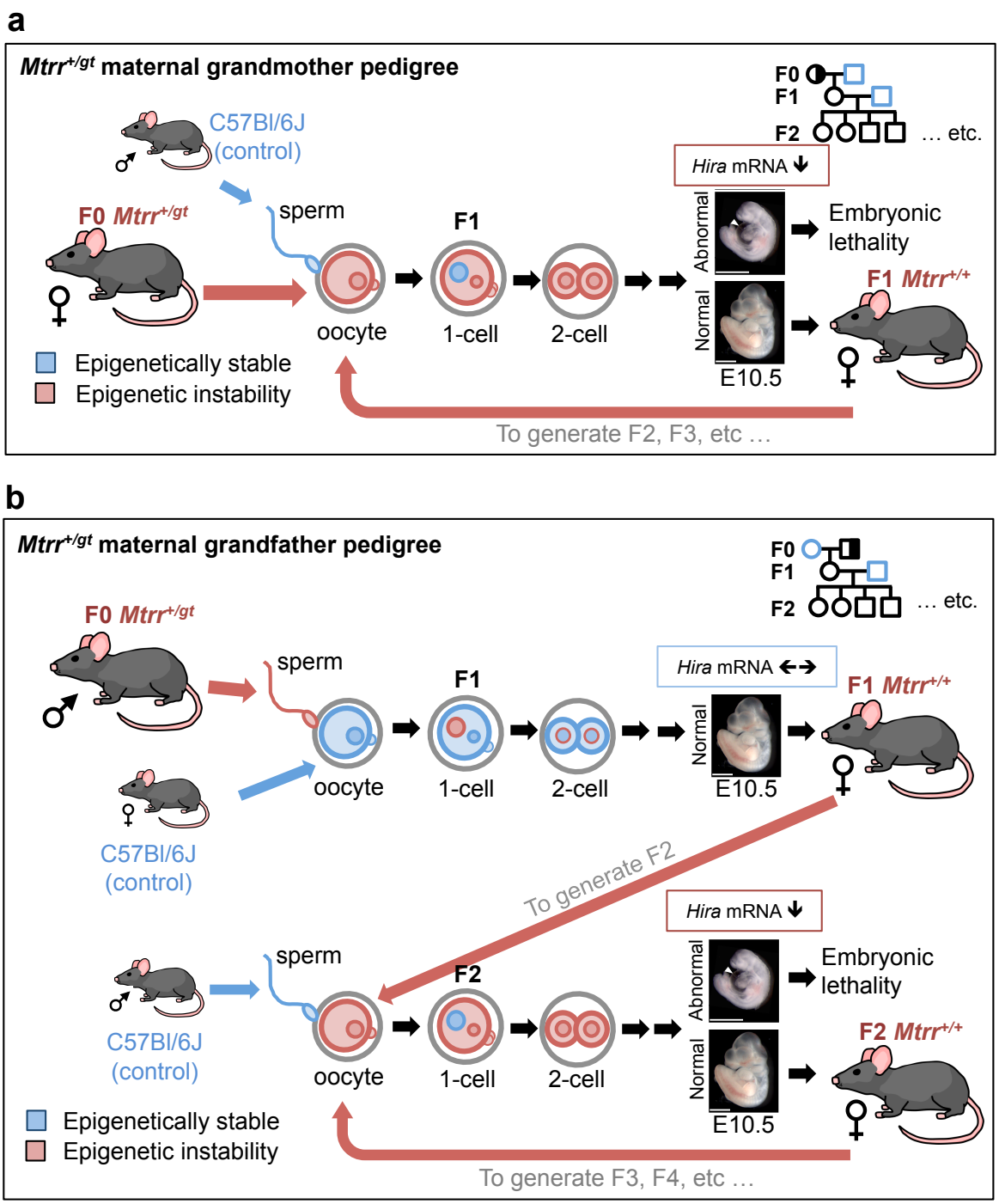


Figure 7. **Proposed model of maternal phenotypic inheritance in *Mtrr*^{gt} model.** A model suggesting how epigenetic instability generated by the *Mtrr*^{gt} mutation might be differently inherited over multiple generations depending upon whether TEI is initiated by **a** an oocyte or **b** sperm of an *Mtrr*$^{+/gt}$ individual. Trends of *Hira* mRNA expression are shown for each breeding scheme. Pedigree legend: circle, female; square, male; blue outline, C57Bl/6J control; black outline, *Mtrr*^{gt} mouse line; white filled, *Mtrr*$^{+/+}$; half-white/half-black filled, *Mtrr*$^{+/gt}$.

The extent to which genetic and epigenetic factors interact in this and other TEI models is unclear. One-carbon metabolism is involved in thymidine synthesis[54], and DNA breaks triggered by folate deficiency-induced uracil misincorporation was demonstrated in erythrocytes of splenectomised patients[55] and prostate adenoma cells[56]. However, WGS of *Mtrr*$^{gt/gt}$ embryos excluded genetic instability in the *Mtrr*^{gt} mouse line because de novo mutations occurred at an expected[28] and

comparable frequency to control embryos. The WGS data also discounted alternative phenotype-causing mutations and, when compared to our sperm methylome data set, showed that differential CpG methylation was unlikely due to underlying genetic variation in the *Mtrr*^{gt} mouse line. Therefore, the epigenetic consequences of the *Mtrr*^{gt} allele rather than genetic instability are more likely to instigate TEI in this model. Generating alternative *Mtrr* mutations and/or backcrossing the *Mtrr*^{gt} allele into a different mouse strain will further assess whether genetics plays a role in TEI.

Our data show that inheritance of sperm DMRs by offspring somatic tissue was unlikely, as have other studies[4,49,50]. Instead, somatic cell lineages might inherit germline epigenetic instability in a broader sense. For instance, despite normal DNA methylation in F1-F3 wildtype embryos and placentas (E10.5) at the genomic locations highlighted by sperm DMRs, epigenetic instability associated with gene misexpression is still evident in F1 and F2 wildtype placentas at several loci[3]. It is possible that abnormalities in the sperm epigenome of F0 *Mtrr*$^{+/gt}$ males might be reprogrammed and then stochastically and abnormally re-established/maintained in other genomic regions in wildtype offspring. This hypothesis might explain inter-individual phenotypic variability in the F2-F3 generations. However, we showed that sperm of *Mtrr*$^{+/+}$ males exhibited several DMRs that overlapped with sperm DMRs in *Mtrr*$^{+/gt}$ males (representing their fathers). Therefore, reconstruction of specific atypical F0 germline methylation patterns in the F1 wildtype germline is possible in the *Mtrr*^{gt} model. Vinclozolin toxicant exposure model of TEI shows dissimilar DMRs in spermatozoa of F1 and F3 offspring[8], which suggests that epigenetic patterns might shift as generational distance from the F0 individual increases.

Dietary folate deficiency causes differential methylation in sperm and craniofacial defects in the immediate offspring[49], though whether it leads to TEI is unknown. There was no overlap of DMRs when sperm methylomes were compared between the diet model and *Mtrr*$^{+/gt}$ males. This result disputes the existence of folate-sensitive epigenomic hotspots in sperm. However, severity of insult or technical differences (e.g., MeDIP-array[49] versus MeDIP-seq) might explain the discrepancy. Why only one *Mtrr*^{gt} allele sufficiently initiates TEI is unknown since a direct paramutation effect was not evident[3] and *Mtrr*$^{+/gt}$ mice do not display similar metabolic derangement to *Mtrr*$^{gt/gt}$ mice[3,12].

Whether specific DNA methylation patterns observed in F0 sperm of the maternal

grandfather pedigree are reconstructed[57,58] in F1 oocytes is yet-to-be determined. Extensive differences between sperm and oocytes methylomes[59] will make this difficult to resolve and yet also emphasizes that additional epigenetic mechanisms are likely involved, such as histone modifications[10], sncRNA expression[1,9] and/or changes in nucleosome composition and spacing that alter nuclear architecture[59]. Several of these mechanisms implicate regulators like HIRA[25-27] (see below). Future studies will explore the extent to which differential epigenetic marks in oocytes from F0 *Mtrr*$^{+/gt}$ females are reconstructed in oocytes of subsequent generations, though embryo transfer experiments will be required to exclude the confounding effects of the F0 uterine environment[60].

Several recent studies have assessed sncRNA expression (e.g., microRNA or tRNA fragments [tsRNA])[1,2,9] in sperm as a mechanism in direct epigenetic inheritance of disease. For example, manipulating paternal diet in mice is sufficient to alter tsRNA content in spermatozoa leading to altered expression of genes associated with MERVL elements in early F1 embryos[9] and metabolic disease in F1 adult offspring[1]. It is possible that sncRNAs in sperm from F0 *Mtrr*$^{+/gt}$ males are misexpressed and/or abnormally modified[61], and that this might contribute to TEI mechanisms. While it is currently unclear how sperm ncRNA content causes phenotypic inheritance beyond the F1 generation, genes involved in development of the primordial germ cell population within the F1 wildtype female embryo might be among those affected. Exploring sncRNA content in germ cells of the *Mtrr*^{gt} model will help to better understand this question.

Several sperm DMRs in *Mtrr*$^{+/gt}$ males (e.g., *Cwc27*, *Tshz3*, *Hira*) demonstrated transcriptional memory in F2-F3 wildtype embryos and adult liver. We focused our analysis on the *Hira* DMR due to its responsiveness to the severity of the *Mtrr*^{gt} genotype, the fact that *Hira*$^{-/-}$ embryos[52] phenocopy embryos in the *Mtrr*^{gt} model[3], and because the potential consequences of HIRA dysfunction are multifaceted and implicated in epigenetic inheritance. The HIRA histone chaperone complex regulates histone deposition/recycling required to maintain chromatin integrity in the oocyte and zygote[25,26,62]. Additionally, HIRA is implicated in rRNA transcription and ribosome assembly in the mouse zygote[27]. Therefore, dysregulation of HIRA expression and/or function in the *Mtrr*^{gt} model might alter nucleosome spacing and ribosome heterogeneity[63,64] with substantial implications for regulation of transcriptional pathways during germ cell and embryo development.

Since the F1 progeny differ phenotypically when derived from an F0 *Mtrr*$^{+/gt}$ male versus F0 *Mtrr*$^{+/gt}$ female[3], the mechanism initiating TEI potentially differs between sperm and oocytes. The extent of mechanistic overlap is not well understood. Certainly, paternally-inherited epigenetic factors[1,2,4,9,10,49,61,65] are better studied than maternally-inherited factors[60,66]. The difference in phenotypic severity in the *Mtrr*^{gt} model might relate to a cytoplasmically-inherited factor in the F1 wildtype zygote (Fig. 7). HIRA is a suitable candidate since maternal HIRA[25,67] is involved in protamine replacement by histones in the paternal pronucleus[27] thus perpetuating epigenetic instability. Therefore, in the case of the *Mtrr*^{gt} model, dysregulation of *Hira* and/or other maternal factors in oocytes might have a greater impact on epigenetic integrity in the early embryo than when dysregulated in sperm.

Overall, we show the long-term transcriptional and phenotypic impact of abnormal folate metabolism on germline DNA methylation and emphasize the complexity of epigenetic mechanisms involved in TEI. Ultimately, our data indicate the importance of normal folate metabolism in both women and men of reproductive age for healthy pregnancies in their daughters and granddaughters.

METHODS

Ethics statement

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.

Animal model

Mtrr$^{Gt(XG334)Byg}$ (MGI:3526159), referred to as *Mtrr*^{gt} mice, were generated as previously described[3]. Briefly, a *β-geo* gene-trap (gt) vector was inserted into intron 9 of the *Mtrr* gene in 129P2Ola/Hsd embryonic stem cells (ESCs). *Mtrr*^{gt} ESCs were injected into C57Bl/6J blastocysts. Upon germline transmission, the *Mtrr*^{gt} allele was backcrossed into the C57Bl/6J genetic background for at least eight generations[3]. *Mtrr*$^{+/+}$ and *Mtrr*$^{+/gt}$ mice were produced from *Mtrr*$^{+/gt}$ intercrosses (Supplementary Fig. 1e). *Mtrr*$^{gt/gt}$ mice were produced from *Mtrr*$^{gt/gt}$ intercrosses (Supplementary Fig. 1f). C57Bl/6J mice from The Jackson Laboratories (www.jaxmice.jax.org) and 129P2Ola/Hsd from Envigo (previously Harlan Laboratories [www.envigo.com]) were used as controls and were bred in house and separately from the *Mtrr*^{gt} mouse line (Supplementary Fig. 1c). All mice were fed a normal chow diet (Rodent No. 3 breeding chow, Special Diet Services) *ad libitum* from weaning. A detailed breakdown of the diet was reported previously[3]. Mice were euthanized via cervical dislocation. Genotyping for the *Mtrr*^{gt} allele and/or sex was performed using PCR on DNA extracted from ear tissue or yolk sac as previously described[3,68,69].

To determine the multigenerational effects of the *Mtrr*^{gt} allele in the maternal grandfather, the following mouse pedigree was established (Supplementary Fig. 1b). For the F1 generation, F0 *Mtrr*$^{+/gt}$ males were mated with C57Bl/6J females and the resulting *Mtrr*$^{+/+}$ progeny were analysed. For the F2 generation, F1 *Mtrr*$^{+/+}$ females were mated with C57Bl/6J males and the resulting *Mtrr*$^{+/+}$ progeny were analysed. For the F3 generation, F2 *Mtrr*$^{+/+}$ females were mated with C57Bl/6J males and the resulting *Mtrr*$^{+/+}$ progeny was analysed. A similar pedigree was established to assess the effects of the *Mtrr*^{gt} allele in the maternal grandmother with the exception of the F0 generation, which involved the mating of an *Mtrr*$^{+/gt}$ female with a C57Bl/6J male.

Tissue dissection and phenotyping

For embryo and placenta collection, timed matings were established and noon on the day that the vaginal plug was detected was considered embryonic day (E) 0.5. Embryos and placentas were dissected at E10.5 in cold 1x phosphate buffered saline and were scored for phenotypes (see below), photographed, weighed, and snap frozen in liquid nitrogen for storage at -80°C. All conceptuses were dissected using a Zeiss SteREO Discovery V8 microscope with an AxioCam MRc5 camera (Carl Zeiss). Livers were collected from pregnant female mice (gestational day 10.5), weighed and snap frozen in liquid nitrogen for storage at -80°C. Both male and female conceptuses at E10.5 were assessed, as no sexual dimorphism is apparent at this stage[70]. While many individual mice were assessed over the course of this study, some of the individual tissue samples were assessed for the expression of multiple genes or for methylation at multiple DMRs.

A rigorous phenotyping regime was performed at E10.5 as previously described[3]. Briefly, all conceptuses were scored for one or more congenital malformation including failure of the neural tube to close in the cranial or spinal cord region, malformed branchial arches, pericardial edema, reversed heart looping, enlarged heart, and/or off-centered chorioallantoic attachment. Twinning or haemorrhaging was also scored as a severe abnormality. Conceptuses with severe abnormalities were categorized separately from resorptions, the latter of which consisted of maternal decidua surrounding an indistinguishable mass of fetally-derived tissue. Resorptions were not assessed in this study. Embryos with <30 somite pairs were considered developmentally delayed. Embryos with 30-39 somite pairs but a crown-rump length more than two standard deviations (sd) from the mean crown-rump length of C57Bl/6J control embryos were considered growth restricted or growth enhanced. Conceptuses were considered phenotypically-normal if they were absent of congenital malformations, had 30-39 somite pairs and had crown-rump lengths within two sd of controls. AxioVision 4.7.2 imaging software was used to measure crown-rump lengths (Carl Zeiss). Conceptus size at E10.5 was unaffected by litter size in all *Mtrr*^{gt} pedigrees assessed[70].

Spermatozoa collection

Spermatozoa from cauda epididymides and vas deferens were collected from 16-20 week-old fertile mice using a swim-up procedure as previously described[38] with the following amendments.

Spermatozoa were released for 20 minutes at 37°C in Donners Medium (25 mM NaHCO3, 20 mg/ml bovine serum albumin (BSA), 1 mM sodium pyruvate and 0.53% (vol/vol) sodium dl-lactate in Donners stock (135 mM NaCl, 5 mM KCl, 1 mM MgSO4, 2 mM $CaCl_2$ and 30 mM HEPES)). Samples were centrifuged at 500 x g (21°C) for 10 minutes. The supernatant was transferred and centrifuged at 1,300 x *g* (4°C) for 15 minutes. After the majority of supernatant was discarded, the samples were centrifuged at 1,300 x *g* (4°C) for 5 minutes. Further supernatant was discarded and the remaining spermatozoa were centrifuged at 12,000 x *g* for 1 minute and stored at -80°C.

Nucleic Acid Extraction

For embryo, trophoblast and liver tissue, genomic DNA (gDNA) was extracted using DNeasy Blood and Tissue kit (Qiagen) according to the manufacturer's instructions. RNA was extracted from tissues using the AllPrep DNA/RNA Mini Kit (Qiagen). For sperm, 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 samples followed by RNAse A incubation (37°C, 1 hour) and Proteinase K incubation (55°C, overnight) as was previously described[4]. DNA was extracted using phenol/chloroform/isoamyl alcohol mix (25:24:1) (Sigma-Aldrich) as per the manufacturer's instructions. DNA was precipitated using 10 M ammonium acetate, 0.1 mg/ml glycogen, and 100% ethanol and incubated at -80°C for at least 30 minutes. DNA was collected by centrifugation (13,000 rpm, 30 minutes). The pellet was washed twice in 70% ethanol, air-dried, and re-suspended in TE buffer. DNA quality and quantity was confirmed using gel electrophoresis and QuantiFluor dsDNA Sample kit (Promega) as per the manufacturer's instructions.

Whole Genome Sequencing (WGS)

DNA was extracted from two whole C57Bl/6J embryos at E10.5 (one male, one female) and six whole $Mtrr^{gt/gt}$ embryos with congenital malformations at E10.5 (four males, two females). Non-degraded gDNA was sent to BGI (Hong Kong) for library preparation and sequencing. The libraries of each embryo were sequenced separately. Sequencing was performed with 150 bp paired-end reads on an Illumina HiSeq X machine. Quality control of reads assessed with FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Adapters and low quality bases

removed using Trim Galore (https://www.bioinformatics.babraham.ac.uk/projects/trim galore/). Summary metrics were created across all samples using the MultiQC package (http://multiqc.info)[71]. Sequencing reads were aligned to the C57Bl/6J reference genome (GRCm38, mm10) using BowTie2 with default parameters (http://bowtie-bio.sourceforge.net/bowtie2/index.shtml)[72]. Duplicates were marked using Picard (http://broadinstitute.github.io/picard).

Structural variant analysis was performed using Manta[73]. Structural variants (SVs) were filtered using vcftools (version 0.1.15)[74]. In order to identify single nucleotide polymorphisms (SNPs), the data was remapped to the *mm10* reference mouse genome using BWA (version 0.7.15-r1144- dirty)[75]. Reads were locally realigned and SNPs and short indels identified using GenomeAnalysisTK (GATK, version 3.7)[76]. 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. Two rounds of filtering of variants were performed as follows. Firstly, low quality and biased variant calls were removed. Secondly, variants with: i) simple repeats with a periodicity <9 bp, ii) homopolymer repeats >8 bp, iii) dinucleotide repeats >14 bp, iv) low mapping quality (<40), v) overlapping annotated repeats or segmental duplications, and vi) >3 heterozygous variants fell within a 10 kb region were removed using vcftools (version 0.1.15) as was previously described[77]. The 129P2/OlaHsd mouse genome variation data was downloaded from Mouse Genomes Project[78].

Methylated DNA immunoprecipitation and sequencing (MeDIP-Seq)

MeDIP-seq was carried out as described previously[79]. Briefly, 3 µg of sperm gDNA was sonicated using a Diagenode Bioruptor UCD-200 to yield 200-700 bp fragments that were end-repaired and dA-tailed. Illumina adaptors for paired-end sequencing were ligated using the NEB Next DNA Library Prep Master Mix for Illumina kit (New England Biolabs). After each step, the DNA was washed using Agencourt AMPure XP SPRI beads (Beckman Coulter). Immunoprecipitations (IPs) were performed in triplicate using 500 ng of DNA per sample, 1.25 µl of mouse anti-human 5mC antibody (clone 33D3; Eurogentec Ltd., Cat No. BI-MECY, RRID:AB_2616058), and 10 µl of Dynabeads coupled with M-280 sheep anti-mouse IgG bead (Invitrogen). The three IPs were

pooled and purified using MinElute PCR Purification columns (Qiagen). Libraries were amplified by PCR (12 cycles) using Phusion High-Fidelity PCR Master Mix and adaptor specific iPCR tag primers (Supplementary Table 7), and purified using Agencourt AMPure XP SPRI beads. The efficiency of the IP was verified using qPCR to compare the enrichment for DNA regions of known methylation status (e.g., methylated in sperm: *H19* and *Peg3* ICR, Supplementary Fig. 4a)[49] in the pre-amplification input and the IP fractions. MeDIP library DNA concentrations were estimated using the Kapa Library Quantification kit (Kapa Biosystems) and were further verified by running on an Agilent High Sensitivity DNA chip on an Agilent 2100 BioAnalyzer. Sequencing of MeDIP libraries was performed using 100 bp paired-end reads on an Illumina HiSeq platform at the Babraham Institute Next Generation Sequencing Facility (Cambridge, UK).

Quality assessment of the sequencing reads was performed using FastQC (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)[72]. All programmes were run with default settings unless otherwise stated. Sample clustering was assessed using principle component analysis (PCA), using the 500 most variable windows with respect to read coverage (as a proxy for methylation) for 5 kb window across all samples. Further data quality checks and differential methylation analysis was performed using the MEDIPS package in R[33]. The following key parameters were defined: BSgenome = BSgenome.Mmusculus.UCSC.mm10, uniq = 1e-3, extend = 300, ws = 500, shift=0. Differentially methylated regions (DMRs) were defined as windows (500 bp) in which there was at least 1.5-fold difference in methylation (reads per kilobase million mapped reads (RPKM)) between C57Bl/6J and *Mtrr* sperm methylation level with a p-value <0.01. Adjacent windows were merged using BEDTools (version 2.27.0)[80]. The background methylome was defined as all 500 bp windows across the genome at which the sum of the average RPKM per genotype group was >1.0. The genomic localisations of DMRs including association with coding/non-coding regions and CpG islands was determined using annotation downloaded from University of California, Santa Cruz (UCSC)[81]. The percentage of DMRs associated with repetitive regions of the genome was calculated using RepeatMasker software (http://www.repeatmasker.org).

Enrichment analysis of published ChIP-seq and ATAC-seq data

Mean enrichment of specific histone modifications and THSS, CTCF, H3.3, and PRM1 binding in the DMR regions were determined using published data sets including ChIP-seq data in CD1 spermatozoa collected from cauda epididymis[42], ESCs[35] and TSCs[35], and ATAC-seq data in CD1 spermatozoa collected from cauda epididymis[42] and B6D2F1 epiblast and extraembryonic ectoderm at E6.5[43]. The source and accession numbers of processed ChIP-seq and ATAC-seq wig/bigwig files are shown in Supplementary Table 8 and accessible on GitHub (https://github.com/CTR-BFX/Blake_Watson). To ensure that the analysis was consistent across public data sets, all wig files were converted to bigwig using UCSC tools "wigToBigWig -clip" (http://hgdownload.soe.ucsc.edu/admin/exe/).

DMRs identified in sperm from all three *Mtrr* genotypes (MeDIP-seq analysis) were combined to generate a list of 893 DMRs. To prevent the inclusion of DMRs associated with genomic variation, the 20 Mb region surrounding the *Mtrr* gene (Chr13:58060780-80060780) that was identified as 129P2Ola/Hsd genomic sequence was masked. This resulted in 459 DMRs for subsequent analysis. The ChIP-seq and ATAC-seq files in sperm[42] were originally aligned to mouse reference genome mm9. The files were converted to mouse reference genome mm10 to match the other published data sets in this analysis using hgLiftOver with the mm9ToMm10.over.chain (http://genome.ucsc.edu/cgi-bin/hgLiftOver). To identify the baseline enrichment profiles around the DMRs for specific histone modifications, THSS, CTCF, H3.3, or PRM1, a similar number of genomic regions were randomly selected using bedtools (v2.26.0)[80] with the following command: "bedtools2/2.26.0, bedtools shuffle -i DMRs.bed -g Mus_GRCm38_chrall.fa.fai -chrom -seed 27442958 -noOverlapping -excl Mtrr_mask20Mb.bed". The DMR profiles were created using deeptools (version 2.3.1)[82], via computeMatix 3 kb scaled windows, flanking regions of 6 kb, and a bin size of 200 bp, and plotted with plotProfile.

Enhancer analysis

To determine whether genetic variants or DMRs overlapped with known enhancer regions, FANTOM5 enhancer database[83] for GRCm38 mouse genome (https://fantom.gsc.riken.jp/5/;

F5.mm10.enhancers.bed.gz) was used. All .bed files were applied to UCSC Genome Browser[81] to check for region specific features. The distance in base pairs between DMRs and genes or between enhancers and genes was calculated using the closest coordinates (the start/end of DMR or enhancer) to the transcriptional start site (TSS) minus the TSS and then plus 1. To determine further interactions between enhancers and promoters, we analysed public Hi-C data sets[35] for ESC and TSC (E-MTAB-6585; ESC_promoter-other_interactions_GOTHiC.txt and TSC_promoter-other_interactions_GOTHiC.txt) using WashU Epigenome Browser[84].

Quantitative reverse transcription PCR (RT-qPCR)

For RNA expression analysis, cDNA was synthesised using RevertAid H Minus reverse transcriptase (Thermo Scientific) and random hexamer primers (Thermo Scientific) using 1-2 µg of RNA in a 20-µl reaction according to manufacturer's instructions. PCR amplification was conducted using MESA Green qPCR MasterMix Plus for SYBR Assay (Eurogentec Ltd.) on a DNA Engine Opticon2 thermocycler (BioRad). The following cycling conditions were used: 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 *Hprt* and/or *Gapdh* RNA levels. Relative cDNA expression levels were analysed as previously described[85]. Transcript levels in C57Bl/6J tissue were normalized to 1. For primer sequences and concentrations, refer to Supplementary Table 9.

Bisulfite mutagenesis and pyrosequencing

Between 250 ng and 2 µg of gDNA extracted from each tissue sample was bisulfite treated with an Imprint DNA Modification Kit (Sigma). Control samples lacking DNA template were run to ensure that there was no contamination during the bisulfite conversion of DNA. To quantify DMR CpG methylation, pyrosequencing was performed. 50 ng of bisulfite-converted DNA was used as template for PCR, together with 0.2 uM of each biotinylated primer and 0.25 units of HotStarTaq PlusDNA Polymerase (Qiagen). Refer to Supplementary Table 10 for primer sequences, which were designed using PyroMark Assay Design Software 2.0 (Qiagen). PCR was performed in triplicate using the following conditions: 95°C for 5 minutes, 40 cycles of 94°C for 30 seconds, 56°C for 30 seconds, 72°C for 55 seconds, and then 72°C for 5 minutes. PCR products were

purified using Strepdavidin Sepharose High Performance beads (GE healthcare). The beads bound to DNA were washed in 70% ethanol, 0.4 M NaOH and 10 mM Tris-acetated (pH 7.6) and then hybridized to the sequencing primer in PyroMark annealing buffer (Qiagen) according to the manufacturer's instructions. Pyrosequencing was conducted using PyroMark Gold reagents kit (Qiagen) on a PyroMark MD pyrosequencer (Biotage). The mean CpG methylation was calculated using three to eight biological replicates and at least two technical replicates. Analysis of methylation status was performed using Pyro Q-CpG software (Biotage).

Mass spectrometry

Sperm gDNA 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. 100 ng of degraded DNA per individual was sent to the Babraham Institute Mass Spectrometry Facility (Cambridge, UK), where global cytosine, 5mC and 5- 5hmC was determined by liquid chromatography-tandem mass spectrometry (LC-MS/MS) as previously described[86]. Sperm of nine males were assessed per genotype and analysed in pools each containing three unique individuals. All pooled samples were analysed in triplicate. Global 5mC and 5hmC levels are reported as percentages relative to C.

Western blotting

Embryos and placentas at E10.5 were homogenised 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 β-glycerolphosphate, 1 mM Na_3VO_4 and complete mini EDTA-free proteases inhibitor cocktail [Roche Diagnostics]) with Lysing Matrix D ceramic beads (MP Biomedical) using a MagNA Lyser (Roche Diagnostics) at 5,500 rpm for 20 seconds. Samples were incubated on ice for 5 minutes and then homogenized again at 5,500 rpm for 20 seconds. Homogenates were then incubated on ice for 20 minutes with brief intervening vortexing steps occurring every 5 minutes. Samples were then centrifuged at 10,000 x *g* for 5 minutes. Supernatant from each sample was transferred to a new tube and centrifuged again at 10,000 x *g* for 5 minutes to ensure that all residual tissue was removed. Protein concentration of tissue lysates was determined using bicinchoninic acid (Sigma-

Aldrich). Proteins were denatured with gel loading buffer (50 mM Tris [pH 6.8], 100 mN DTT, 2% SDS, 10% glycerol and trace amount of bromophenol blue) at 70°C for 10 minutes. Equivalent amounts of protein were resolved by 8-10% SDS-PAGE and blotted onto nitrocellulose (0.2 μm, Amersham Protran) with a semi-dry blotter (GE Healthcare). The membrane was stained with Ponceau S solution (Sigma-Aldrich) and the resulting scanned image was used as a loading control. After washing, the membrane was blotted with 5% skimmed milk in Tris buffered saline containing 0.1% Tween-20 (TBS-T) before incubation with 1:1,000 dilution of monoclonal rabbit anti-human HIRA (clone D2A5E, Cell Signalling Technology, Cat. No. 13307, RRID:AB_2798177) overnight at 4°C or 1:10,000 dilution of monoclonal mouse anti-human β-actin (clone AC-74, Sigma-Aldrich, Cat. No. A2228, RRID:AB_476997) for 1 hour at room temperature. Primary antibodies were diluted in TBS-T. The membrane was incubated with 1:10,000 dilution of anti-rabbit IgG conjugated to horse radish peroxidase (HRP; GE Healthcare) diluted in 2.5% skimmed milk in TBS-T or anti-mouse IgG conjugated to HRP (GE Healthcare) diluted in TBS-T. The signal of resolved protein was visualized by Amersham enhanced chemiluminescence (ECL) Western Blotting Analysis System (GE Healthcare) using Amersham Hyperfilm ECL (GE Healthcare). A flat-bed scanner (HP Scanjet G4050) was used to scan films. Band intensities were determined with background subtraction using ImageJ (64-bit) software (NIH, USA).

Statistical analysis

Statistical analysis was performed using GraphPad Prism software (version 7). RT-qPCR data were analysed by independent unpaired t tests or ordinary one-way ANOVA with Dunnett's or Sidak's multiple comparison testing. SV and SNP data were analysed by independent unpaired t tests with Welch's correction. Bisulfite pyrosequencing data, SV chromosome frequency and DMR distribution at repetitive elements were analysed by two-way ANOVAs with Dunnett's, Sidak's or Tukey's multiple comparisons tests. A binomial test (Wilson/Brown) was used to compare the observed frequency of nucleosome occupancy at DMRs to expected values. Western blot data were analysed using a non-parametric Kruskal-Wallis test with Dunn's multiple comparisons test. $p<0.05$ was considered significant unless otherwise stated.

DATA AVAILABILITY

All relevant data are available from the corresponding author upon reasonable request. WGS data have been deposited in ArrayExpress database at EMBL-EBI under accession number E-MTAB-8513 (https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-8513) and MeDIP-Seq data accession number is E-MTAB-8533 (https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-8533).

CODE AVAILABILITY

The in-house scripts used for the analysis can be found in the following online repository: https://github.com/CTR-BFX/Blake_Watson.

REFERENCES

1. Chen, Q. *et al.* Sperm tsRNAs contribute to intergenerational inheritance of an acquired metabolic disorder. *Science* **351**, 397-400 (2016).
2. Gapp, K. *et al.* Implication of sperm RNAs in transgenerational inheritance of the effects of early trauma in mice. *Nat Neurosci* **17**, 667-9 (2014).
3. Padmanabhan, N. *et al.* Mutation in folate metabolism causes epigenetic instability and transgenerational effects on development. *Cell* **155**, 81-93 (2013).
4. Radford, E.J. *et al.* In utero effects. In utero undernourishment perturbs the adult sperm methylome and intergenerational metabolism. *Science* **345**, 1255903 (2014).
5. Anway, M.D., Cupp, A.S., Uzumcu, M. & Skinner, M.K. Epigenetic transgenerational actions of endocrine disruptors and male fertility. *Science* **308**, 1466-9 (2005).
6. Blake, G.E. & Watson, E.D. Unravelling the complex mechanisms of transgenerational epigenetic inheritance. *Curr Opin Chem Biol* **33**, 101-7 (2016).
7. Heard, E. & Martienssen, R.A. Transgenerational epigenetic inheritance: myths and mechanisms. *Cell* **157**, 95-109 (2014).
8. Beck, D., Sadler-Riggleman, I. & 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* **3**(2017).
9. Sharma, U. *et al.* Biogenesis and function of tRNA fragments during sperm maturation and fertilization in mammals. *Science* **351**, 391-396 (2016).
10. Siklenka, K. *et al.* Disruption of histone methylation in developing sperm impairs offspring health transgenerationally. *Science* **350**, aab2006 (2015).
11. Jimenez-Chillaron, J.C. *et al.* Intergenerational transmission of glucose intolerance and obesity by in utero undernutrition in mice. *Diabetes* **58**, 460-8 (2009).
12. Elmore, C.L. *et al.* Metabolic derangement of methionine and folate metabolism in mice deficient in methionine synthase reductase. *Mol Genet Metab* **91**, 85-97 (2007).
13. Leclerc, D. *et al.* Cloning and mapping of a cDNA for methionine synthase reductase, a flavoprotein defective in patients with homocystinuria. *Proc Natl Acad Sci U S A* **95**, 3059-64 (1998).
14. Yamada, K., Gravel, R.A., Toraya, T. & Matthews, R.G. Human methionine synthase reductase is a molecular chaperone for human methionine synthase. *Proc Natl Acad Sci U S A* **103**, 9476-81 (2006).

15. Field, M.S., Kamynina, E., Chon, J. & Stover, P.J. Nuclear Folate Metabolism. *Annu Rev Nutr* **38**, 219-243 (2018).
16. Shane, B. & Stokstad, E.L. Vitamin B12-folate interrelationships. *Annu Rev Nutr* **5**, 115-41 (1985).
17. Ghandour, H., Lin, B.F., Choi, S.W., Mason, J.B. & Selhub, J. Folate status and age affect the accumulation of L-isoaspartyl residues in rat liver proteins. *J Nutr* **132**, 1357-60 (2002).
18. Jacob, R.A. *et al.* Moderate folate depletion increases plasma homocysteine and decreases lymphocyte DNA methylation in postmenopausal women. *J Nutr* **128**, 1204-12 (1998).
19. Wainfan, E., Moller, M.L., Maschio, F.A. & Balis, M.E. Ethionine-induced changes in rat liver transfer RNA methylation. *Cancer Res* **35**, 2830-5 (1975).
20. Wilson, A. *et al.* A common variant in methionine synthase reductase combined with low cobalamin (vitamin B12) increases risk for spina bifida. *Mol Genet Metab* **67**, 317-23 (1999).
21. Rosenblatt, D.S., Cooper, B.A., Schmutz, S.M., Zaleski, W.A. & Casey, R.E. Prenatal vitamin B12 therapy of a fetus with methylcobalamin deficiency (cobalamin E disease). *Lancet* **1**, 1127-9 (1985).
22. Schuh, S. *et al.* Homocystinuria and megaloblastic anemia responsive to vitamin B12 therapy. An inborn error of metabolism due to a defect in cobalamin metabolism. *N Engl J Med* **310**, 686-90 (1984).
23. Koury, M.J. & Ponka, P. New insights into erythropoiesis: the roles of folate, vitamin B12, and iron. *Annu Rev Nutr* **24**, 105-31 (2004).
24. Padmanabhan, N. *et al.* Abnormal folate metabolism causes age-, sex- and parent-of-origin-specific haematological defects in mice. *J Physiol* **596**, 4341-4360 (2018).
25. Nashun, B. *et al.* Continuous Histone Replacement by Hira Is Essential for Normal Transcriptional Regulation and De Novo DNA Methylation during Mouse Oogenesis. *Mol Cell* **60**, 611-25 (2015).
26. Pchelintsev, N.A. *et al.* Placing the HIRA histone chaperone complex in the chromatin landscape. *Cell Rep* **3**, 1012-9 (2013).
27. Lin, C.J., Koh, F.M., Wong, P., Conti, M. & Ramalho-Santos, M. Hira-mediated H3.3 incorporation is required for DNA replication and ribosomal RNA transcription in the mouse zygote. *Dev Cell* **30**, 268-79 (2014).
28. Simon, M.M. *et al.* A comparative phenotypic and genomic analysis of C57BL/6J and C57BL/6N mouse strains. *Genome Biol* **14**, R82 (2013).
29. Kim, S. *et al.* PRMT5 protects genomic integrity during global DNA demethylation in primordial germ cells and preimplantation embryos. *Mol Cell* **56**, 564-79 (2014).
30. Deniz, O., Frost, J.M. & Branco, M.R. Regulation of transposable elements by DNA modifications. *Nat Rev Genet* **20**, 417-431 (2019).
31. Blake, G.E.T., Hall, J., Petkovic, G.E. & Watson, E.D. Analysis of spermatogenesis and fertility in adult mice with a hypomorphic mutation in the Mtrr gene. *Reprod Fertil Dev* (2019).
32. Taiwo, O. *et al.* Methylome analysis using MeDIP-seq with low DNA concentrations. *Nat Protoc* **7**, 617-36 (2012).
33. Lienhard, M., Grimm, C., Morkel, M., Herwig, R. & Chavez, L. MEDIPS: genome-wide differential coverage analysis of sequencing data derived from DNA enrichment experiments. *Bioinformatics* **30**, 284-6 (2014).
34. Orozco, L.D. *et al.* Epigenome-wide association of liver methylation patterns and complex metabolic traits in mice. *Cell Metab* **21**, 905-17 (2015).
35. Schoenfelder, S. *et al.* Divergent wiring of repressive and active chromatin interactions between mouse embryonic and trophoblast lineages. *Nat Commun* **9**, 4189 (2018).
36. Watkins-Chow, D.E. & Pavan, W.J. Genomic copy number and expression variation within the C57BL/6J inbred mouse strain. *Genome Res* **18**, 60-6 (2008).
37. Casas, E. & Vavouri, T. Sperm epigenomics: challenges and opportunities. *Front Genet* **5**, 330 (2014).
38. Hisano, M. *et al.* Genome-wide chromatin analysis in mature mouse and human spermatozoa. *Nat Protoc* **8**, 2449-70 (2013).

39. Brunner, A.M., Nanni, P. & Mansuy, I.M. Epigenetic marking of sperm by post-translational modification of histones and protamines. *Epigenetics Chromatin* **7**, 2 (2014).
40. Carone, B.R. *et al.* High-resolution mapping of chromatin packaging in mouse embryonic stem cells and sperm. *Dev Cell* **30**, 11-22 (2014).
41. Erkek, S. *et al.* Molecular determinants of nucleosome retention at CpG-rich sequences in mouse spermatozoa. *Nat Struct Mol Biol* **20**, 868-75 (2013).
42. Jung, Y.H. *et al.* Chromatin States in Mouse Sperm Correlate with Embryonic and Adult Regulatory Landscapes. *Cell Rep* **18**, 1366-1382 (2017).
43. Smith, Z.D. *et al.* Epigenetic restriction of extraembryonic lineages mirrors the somatic transition to cancer. *Nature* **549**, 543-547 (2017).
44. Seah, M.K.Y. & Messerschmidt, D.M. From Germline to Soma: Epigenetic Dynamics in the Mouse Preimplantation Embryo. *Curr Top Dev Biol* **128**, 203-235 (2018).
45. Tang, W.W., Kobayashi, T., Irie, N., Dietmann, S. & Surani, M.A. Specification and epigenetic programming of the human germ line. *Nat Rev Genet* **17**, 585-600 (2016).
46. Hackett, J.A. *et al.* Germline DNA demethylation dynamics and imprint erasure through 5-hydroxymethylcytosine. *Science* **339**, 448-52 (2013).
47. Kobayashi, H. *et al.* Contribution of intragenic DNA methylation in mouse gametic DNA methylomes to establish oocyte-specific heritable marks. *PLoS Genet* **8**, e1002440 (2012).
48. Smallwood, S.A. *et al.* Dynamic CpG island methylation landscape in oocytes and preimplantation embryos. *Nat Genet* **43**, 811-4 (2011).
49. Lambrot, R. *et al.* Low paternal dietary folate alters the mouse sperm epigenome and is associated with negative pregnancy outcomes. *Nat Commun* **4**, 2889 (2013).
50. Murphy, P.J. *et al.* NRF2 loss recapitulates heritable impacts of paternal cigarette smoke exposure. *PLoS Genet* **16**, e1008756 (2020).
51. Gardiner-Garden, M. & Frommer, M. CpG islands in vertebrate genomes. *J Mol Biol* **196**, 261-82 (1987).
52. Roberts, C. *et al.* Targeted mutagenesis of the Hira gene results in gastrulation defects and patterning abnormalities of mesoendodermal derivatives prior to early embryonic lethality. *Mol Cell Biol* **22**, 2318-28 (2002).
53. Batista, P.J. & Chang, H.Y. Long noncoding RNAs: cellular address codes in development and disease. *Cell* **152**, 1298-307 (2013).
54. Stover, P.J. One-carbon metabolism-genome interactions in folate-associated pathologies. *J Nutr* **139**, 2402-5 (2009).
55. Blount, B.C. *et al.* Folate deficiency causes uracil misincorporation into human DNA and chromosome breakage: implications for cancer and neuronal damage. *Proc Natl Acad Sci U S A* **94**, 3290-5 (1997).
56. Bistulfi, G., Vandette, E., Matsui, S. & Smiraglia, D.J. Mild folate deficiency induces genetic and epigenetic instability and phenotype changes in prostate cancer cells. *BMC Biol* **8**, 6 (2010).
57. Jablonka, E. Epigenetic inheritance and plasticity: The responsive germline. *Prog Biophys Mol Biol* **111**, 99-107 (2013).
58. Miska, E.A. & Ferguson-Smith, A.C. Transgenerational inheritance: Models and mechanisms of non-DNA sequence-based inheritance. *Science* **354**, 59-63 (2016).
59. Messerschmidt, D.M., Knowles, B.B. & Solter, D. DNA methylation dynamics during epigenetic reprogramming in the germline and preimplantation embryos. *Genes Dev* **28**, 812-28 (2014).
60. Watson, E.D. & Rakoczy, J. Fat eggs shape offspring health. *Nat Genet* **48**, 478-9 (2016).
61. Zhang, Y. *et al.* Dnmt2 mediates intergenerational transmission of paternally acquired metabolic disorders through sperm small non-coding RNAs. *Nat Cell Biol* **20**, 535-540 (2018).
62. Torne, J. *et al.* Two HIRA-dependent pathways mediate H3.3 de novo deposition and recycling during transcription. *Nat Struct Mol Biol* **27**, 1057-1068 (2020).
63. Zhang, Y., Shi, J., Rassoulzadegan, M., Tuorto, F. & Chen, Q. Sperm RNA code programmes the metabolic health of offspring. *Nat Rev Endocrinol* **15**, 489-498 (2019).
64. Holland, M.L. *et al.* Early-life nutrition modulates the epigenetic state of specific rDNA genetic variants in mice. *Science* **353**, 495-8 (2016).

65. Gapp, K. *et al.* Alterations in sperm long RNA contribute to the epigenetic inheritance of the effects of postnatal trauma. *Mol Psychiatry* (2018).
66. Branco, M.R. *et al.* Maternal DNA Methylation Regulates Early Trophoblast Development. *Dev Cell* **36**, 152-63 (2016).
67. Park, S.J., Shirahige, K., Ohsugi, M. & Nakai, K. DBTMEE: a database of transcriptome in mouse early embryos. *Nucleic Acids Res* **43**, D771-6 (2015).
68. Jiang, Z. *et al.* Zic3 is required in the extra-cardiac perinodal region of the lateral plate mesoderm for left-right patterning and heart development. *Hum Mol Genet* **22**, 879-89 (2013).
69. Tunster, S.J. Genetic sex determination of mice by simplex PCR. *Biol Sex Differ* **8**, 31 (2017).
70. Padmanabhan, N. *et al.* Multigenerational analysis of sex-specific phenotypic differences at midgestation caused by abnormal folate metabolism. *Environ Epigenet* **3**, dvx014 (2017).
71. Ewels, P., Magnusson, M., Lundin, S. & Kaller, M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047-8 (2016).
72. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-9 (2012).
73. Chen, X. *et al.* Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics* **32**, 1220-2 (2016).
74. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156-8 (2011).
75. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-60 (2009).
76. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297-303 (2010).
77. 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* **8**, 54 (2015).
78. Keane, T.M. *et al.* Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature* **477**, 289-94 (2011).
79. Senner, C.E., Krueger, F., Oxley, D., Andrews, S. & Hemberger, M. DNA methylation profiles define stem cell identity and reveal a tight embryonic-extraembryonic lineage boundary. *Stem Cells* **30**, 2732-45 (2012).
80. Quinlan, A.R. & Hall, I.M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-2 (2010).
81. Meyer, L.R. *et al.* The UCSC Genome Browser database: extensions and updates 2013. *Nucleic Acids Res* **41**, D64-9 (2013).
82. Ramirez, F., Dundar, F., Diehl, S., Gruning, B.A. & Manke, T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res* **42**, W187-91 (2014).
83. Andersson, R. *et al.* An atlas of active enhancers across human cell types and tissues. *Nature* **507**, 455-461 (2014).
84. Li, D., Hsu, S., Purushotham, D., Sears, R.L. & Wang, T. WashU Epigenome Browser update 2019. *Nucleic Acids Res* **47**, W158-W165 (2019).
85. Livak, K.J. & Schmittgen, T.D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25**, 402-8 (2001).
86. Ficz, G. *et al.* FGF signaling inhibition in ESCs drives rapid genome-wide demethylation to the epigenetic ground state of pluripotency. *Cell Stem Cell* **13**, 351-9 (2013).

ACKNOWLEDGEMENTS

We thank Drs Nozomi Takahashi and Tessa Bertozzi for their technical support, Dr Claire Senner for critical discussion, and Drs Jamie Hackett and Miguel Branco for .bed files. The MeDIP-seq and LC-MS/MS was performed at the Babraham Institute (UK). WGS was performed by BGI (Hong Kong). This work was funded by grants from Lister Institute for Preventative Medicine and the Centre for Trophoblast Research (CTR) to E.D.W and from the MRC MR/J001597 and Wellcome Trust WT095606 (to A.C.F.-S). The following support was provided: Wellcome Trust 4-year DTP in Developmental Mechanisms (to G.E.T.B.) and CTR funding (to G.E.T.B., H.W.Y., X.Z., R.S.H.).

AUTHOR CONTRIBUTIONS

G.E.T.B. collected sperm, performed DNA/RNA extractions, generated WGS and MeDIP libraries, and performed RT-qPCR and bisulfite pyrosequencing analyses. G.E.T.B. and E.D.W. dissected tissue and phenotyped conceptuses. G.E.T.B. and E.D.W. collected and analysed the data. X.Z., R.S.H., and G.E.T.B. designed and performed bioinformatics analyses. H.W.Y. performed the western blotting analysis. E.D.W. conceived the project. G.E.T.B., A.C.F.-S., G.J.B., and E.D.W. designed the experiments and interpreted the results. E.D.W. and G.E.T.B. wrote the manuscript. All authors read and revised the manuscript.

COMPETING INTERESTS

The authors declare no competing interests.

SUPPLEMENTARY INFORMATION

Defective folate metabolism causes germline epigenetic instability and distinguishes *Hira* as a phenotype inheritance biomarker

Georgina E.T. Blake[1,2,†], Xiaohui Zhao[1,2], Hong wa Yung[1,2], Graham J. Burton[1,2], Anne C. Ferguson-Smith[2,3], Russell S. Hamilton[2,3], Erica D. Watson[1,2,*]

[1]Department of Physiology, Development and Neuroscience, University of Cambridge, Cambridge UK

[2]Centre for Trophoblast Research, University of Cambridge, Cambridge UK

[3]Department of Genetics, University of Cambridge, Cambridge, UK

[†]Current address: College of Medicine and Health, University of Exeter Medical School, Exeter, UK

*Corresponding author: edw23@cam.ac.uk

Supplementary Table 1. **Common structural variants (SVs)* in all *Mtrr*$^{gt/gt}$ embryos that were absent in control C57Bl/6J embryos.**

Chr	Location	Size of SV (bp)	SV type	Repeat region?	Nearest gene	Gene location	Distance: SV and TSS (bp)	SV/ enhancer overlap?**
1	80828728-80828784	56	Deletion	Yes, simple	*Dock10*	80501073-80758527	70,201	N
11	32122791-32122854	63	Deletion	Yes, simple	*Nsg2*	32000463-32059202	63,589	N
					Il9r	32187541-32200279	77,488	N
15	83975384-83975519	135	Deletion	No	*Efcab6*	83866712-84065379	89,995	N
19	36911361-37379467	468,106	Known copy number variant in C57Bl/6J strain[1]	Tandem duplication	*Marchf5*	37207543-37222151	-	Y
					Kif11	37376403-37421859		
					Fgfbp3	36917550-36919599		
					Cpeb3	37021291-37208601		
					Btaf1	36926079-37012752		
					Ide	37268743-37334544		

*Excludes 20 Mb region surrounding *Mtrr*^{gt} locus. **Known enhancers identified via FANTOM5 software in GRCm38 mouse genome. Chr, chromosome; SV, structural variant; bp, base pairs; TSS, transcriptional start site.

Supplementary Table 2 **Characteristics of common small nucleotide polymorphisms (SNPs)* in all *Mtrr*$^{gt/gt}$ embryos that were absent in control C57Bl/6J embryos.**

Chr	Location	SNP	Context	Nearest gene	Gene location	Distance between SNP and TSS (bp)	SNP/ enhancer overlap?**
1	23478171	TC→T	Intergenic	*Ogfrl1*	23366424-23383201	94,970	N
1	130763137	A→C	Intergenic	*AA986860*	130731976-130744622	31,161	N
6	114014496	T→TCTCC CCTCCC	Intronic	*Atp2b2*	113743831-114042613	28,117	N
6	121526732	G→T	Intergenic	*Iqsec3*	121372933-121473678	53,054	N
8	98943351	A→G	Intergenic	*Gm15679*	99011886-99032740	68,535	N
8	104303432	C→G	Intronic	*Cmtm1*	104293542-104310145	6,713	N
13	58081242	C→G	Intronic	*Klhl3*	58000228-58113592	32,350	N
13	81473683	C→G	Intronic	*Adgrv1*	81095068-81633154	159,471	N
13	81473686	T→C	Intronic			159,468	N
13	81473693	C→T	Intronic			159,461	N
13	81473695	C→G	Intronic			159,459	N
13	81473744	C→T	Intronic			159,410	N
13	81473746	C→G	Intronic			159,408	N
13	81473756	C→T	Intronic			159,398	N
14	119813417	ACCCC→ ACCCCC	Intronic	*Hs6st3*	119138341-119869815	675,076	N
16	85907608	GC→C	Intergenic	*Adamts5*	85858157-85901125	6,483	N
16	89162602	G→A	Intergenic	*Krtap20-2*	89205861-89206394	43,259	N
18	87587984	T→C	Intergenic	*Cbln2*	86711110-86718283	876,874	N
X	68533748	TC→C	Intergenic	*Fmr1os*	68667514-68678399	144,651	N
X	71514412	C→C$(A)_{45}$	Intergenic	*Cd99l2*	71420060-71492849	21,563	N
X	169366447	TC→T	Intergenic	*Hccs*	169311530-169320374	46,073	N

*Excludes 20 Mb region surrounding the *Mtrr*^{gt} locus. **Known enhancers identified via FANTOM5 software in GRC38m mouse genome. Chr, chromosome; SNP, small nucleotide polymorphism; TSS, transcriptional start site of nearest gene.

Supplementary Table 3 **Summary of validation of sperm DMRs by bisulfite pyrosequencing**

Male genotype*	Number of DMRs** that validated (%)		Total (%)
	Hypomethylated	Hypermethylated	
Mtrr$^{+/+}$	0/0	3/3 (100%)	3/3 (100%)
Mtrr$^{+/gt}$	3/4 (75%)	10/20 (50%)	13/24 (54.2%)	
Mtrr$^{gt/gt}$	13/13 (100%)	8/13 (61.5%)	21/26 (80.8%)
Total	16/17 (94.1%)	21/36 (58.3%)	37/53 (69.8%)

*N=8 males per experimental group including four males from MeDIP-seq analysis and four unique males. **Relative to C57Bl/6J controls. DMR, differential methylated region.

Supplementary Table 4 **Sperm DMRs identified via MeDIP-seq that were associated with known enhancer regions**

Genotype(s) in which sperm DMR identified	Chr	DMR location	Enhancer location	Nearest gene	Gene location (strand)	Distance: DMR and TSS (bp)	Distance: enhancer and TSS (bp)*
Mtrr$^{+/+}$, *Mtrr*$^{+/gt}$, *Mtrr*$^{gt/gt}$	19	37238001-37239000	37238210-37238404	*Gm25268*	37233382-37233484 (-)	4518	4727
Mtrr$^{+/gt}$, *Mtrr*$^{gt/gt}$	13	104660501-104661000	104660659-104660787	*Cwc27*	104631140-104817142 (-)	156143	156356
Mtrr$^{+/+}$	13	119596001-119596500	119596241-119596530	*Tmem267*	119488039-119611059 (+)	107963	108203
	13	119597501-119598000	119597731-119598237	*Tmem267*	119488039-119611059 (+)	109463	109603
Mtrr$^{gt/gt}$	4	147275001-147275500	147275382-147275778	*Zfp988*	147305674-147333734 (+)	30175	29897
	4	147738501-147739000	147738791-147739184	*Gm13157*	147753974-147809788 (-)	70789	70605
	8	75083501-75084000	75083800-75084239	*Hmox1*	75093621-75100589 (+)	9622	9383
	13	67797501-67798500	67797828-67797962	*4930525 G20Rik*	67796594-67830985 (-)	32486	33024

*Known enhancers identified using FANTOM5 software in CRG38m mouse genome.

Supplementary Table 5 **Common sperm DMRs in *Mtrr*$^{+/+}$, *Mtrr*$^{+/gt}$ and *Mtrr*$^{gt/gt}$ males as determined by MeDIP-seq**

Chr	Start location	Stop location	Genomic location	Nearest gene	Gene location	DMR/ enhancer overlap?*
1	79722501	79723500	intronic	*Wdfy1*	79,702,262-79,776,143	N
5	14905001	14905500	intergenic	*Gm9758*	14,910,122-14,914,899	N
5	14908501	14909000	intergenic	*Gm9758*	14,910,122-14,914,899	N
5	14914501	14915000	exonic	*Gm9758*	14,910,122-14,914,899	N
5	14933501	14934000	exonic	*Speer4e*	14,933,221-14,938,429	N
5	15006001	15007000	intergenic	*Gm10354*	14,974,113-14,978,935	N
5	15031501	15032000	exonic	*Gm17019*	15,028,950-15,032,998	N
5	15040001	15040500	intergenic	*Gm17019*	15,028,950-15,032,998	N
5	15459501	15460000	intergenic	*Gm43391*	15,464,171-15,476,857	N
5	15462501	15464000	intergenic	*Gm43391*	15,464,171-15,476,857	N
5	15472501	15473500	intronic	*Gm43391*	15,464,171-15,476,857	N
5	15509001	15509500	intergenic	*Gm21847*	15,516,489-15,656,679	N
5	15522001	15522500	intronic	*Gm21847*	15,516,489-15,656,679	N
5	15527501	15528000	intronic	*Gm21847*	15,516,489-15,656,679	N
5	15528501	15529500	intronic	*Gm21847*	15,516,489-15,656,679	N
5	15596001	15597000	intronic	*Gm21847*	15,516,489-15,656,679	N
5	15601001	15601500	intronic	*Gm21847*	15,516,489-15,656,679	N
5	15630501	15631000	intronic	*Gm21847*	15,516,489-15,656,679	N
5	15632001	15632500	intronic	*Gm21847*	15,516,489-15,656,679	N
5	15637001	15638000	intronic	*Gm21847*	15,516,489-15,656,679	N
5	15643501	15644000	intronic	*Gm21847*	15,516,489-15,656,679	N
5	15670501	15671500	intergenic	*Speer4cos*	15,680,710-15,714,596	N
5	15680501	15681000	exonic	*Speer4cos*	15,680,710-15,714,596	N
5	15690501	15691000	intronic	*Speer4cos*	15,680,710-15,714,596	N
5	15713501	15714500	exonic	*Speer4cos*	15,680,710-15,714,596	N
10	122886001	122886500	intronic	*Ppm1h*	122,678,762-122,945,795	N
13	119609001	119609500	exonic	*Tmem267*	119,488,086-119,610,459	N
14	53067001	53067500	intergenic	*Trav13d-4*	53,073,004-53,073,267	N
14	53073501	53074000	intergenic	*Trav13d-4*	53,073,004-53,073,267	N
14	53624501	53625000	intergenic	*Trav12-3*	53,621,657-53,622,245	N
17	6441501	6442000	intronic	*Tmem181b-ps*	6,438,524-6,449,745	N
17	6445001	6446500	intronic	*Tmem181b-ps*	6,438,524-6,449,745	N
17	6449501	6450500	exonic	*Tmem181b-ps*	6,438,524-6,449,745	N
17	6451001	6452000	exonic	*Tmem181b-ps*	6,438,524-6,449,745	N
17	6455501	6456500	intergenic	*Tmem181b-ps*	6,438,524-6,449,745	N
17	6457001	6457500	intergenic	*Tmem181b-ps*	6,438,524-6,449,745	N
17	6463001	6464000	intergenic	*Tmem181b-ps*	6,438,524-6,449,745	N
17	6465001	6465500	intergenic	*Tmem181b-ps*	6,438,524-6,449,745	N
17	6473001	6474000	intergenic	*Tmem181b-ps*	6,438,524-6,449,745	N
17	6479001	6479500	intergenic	*Tmem181b-ps*	6,438,524-6,449,745	N
17	6502001	6502500	intergenic	*Tmem181b-ps*	6,438,524-6,449,745	N

17	6506001	6507000	intergenic	*Tmem181b-ps*	6,438,524-6,449,745	N
17	6522501	6523000	intergenic	*Tmem181b-ps*	6,438,524-6,449,745	N
17	6583501	6584000	intergenic	*Dynlt1c*	6,601,671-6,609,679	N
17	6592001	6592500	intergenic	*Dynlt1c*	6,601,671-6,609,679	N
17	58728501	58729000	intronic	*A330072L02Rik*	58,680,347-58,836,041	N
17	58730001	58731500	intronic	*A330072L02Rik*	58,680,347-58,836,041	N
19	36911001	36912000	intergenic	*Fgfbp3*	36,917,550-36,919,615	N
19	37238001	37239000	intergenic	*Gm25268*	37,207,543-37,222,151	**Y**
19	37247501	37248000	intergenic	*4931408D14Rik*	37,258,423-37,264,095	N
19	37255001	37255500	intergenic	*4931408D14Rik*	37,258,423-37,264,095	N
19	37269001	37270000	exonic	*Ide*	37,268,743-37,330,613	N
19	37277501	37278500	exonic	*Ide*	37,268,743-37,330,613	N
19	37351501	37352500	intergenic	*Ide*	37,268,743-37,330,613	N

*Enhancers identified via FANTOM5 software in GRC38m mouse genome. See also Supplementary Fig. 5.

Supplementary Table 6 **Characteristics of sperm DMRs from *Mtrr*$^{+/gt}$ males that were assessed in F1-F2 wildtype somatic tissue.**

DMR (Coordinates)	Genomic location	Repro-gramming resistant region[2,3]	Status	Nearest gene	Gene function	DMR/ enhancer overlap?*
DMR A10 (Chr16:18976001-18977000)	Intergenic	Yes	Hyper	*Hira*	Histone H3.3 chaperone	No
DMR D20 (Chr13: 104660501-104661000)	Intragenic	No	Hypo	*Cwc27*	Spliceosome-associated protein; isomerase	Yes
DMR E115 (Chr7:36770501-36772000)	Intragenic	No	Hypo	*Tshz3*	Zinc finger transcription factor	No
DMR E52 Chr17:6324501-6325500)	Intergenic	Yes	Hyper	*Dynlt1a*	Dynein light chain	No
DMR D87 (Chr17:6324501-6325500)	Intergenic	Yes	Hyper	*4930572 O03Rik*	Unknown	No
				Speer7-cos	Spermatogenesis associated glutamate (E)-rich protein	
DMR E112 (Chr16:33270501-33271000)	Intragenic	Yes	Hypo	*Exoc4*	Exocyst complex component	No
DMR E109 (Chr5:28168501-28169000)	Intragenic	No	Hypo	*En2*	Homeobox transcription factor	No
DMR E28 (Chr14:13512001-13512500)	Intragenic	No	Hyper	*Synpr*	Synaptoporin	No

*Known enhancers identified via FANTOM5 in GRC38m mouse genome. DMR, differentially methylated region; Hyper, hypermethylated; hypo, hypomethylated.

Supplementary table 7 **iPCR Tag primers used in the MeDIP-seq analysis.**

iPCR Tag primer	Primer Sequence	Barcode	Sequence Obtained
iPCRtag1	CAAGCAGAAGACGGCATACG**AGAT**AACGTGAT**GAGA**T CGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATC	AACGTGAT	ATCACGTT
iPCRtag2	CAAGCAGAAGACGGCATACGAGATAAACATCGGAGAT CGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATC	AAACATCG	CGATGTTT
iPCRtag3	CAAGCAGAAGACGGCATACGAGATATGCCTAAGAGAT CGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATC	ATGCCTAA	TTAGGCAT
iPCRtag4	CAAGCAGAAGACGGCATACGAGATAGTGGTCAGAGAT CGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATC	AGTGGTCA	TGACCACT
iPCRtag5	CAAGCAGAAGACGGCATACGAGATACCACTGTGAGAT CGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATC	ACCACTGT	ACAGTGGT
iPCRtag6	CAAGCAGAAGACGGCATACGAGATACATTGGCGAGAT CGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATC	ACATTGGC	GCCAATGT
iPCRtag7	CAAGCAGAAGACGGCATACGAGATCAGATCTGGAGAT CGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATC	CAGATCTG	CAGATCTG
iPCRtag8	CAAGCAGAAGACGGCATACGAGATCATCAAGTGAGAT CGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATC	CATCAAGT	ACTTGATG

Supplementary table 8 **Source and accession numbers of processed ChIP-seq and ATAC-seq wig/bigwig files**

ID	Experiment	Sample	Target	GEO Accession	Source	Ref.	Data
1	ChIP-seq	ESC	H3K27ac	GSM1000099	ENCODE/PMID: 30305613	Ref.[4]	raw/ processed*
2	ChIP-seq	ESC	H3K4me3	GSM769008	ENCODE/PMID: 30305613	Ref.[4]	raw/ processed*
3	ChIP-seq	ESC	H3K9me3	GSM1000147	ENCODE/PMID: 30305613	Ref.[4]	raw/ processed*
4	ChIP-seq	ESC	H3K4me1	GSM769009	ENCODE/PMID: 30305613	Ref.[4]	raw/ processed*
5	ChIP-seq	ESC	H3K27me3	GSM1000089	ENCODE/PMID: 30305613	Ref.[4]	raw/ processed*
6	ChIP-seq	ESC	CTCF	GSM918748	ENCODE/PMID: 30305613	Ref.[4]	raw/ processed*
8	ChIP-seq	TSC	H3K27ac	GSM1035380, GSM1035381	PMID: 23396136	Ref.[4]	raw/ processed*
9	ChIP-seq	TSC	H3K4me3	GSM1035382	PMID: 23396136	Ref.[4]	raw/ processed*
10	ChIP-seq	TSC	H3K9me3	GSM1035383, GSM1035384	PMID: 23396136	Ref.[4]	raw/ processed*
11	ChIP-seq	TSC	H3K4me1	GSM1035385	PMID: 23396136	Ref.[4]	raw/ processed*
12	ChIP-seq	TSC	H3K27me3	GSM1035386, GSM1035387	PMID: 23396136	Ref.[4]	raw/ processed*
13	ChIP-seq	TSC	CTCF	GSM967658	PMID: 23178118	Ref.[4]	raw/ processed*
15	ChIP-seq	Sperm	H3K27ac	GSM2088387, GSM2401435	PMID: 28178516	Ref.[5]	processed
16	ChIP-seq	Sperm	H3K4me3	GSM2088391, GSM2401439	PMID: 28178516	Ref.[5]	processed
17	ChIP-seq	Sperm	H3K9me3	GSM2088388, GSM2401436	PMID: 28178516	Ref.[5]	processed
18	ChIP-seq	Sperm	H3K4me1	GSM2088390, GSM2401438	PMID: 28178516	Ref.[5]	processed
19	ChIP-seq	Sperm	H3K27me3	GSM2088386, GSM2401434	PMID: 28178516	Ref.[5]	processed
20	ChIP-seq	Sperm	CTCF	GSM2088382, GSM2088383, GSM2088384	PMID: 28178516	Ref.[5]	processed
21	ChIP-seq	Sperm	H3	GSM2088392	PMID: 28178516	Ref.[5]	processed
22	ChIP-seq	Sperm	PRM1	GSM2088400, GSM2401441	PMID: 28178516	Ref.[5]	processed
23	ATAC-seq	Sperm	THSS	GSM2088376, GSM2088377, GSM2088378	PMID: 28178516	Ref.[5]	processed
24	ATAC-seq	ExE	THSS	GSM2229962, GSM2229963	PMID: 28959968	Ref.[6]	processed
25	ATAC-seq	Epi	THSS	GSM2229960, GSM2229961	PMID: 28959968	Ref.[6]	processed

*We used the processed file from the author directly. ExE, extraembryonic ectoderm (E6.5); Epi, epiblast (E6.5). All files accessible on GitHub: https://github.com/CTR-BFX/Blake_Watson

Supplementary Table 9 **PCR primers used in this study.**

Gene name	Forward primer (5'→3')	Reverse primer (5'→3')	Primer conc. (nm)	Ref.
Cwc27	TGATAATGGCAGCCAGTTTTTCT	CTGTCAGGCGTAGCATGTTGT	200	-
Cts8	TCCTGTGAAGAATCAGGGCAC	GTGCTCAGTGGGACCAGTTT	200	-
Dynlt1a	GAAGACTTCCAGGCCTCCG	GGTTGACTTTGCTGTGCTGG	100	-
En2	GCTGAGTTTCAGACCAACAGGTA	GCTTGTTCTGGAACCAAATCTT	200	-
Exoc4	CACAGCCTACAGGGGCATTG	TTGGCAGCGATTTCAAGAGTC	200	-
Gapdh	CATGGCCTTCCGTGTTCCT	GCGGCACGTCAGATCCA	variable	Ref.[7]
Gas1	CCTCTGCACCACGTGTCTTA	TGGCAGTACCGAGCTTTAGG	200	-
Hira mRNA (set 1)	CTCCATCTTGTCAGGAAGTGAT	GTTCCTGGCACTCAGTAAAGAG	200	-
Hira lncRNA209 (set 2)	TATGAAAACGCCCCGCTCTT	ATTCGCGCTATTGGGCATCT	200	-
Hprt	CAGGCCAGACTTTGTTGGAT	TTGCGCTCATCTTAGGCTTT	variable	Ref.[8]
Hsd17b3	CTGAGCACTTCCGGTGAGAG	ATAAGGGGTCAGCACCTGAAT	200	-
IAP-GAG	AACCAATGCTAATTTCACCTTGGT	GCCAATCAGCAGGCGTTAGT	200	Ref.[9]
IAP-3'LTR	GCACATGCGCAGATTATTTGTT	CCACATTCGCCGTTACAAGAT	100	Ref.[9]
LINE1-5'UTR	GGCGAAAGGCAAACGTAAGA	GGAGTGCTGCGTTCTGATGA	100	Ref.[9]
LINE1-ORF2	GGAGGGACATTTCATTCTCATCA	GCTGCTCTTGTATTTGGAGCATAGA	200	Ref.[9]
March5	TTCACCAGGCTTGTCTCCA	GCATCACTGTCACTGCTCCA	150	-
Nsun2	GGTAAACCATGACGCATCC	CCTTCCACTGTGAGACTCC	100	Ref.[10]
Ptch1	CCTCTGCTCCTTGATTGGCA	TCCCCAGTCCTGTCCTCAAA	200	-
SINEB1	TGAGTTCGAGGCCAGCCTGGTCTA	ACAGGGTTTCTCTGTGTAGCCCTG	100	Ref.[9]
Srd5a1	CTTGAGCCAGTTTGCGGTGTA	GCCTCCCCTGGGTATTTTGTATC	200	-
Tbpba	ACTGGAGTGCCCAGCACAGC	GCAGTTCAGCATCCAACTGCG	200	-
Tshz3	GCGCGCAGCAGCCTATGTTTC	TCAGCCATCCGGTCACTCGTC	300	-
Uqcrb	TCTCAGGTCAAAATGGCGGG	ATCATCTCGCATTAACCCCAGT	200	-

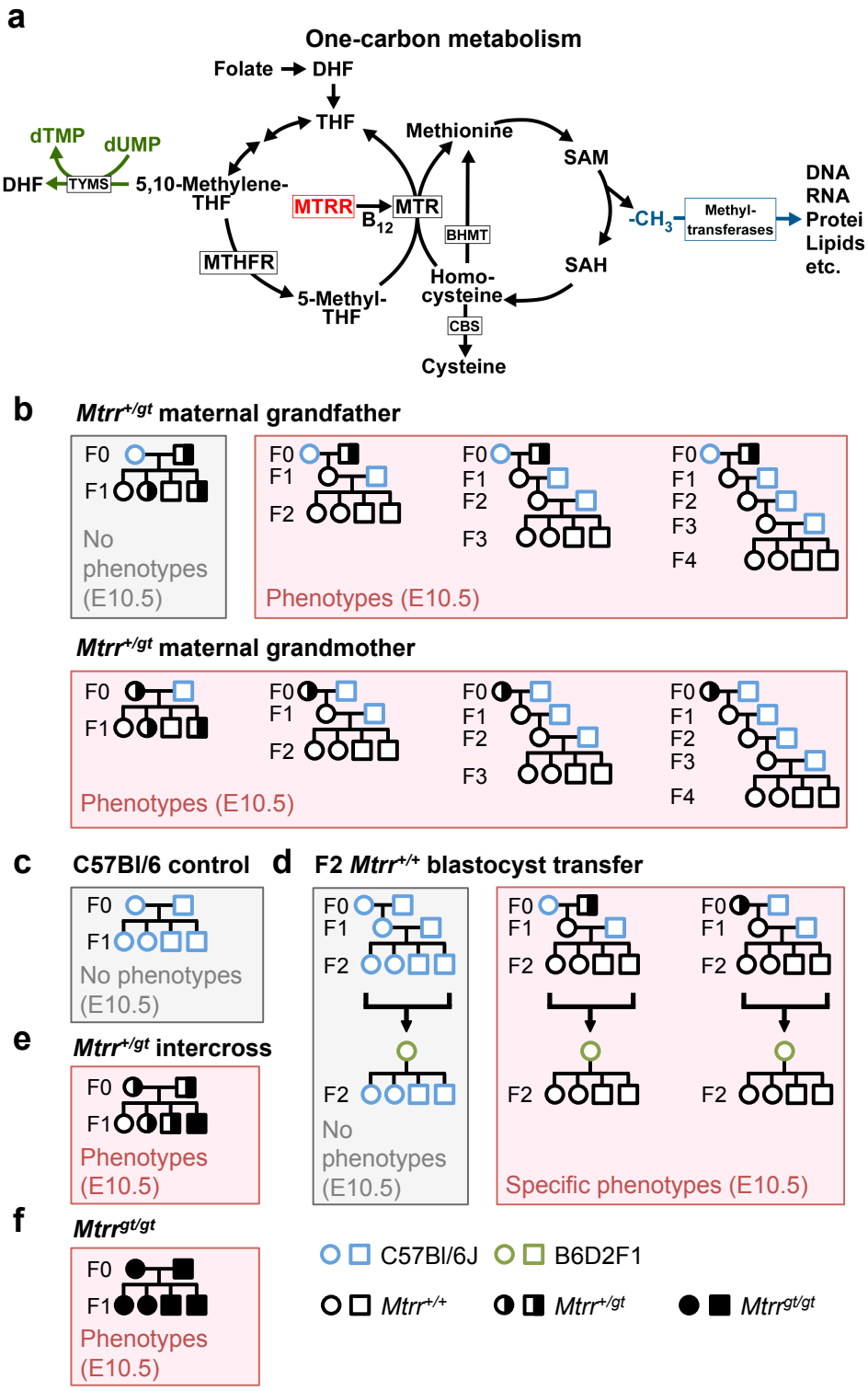
Supplementary Table 10 **Bisulfite pyrosequencing primers used in this study.**

DMR name	Coordinates	Forward primer* (5'→3')	Reverse primer (5'→3')	Sequencing primer** (5'→3')
11	Chr1: 65,104,501-65,105,000	[biotin]-ATAGATAGTGAAAGGATGGAGAGTAAGAA	CTCCAACCACTAAAAACTTAACTCA	CCACTAAAAACTTAACTCAATC
17	Chr2: 144,308,501-144,309,000	AGTTTTTTTGGGTGTGAGAATTAT	[biotin]-ATCCAAAACCACTAAAAAACACAAC	TGGGTGTGAGAATTATT
29	Chr19: 36,911,501-36,912,000	AGGGGAGAATTTTGAATGTGAT	[biotin]-ACCCCACAACTTTCTTATAACTT	GGTTATAGGGGGGAT
60	Chr10: 122,886,001-122,886,500	[biotin]-AGATGTAAAAGAAAGGAAGGTAGT	CAATCCCCCATTCAATACAAAAA	ACAAAAATACCCTCCC
177	Chr19: 37,247,501-37,248,000	[biotin]-AATTTAAGTGAAGAATTTTGGTTTTATG	AACCCTAAATATTCTCCTTTACTCAAC	ATTCTCCTTTACTCAACT
181	Chr19: 37,280,001-37,280,500	AGGAAGTATTTGAGATGTTTAGAGTT	[biotin]-CCACTACAATATATCCTTAACCCTACT	TTATAGGATTATAGGATGTGT
185	Chr3: 122,504,501-122,505,000	[biotin]-TGTTTTATTTGTAGTTGGAGAAGTAAG	ACCTCTCAAAAATCCTATCCATAATAT	CACCCCCTCACTAATCCTCTA
189	Chr4: 156,135,001-156,136,000	TGGGAGTTAGTTATTGGTTTAGTTTGAGG	[biotin]-AACCCAATACTAATCCACCTTTACA	TGAGGTTGGTATAGGAA
220	Chr6: 136,907,001-136,907,500	AGTTATGTGTAGGGTATTAATTTAATGT	[biotin]-TCTCCCCATACCTAACTTCACACT	AGATGTATGTTTTTGAAGTTTA
269	Chr6: 33,270,501-33,271,000	[biotin]-TAGGAAATAAAAATGTGAGGGGTAAT	ACTACTACTCTATCCCTTTTATAACA	ACTCTATCCCTTTTATAACAA
274	Chr7: 75,821,501-75,822,000	[biotin]-GGATGTAGATGAATTTGGAAATTTTAGA	ACTAAAACTCCACTATTCCTTTCCACAATT	CTATTCCTTTCCACAATTAC
278	Chr8: 116,801,001-116,801,500	GTTTGAATTTTTTGAGGGTTATTTAAGGTT	[biotin]-TCCCCCTTTTAAAACTATCTCTTCT	TTTGAGTTTAGTTTTTTATATTGTG
279	Chr8: 119,794,501-119,795,000	GTTTTTTTAGTAGAGTTGGGAGTTT	[biotin]-AAACAAACTAAACCTAAACAAATATAACT	GGGAGTTTTGTTTTTTTTAGAT
280	Chr10: 4,354,501-4,355,000	AGAAGGGGTTATAGGAAGTATTTTAGG	[biotin]-CCCAAAAAATAATCCCTTCCCTTTTC	GGGTTATAGGAAGTATTTTAGGA
281	Chr7: 75,821,501-75,822,000	AGGATGGGGGAATAAAATGATG	[biotin]-ACCCAAACCTATATAATAATAACTTTCC	GGGGAATAAAATGATGTG
282	Chr10: 82,975,501-82,976,500	GAATTGTTTAGGGAGGATTTTTTTATAGT	[biotin]-CCAACATTCTACAAAATCTACTACTCC	TGTTGTGTAGATAGATATTTTGAT
289	Chr11: 44,696,501-44,697,000	GGTTGTGGGTTGTTATTGTAGTTAAG	[biotin]-ATCAAATTTTCTCTTTCCCCATTC	GTTATTGTAGTTAAGGGGT
A10	Chr16: 18,975,501-18,977,000	TGTTATAGGTTGGAGGGTATGAGTT	[biotin]-ATCCAATCCTTCTCAACAATCTCTCTC	AGGAGTAGAGAGTTTTAG
B19	Chr19: 37,252,001-37,252,500	[biotin]-AATTGGAGAATAGGATTTTTAGTGTTTATA	AACAATAATAAATACTTAAACTACA	TCCTAAATCTCCCTATAACACCATCAAT
B21	Chr19: 37,260,501-37,261,000	AGTATAAGTGATGGAGTTTTAGATAAAT	[biotin]-AATCTTAAACCAATACAAATCCATACT	TGATGGAGTTTTAGATAAATTAG

C7	Chr19: 37,238,001-37,239,000	GGTAGTTTTGGATATGTAAGAGTTG	[biotin]-CATATCCCCTCTCTCTTTATCTTTTTA	GGTTTATAGAGTTGAGAATTGTG
D20	Chr13: 104,660,501-104,661,000	GAAATGTTTGAGAGAATGGGTTTTAG	[biotin]-ATACAATAACCTTAACCACTCATAAC	AAATTTTAGTTTAAGGTTTTATATA
D41	Chr19: 37,247,501-37,248,000	[biotin]-AAGTGAAGAATTTTGGTTTTATGTAGAAT	AACCCTAAATATTCTCCTTTACTCAACT	CTCCTTTACTCAACTCT
D87	Chr5: 15,671,001-15,671,500	GGGTGTTGTTTTAGAGGTTTTGTATTTTTG	[biotin]-ATCACCTCATCCTACTAAATACTATAAC	AGATTAATTTTTTAAGTTTTGTTG
E28	Chr14: 13,512,001-13,512,500	TAGGGAGATTGATTGTGTTTTATTAGAT	[biotin]-ACTACCCTCTTCCTCTCTTACTCCTAAAT	AGATTGATTGTGTTTTATTAGATT
E50	Chr15: 78,741,501-78,742,000	AGGTAGATTAGGTTGAATATTTATAGAGA	[biotin]-AAAAACAAAATCTATAAAACCCTAAATCC	GGGTTGAAATTAAAGATAGG
E52	Chr17: 6,324,501-6,325,500	AGTTTTGGAGTTTAGGTTAGGT	[biotin]-ATAACCATATCAATAAAACCTTCTTAACA	AGTTTAGGTTAGGTTAAAG
E66	Chr17: 6,562,501-6,563,000	GGAGAGGGTTGTTTTTTTTGGGATAAGT	[biotin]-ATTTACCCTTCTCCCCACTACCA	GTGTTTTTGTTTTTAGTTAGTTATT
E74	Chr18: 47,505,501-47,506,000	GTAAGTTGGAGGTAGATATAGTTAGAAGT	[biotin]-ACTCCATTTACAACTCTTCACCATACT	ATAAAAGTTTTTAAATTTATTTGGG
E81	Chr2: 135,170,501-135,171,000	TGTAAGGTTTAAGATAGGAGATGTTATT	[biotin]-TACAAACCCATATAACTACAAAACTCAT	TTTAGTTGGTGAGATTGA
E109	Chr5: 28,168,501-28,169,000	TAGAAGTTTTATTGGTTAGGTGTAAGTTA	[biotin]-CAAAAAAATCTTTAAACACCAAAAAACATC	ATGGATGTGAAGAGG
E112	Chr6: 33,270,501-33,271,000	[biotin]-TAGGAAATAAAAATGTGAGGGGTAAT	ACTACTACTCTATCCCTTTTATAACA	ACTCTATCCCTTTTATAACAAT
E114	Chr7: 16,633,501-16,634,000	GTTTTTGGGTTTTAGTAATAGTGTATGG	[biotin]-AACACATATAACTTCCCACACATC	GTTTTTAGAGAGTAGTTTAGG
E115	Chr7: 36,770,501-36,772,000	AGTTTGATATTGAAGGGGTTATTTTGGA	[biotin]-AACATAATATACTTCCCTTCTAAAATCT	GGTTATTTTGGAAGAGGT

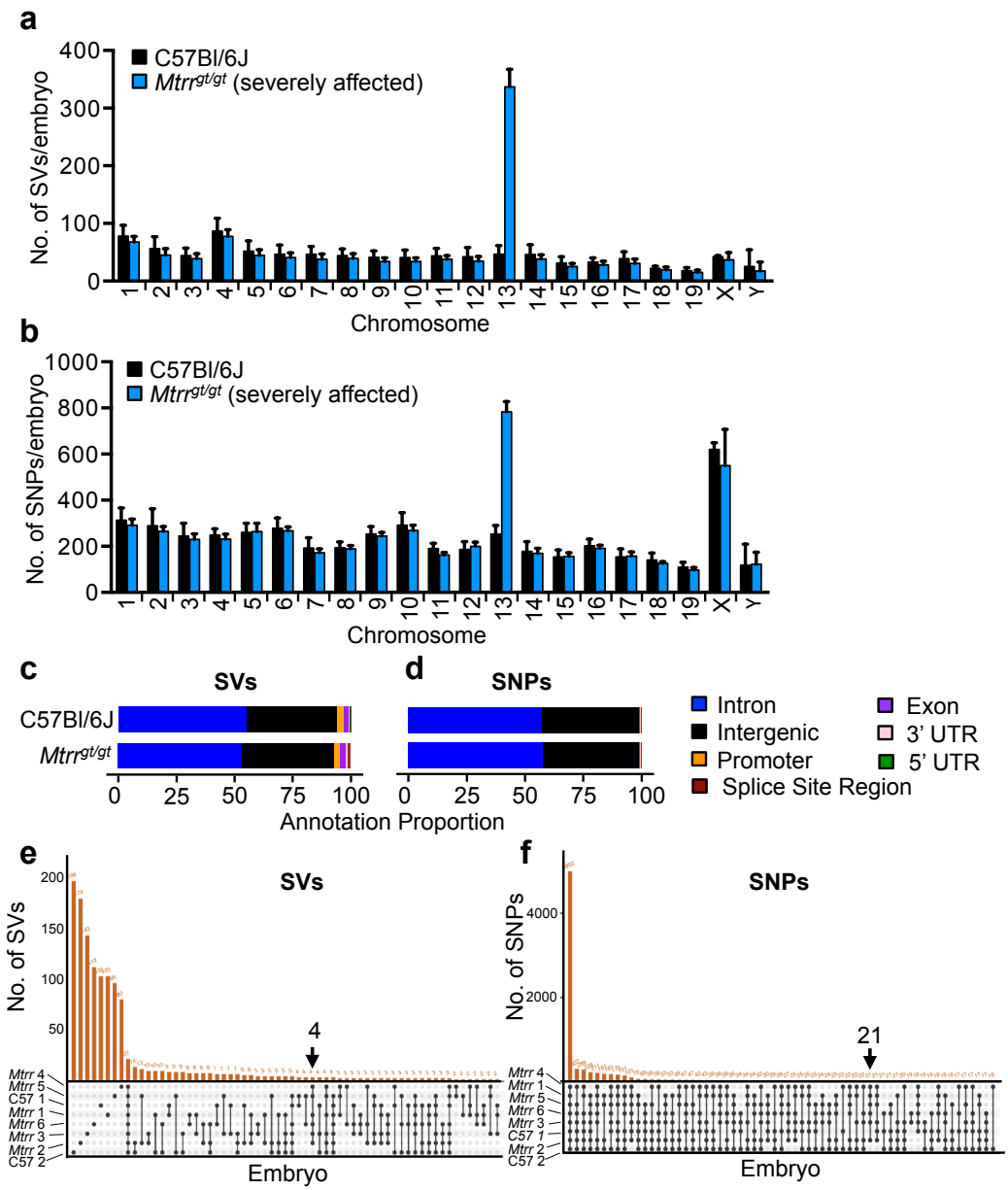
Primer concentration: *PCR primers, 250 nM; **Sequencing primers, 417 nM

SUPPLEMENTARY FIGURES AND LEGENDS

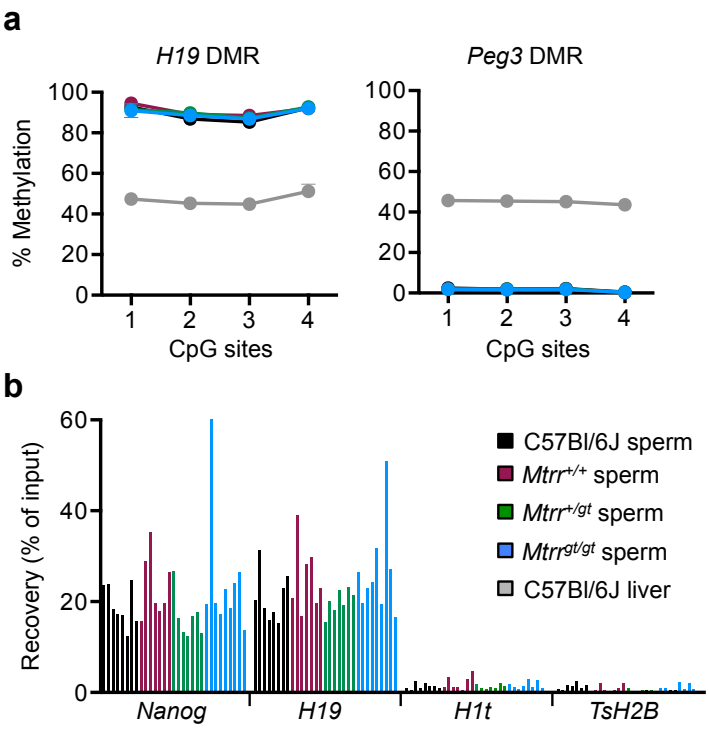


Supplementary Figure 1 **Transgenerational epigenetic inheritance (TEI) in the *Mtrr*^{gt} mouse line occurs via the maternal grandparental lineage. a** Simplified schematic drawing of one-carbon metabolism. MTRR (red) is a key enzyme at the intersection between folate and methionine pathways that is required to activate methionine synthase (MTR) through the reductive methylation of its vitamin B_{12} cofactor. BHMT, betaine-homocysteine S-methyltransferase; CBS,

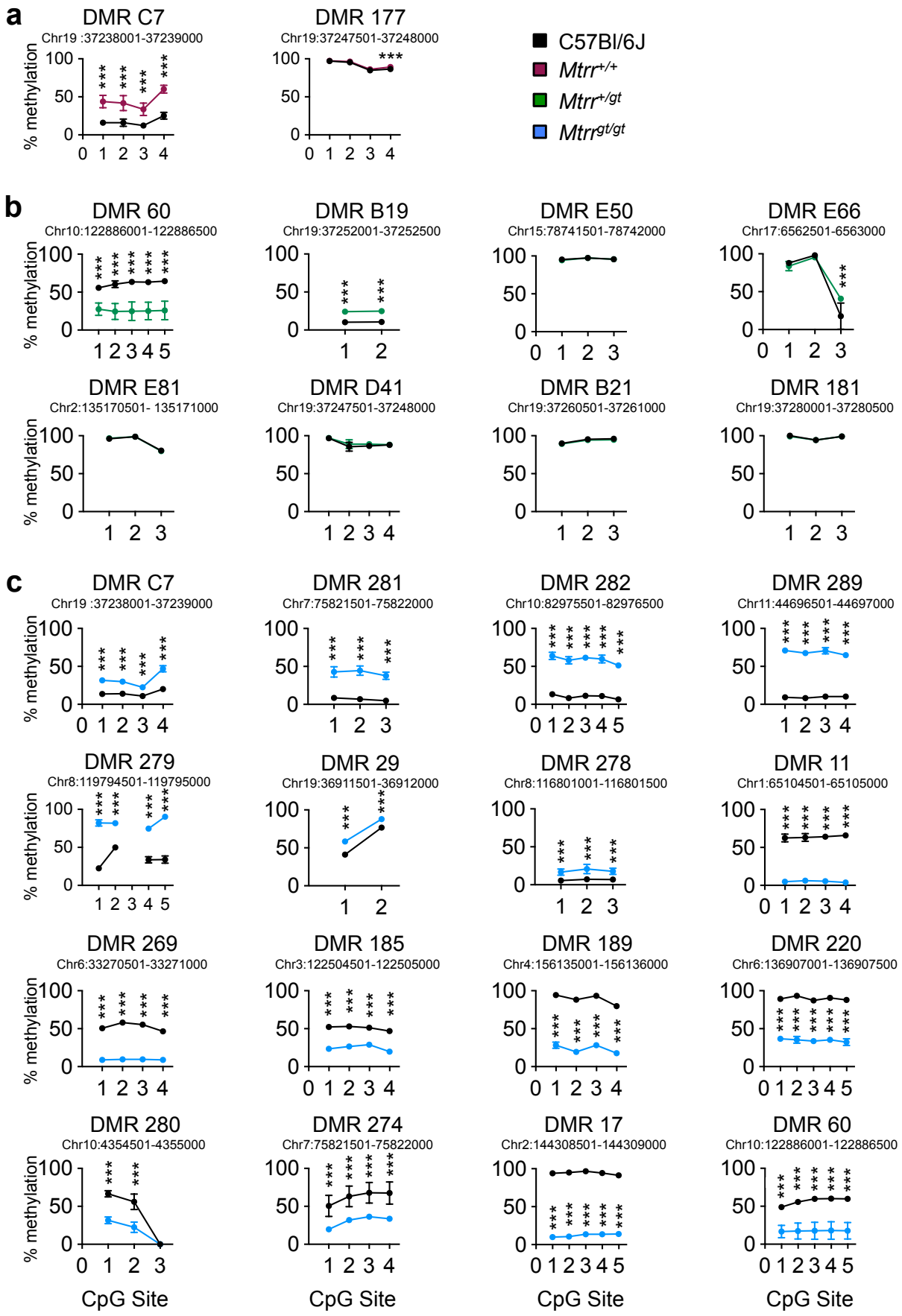
cystathionine beta-synthase; DHF, dihydrofolate; MTHFR, methylenetetrahydrofolate reductase; SAH, S-adenosyl-homocysteine; THF, tetrahydrofolate; TYMS, thymidylate synthase. **b-f** Highly controlled genetic pedigrees that demonstrated TEI in the *Mtrr*^{gt} mouse line[11], some of which are used in this study. Grey shaded boxes, pedigrees that do not display phenotypes at E10.5 in the final generation shown[11]. Pink shaded boxes, pedigrees that display a wide spectrum of phenotypes at E10.5 in the final generation shown[11]. **b** *Mtrr*$^{+/gt}$ maternal grandfather pedigree and *Mtrr*$^{+/gt}$ maternal grandmother pedigree. **c** Control C57Bl/6J pedigree. **d** Embryo transfer experiment demonstrating germline inheritance of a yet-to-be determined epigenetic factor to cause congenital malformation at E10.5. F2 wildtype blastocysts derived from an F0 *Mtrr*$^{+/gt}$ male or female were transferred into control B6D2F1 pseudopregnant females and demonstrated that congenital malformations and not growth phenotypes at E10.5 were independent of the F1 uterine environment[11]. **e** *Mtrr*$^{+/gt}$ intercross pedigree. **f** *Mtrr*$^{gt/gt}$ intercross pedigree. Pedigree legend: circle, female; square, male; blue outline, C57Bl/6J mouse line; black outline, *Mtrr*^{gt} mouse line; pink outline, B6D2F1 mouse line; white fill, *Mtrr*$^{+/+}$; half black-half white fill, *Mtrr*$^{+/gt}$; black fill, *Mtrr*$^{gt/gt}$.



Supplementary Figure 2 **Frequency and location of genetic variants in C57Bl/6J and *Mtrr*$^{gt/gt}$ embryos. a, b** Whole genome sequencing data showing the average frequency of **a** structural variants (SVs) and **b** single nucleotide polymorphisms (SNPs) for each chromosome. Phenotypically normal C57Bl/6J embryos (N=2, black bars) and severely affected *Mtrr*$^{gt/gt}$ embryos (N=6, blue bars) were assessed. Data is presented as mean ± standard deviation (sd). Note that the gene-trap insertion in the *Mtrr* locus is on chromosome 13. **c, d** Genomic location of **c** SVs and **d** SNPs in C57Bl/6J and *Mtrr*$^{gt/gt}$ embryos after masking the region surrounding the gene-trap insertion site. **e, f** UpsetR plots showing intersections of **e** SVs or **f** SNPs between individual C57Bl/6J and *Mtrr*$^{gt/gt}$ embryos. Arrows indicate the variants that are present in all *Mtrr*$^{gt/gt}$ embryos and not C57Bl/6J embryos. See also Supplementary Tables 1 and 2.

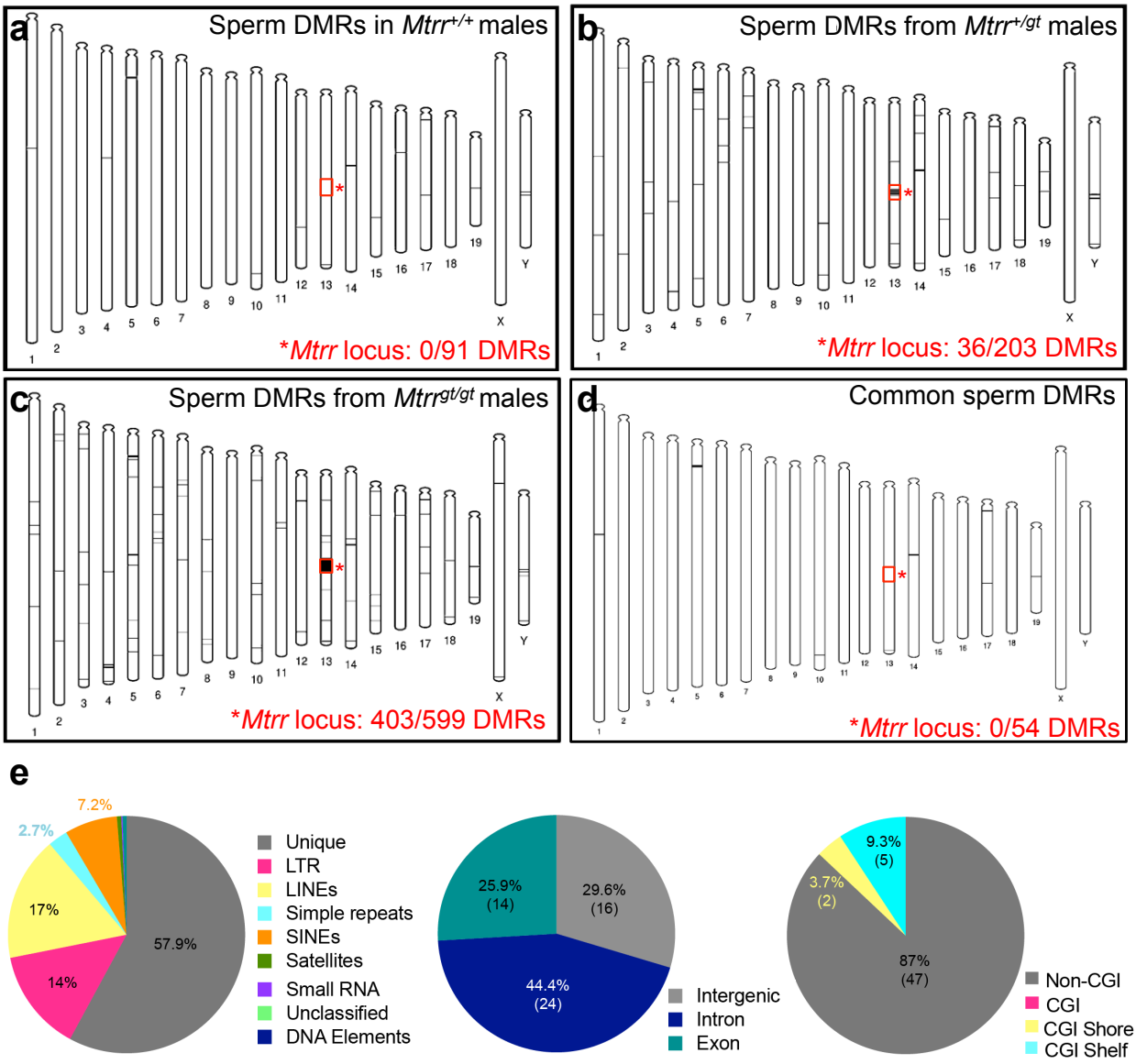


Supplementary Figure 3 **Confirmation of spermatozoa purity and validation of immunoprecipitation. a** Bisulphite pyrosequencing of imprinting control regions in DNA from spermatozoa collected from cauda epididymides to determine sperm purity. Percentage methylation at specific CpG sites in the maternally imprinted *Peg3* differentially methylated region (DMR) and paternally imprinted *H19* DMR were determined in sperm samples isolated from C57Bl/6J (black circles), wildtype ($Mtrr^{+/+}$; purple circles), $Mtrr^{+/gt}$ (green circles) and $Mtrr^{gt/gt}$ (blue circles) mice. C57Bl/6J liver (grey dots) was assessed as a control. Data is represented as mean ± standard deviation (sd) for each CpG site. N=8 males per group. **b** Percentage recovery of DNA input after MeDIP experiment as determined using RT-qPCR to amplify known methylated (*Nanog* and *H19*) and unmethylated (*H1t* and *TsH2B*) regions. MeDIP samples from sperm of C57Bl/6J (black bars), $Mtrr^{+/+}$ (purple bars), $Mtrr^{+/gt}$ (green bars), and $Mtrr^{gt/gt}$ (blue bars) males are shown. Each bar indicates one individual (N=8 males per genotype).

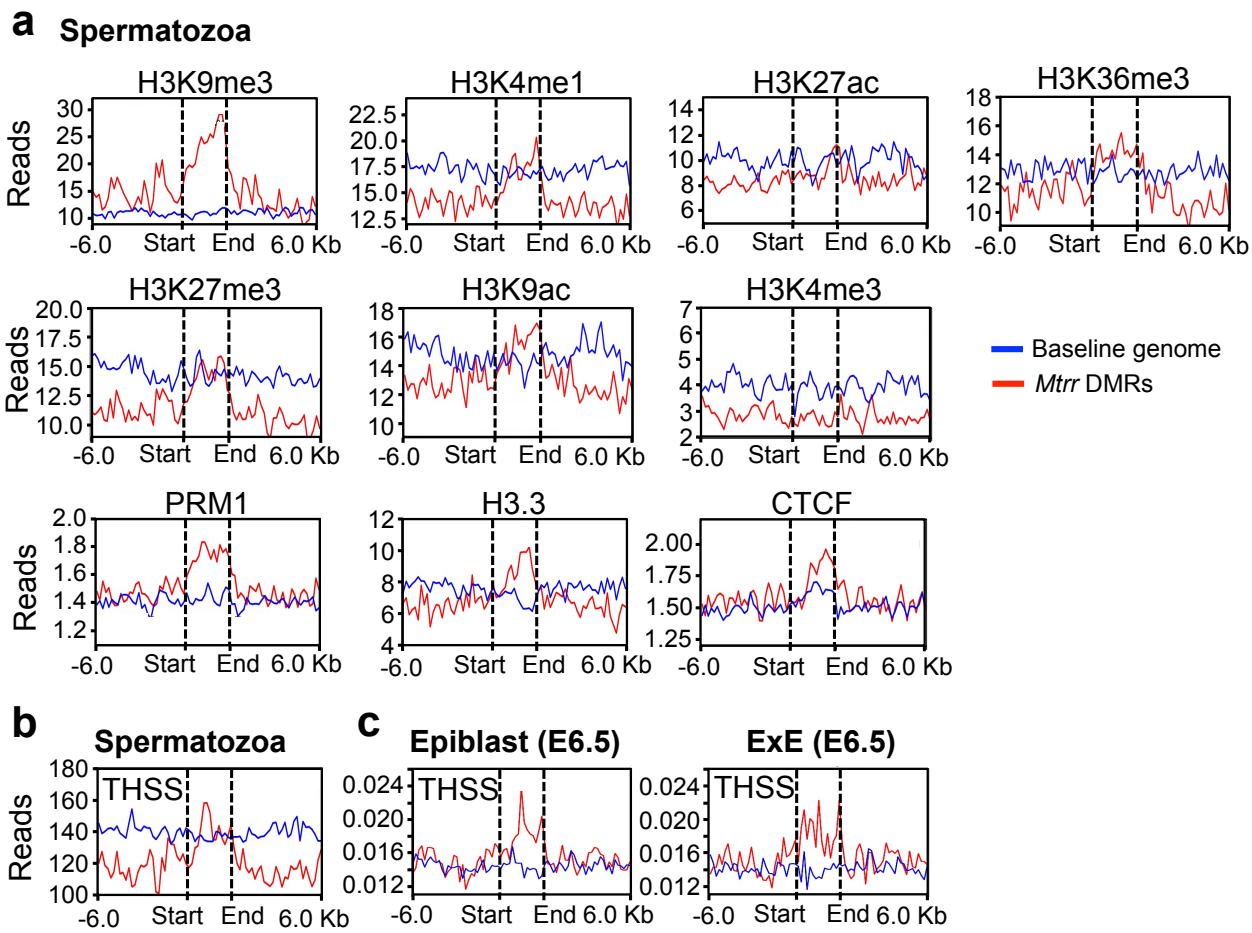


Supplementary Figure 4 **Validation of a panel of sperm DMRs identified in *Mtrr*$^{+/+}$, *Mtrr*$^{+/gt}$ and *Mtrr*$^{gt/gt}$ males. a-c** Validation of differentially methylated regions (DMRs) identified in a MeDIP-seq

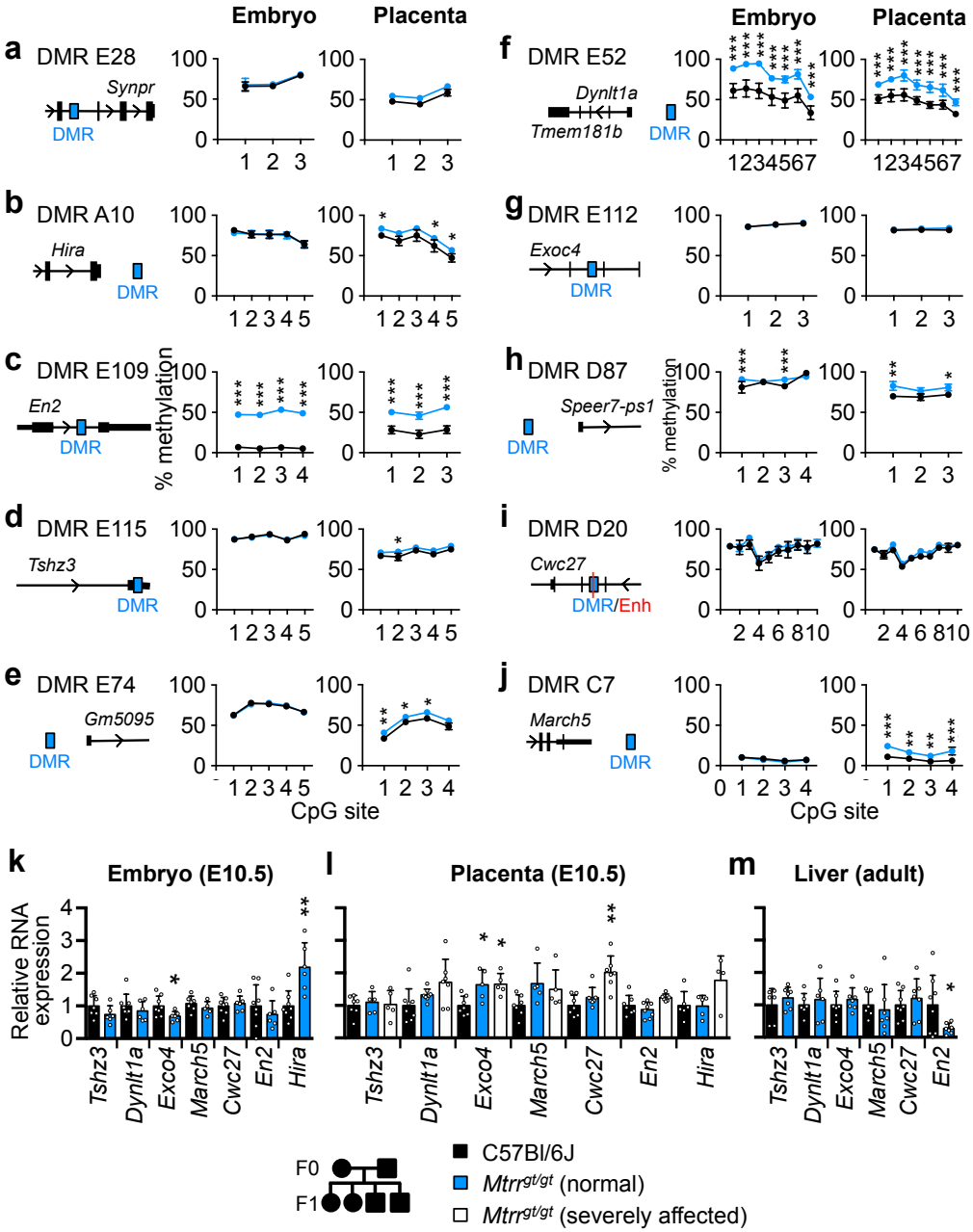
experiment of sperm from wildtype (*Mtrr*$^{+/+}$; purple circles), *Mtrr*$^{+/gt}$ (green circles) and *Mtrr*$^{gt/gt}$ (blue circles) males relative to C57Bl/6J control sperm (black circles). The average percentage methylation at individual CpG sites was determined by bisulfite pyrosequencing. Data is plotted as mean ± standard deviation (sd) at each CpG site. N=8 males per group (four samples from MeDIP-seq analysis plus four unique samples). The coordinates for each DMR are given. Two-way ANOVA, with Sidak's multiple comparisons test, performed on mean methylation per CpG site per genotype group. ***$p<0.001$.



Supplementary Figure 5 **Chromosomal distribution of DMRs in sperm of *Mtrr* males. a-d** Phenograms showing chromosomal location of sperm DMRs (black rectangles) identified via MeDIP-seq analysis in **a** wildtype (*Mtrr*$^{+/+}$), **b** *Mtrr*$^{+/gt}$, and **c** *Mtrr*$^{gt/gt}$ males compared to C57Bl/6J sperm. Red box indicates region on chromosome 13 surrounding *Mtrr* locus identified in Fig. 1a. **d** A phenogram depicting the chromosome location of the 54 common sperm DMRs between *Mtrr* genotypes when compared to C57Bl/6J sperm. **e** Relative genomic distribution of methylated regions among the 54 common sperm DMRs between *Mtrr*$^{+/+}$, *Mtrr*$^{+/gt}$ and *Mtrr*$^{gt/gt}$ males (compared to C57Bl/6 males) with respect to unique sequences and repetitive elements, coding and non-coding regions, and CpG islands (CGIs), shores and shelves. N=8 males per group.

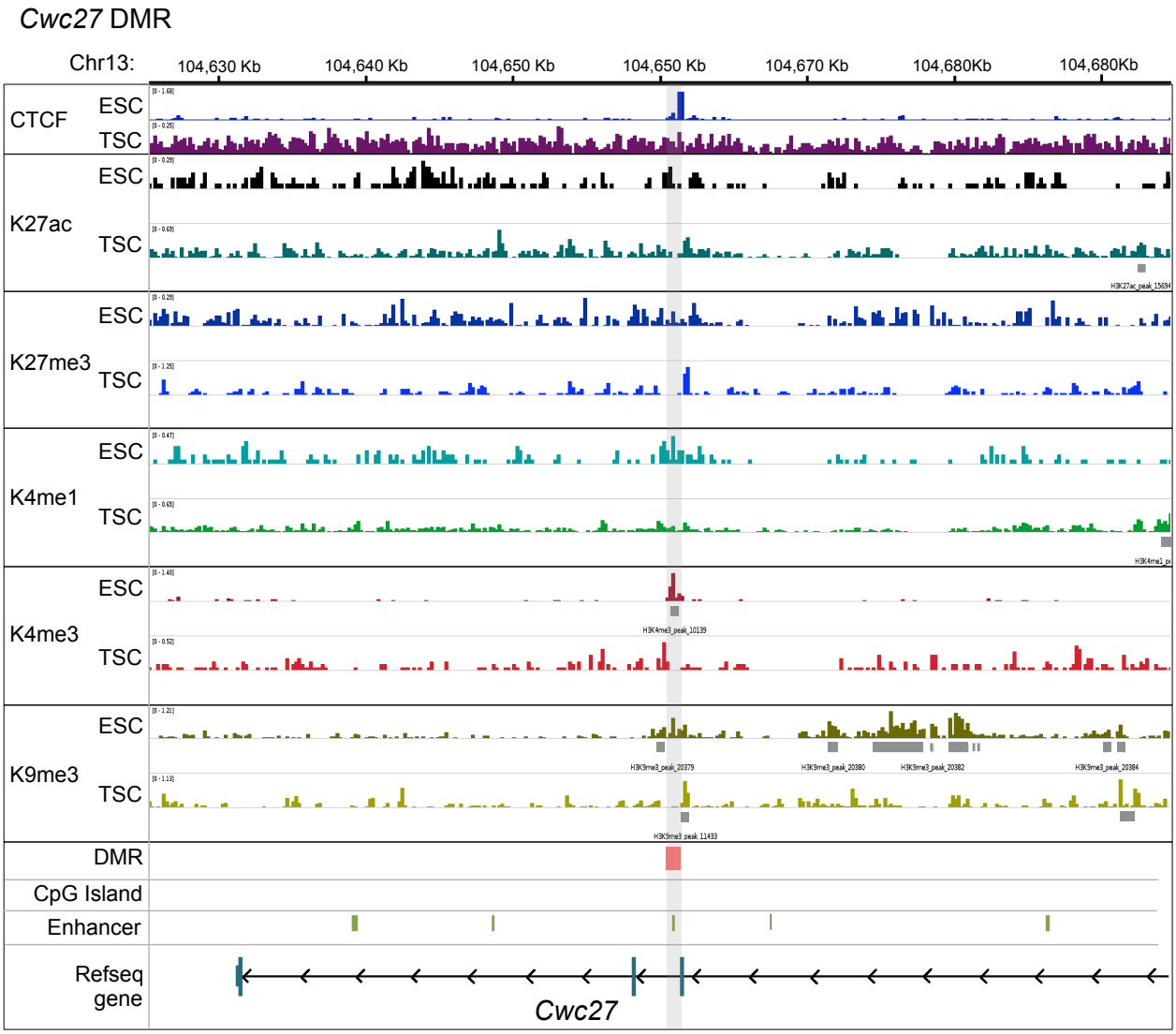


Supplementary Fig. 6 **General epigenetic signature of genomic regions identified as sperm DMRs in *Mtrr* males. a** Using published ChIP-seq data sets in wildtype CD1 spermatozoa[5], mean enrichment of selected histone modifications and DNA binding proteins was determined in the genomic regions identified as differentially methylated in sperm. Differentially methylated regions (DMRs) of all *Mtrr* genotypes, excluding those within the 20 Mb region surrounding the *Mtrr* locus, were combined in this analysis (N= 379 DMRs; red line) and compared to the baseline genome (blue line; see Methods). **b, c** Using published ATAC-seq data sets in **b** wildtype CD1 spermatozoa[5] and **c** wildtype B6D2F1 mouse epiblast and extraembryonic ectoderm (ExE) at embryonic day 6.5[6], mean enrichment of Tn5 transposase sensitive site (THSS) was determined in sperm DMRs of all *Mtrr* genotypes combined (N= 379 DMRs; red line) compared to the baseline genome (blue line). DMRs in the region surrounding the *Mtrr* gene-trap insertion site were masked in both analyses. Dotted lines indicate the start and end of the DMR. Six kilobases (Kb) of DNA surrounding the DMR was also considered.

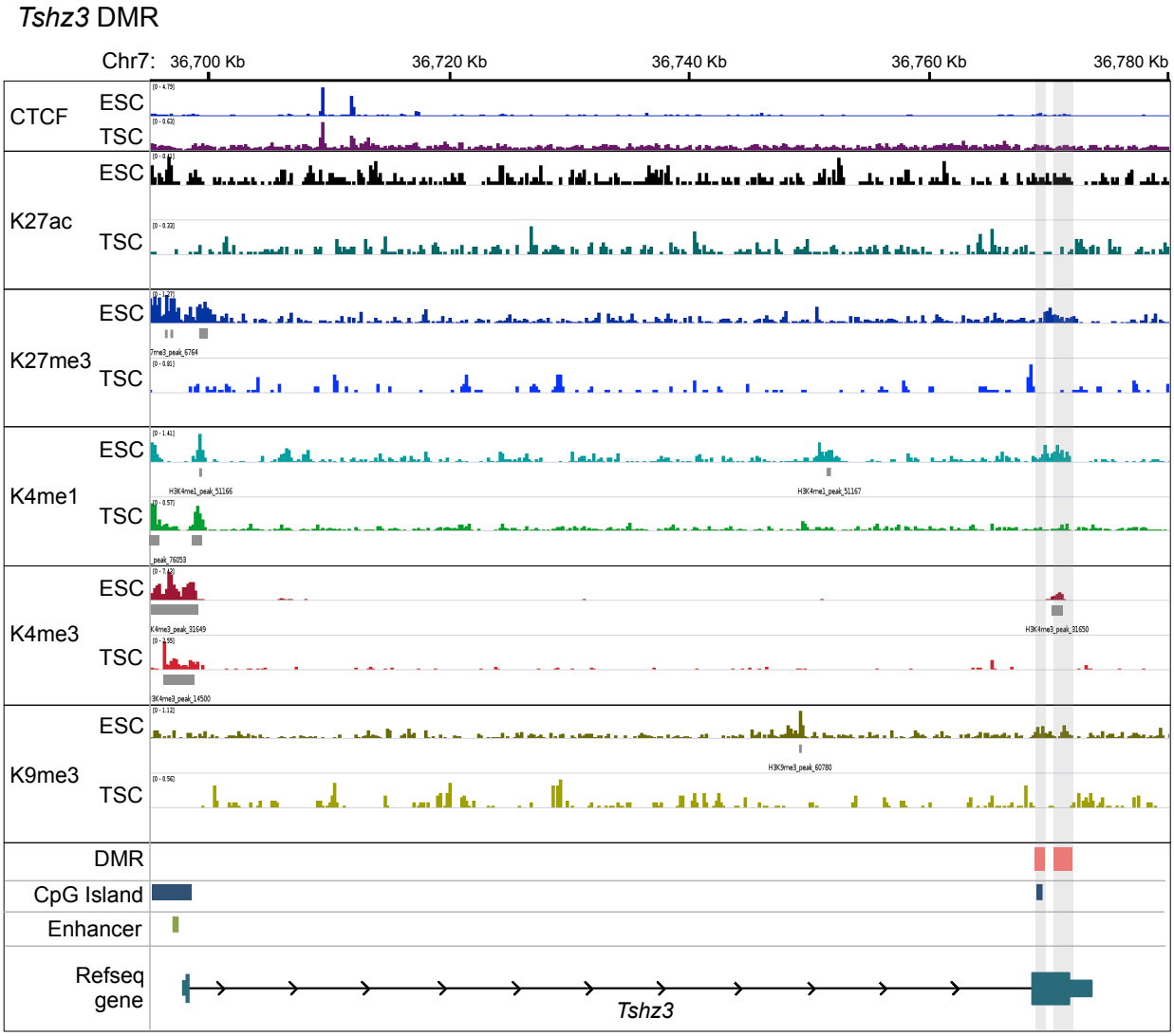


Supplementary Figure 7 **Analysis of DNA methylation and gene expression at sperm DMRs in *Mtrr*$^{gt/gt}$ tissue. a-j** Schematic drawings of differentially methylated regions (DMRs; blue rectangle) identified in *Mtrr*$^{+/gt}$ sperm in relation to the closest gene alongside bisulfite pyrosequencing analysis of CpG methylation at these DMRs in C57Bl/6J control (black circles) and *Mtrr*$^{gt/gt}$ (blue circles) phenotypically normal embryos and placentas at E10.5. The average percentage of methylation at individual CpG sites within the corresponding DMR is shown (mean ± standard deviation (sd) for each CpG site). Enhancer (Enh) overlap is shown in red. N=4 placentas and N=6-8 embryos assessed per experimental group. Two-way ANOVA, with Sidak's multiple

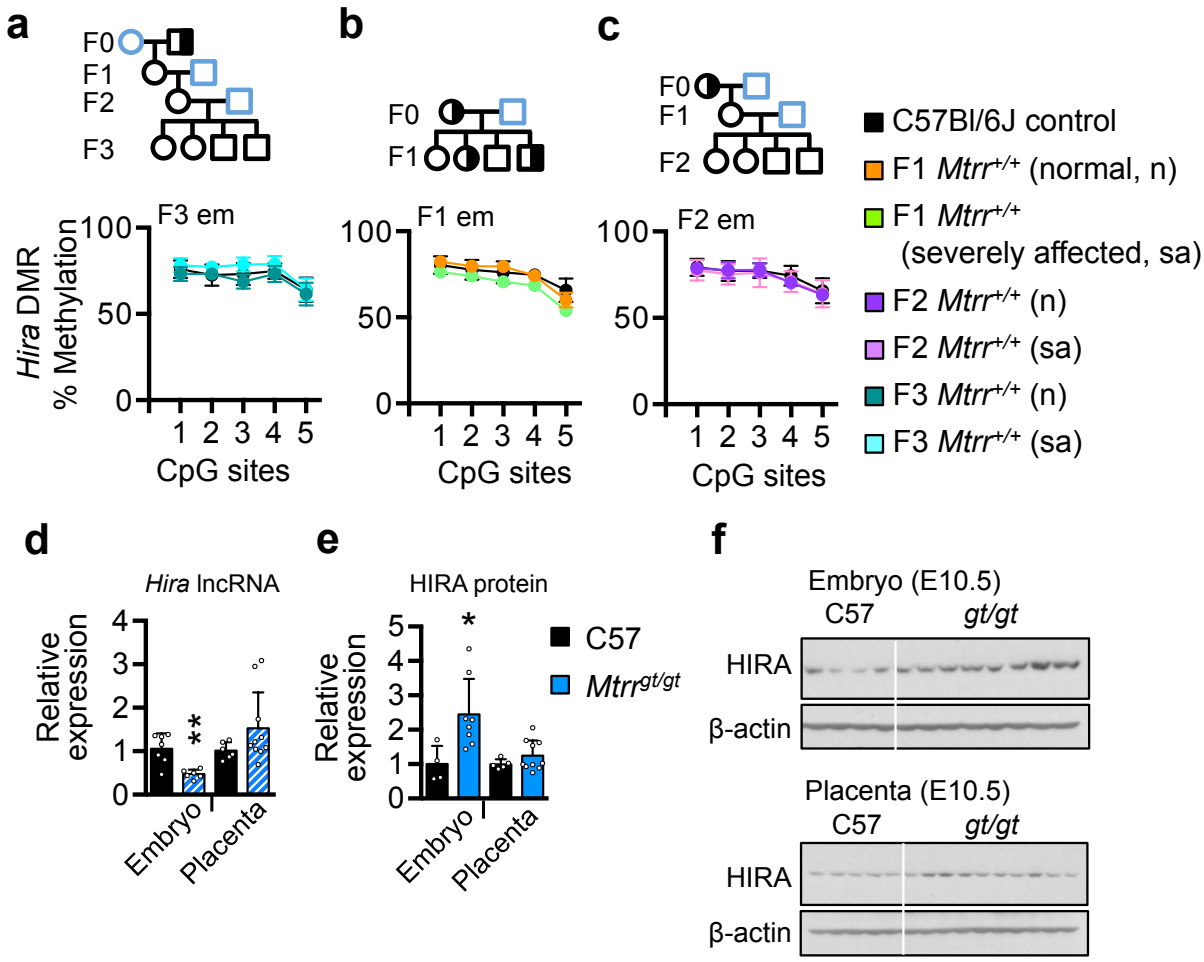
comparisons test, performed on mean methylation per CpG site per genotype group. *p<0.05, **p<0.01, ***p<0.001. **k-m** RT-qPCR analysis of mRNA expression of some of the genes proximal to or overlapping with sperm DMRs (shown in a-f). Gene expression was assessed in C57Bl/6J control (black bars) and *Mtrr*$^{gt/gt}$ (blue and white bars) **k** embryos and **l** placentas at E10.5, and **m** *Mtrr*$^{gt/gt}$ adult livers. N=4-8 individuals per experimental group. Tissue from phenotypically normal (blue bars) and severely affected (white bars) *Mtrr*$^{gt/gt}$ conceptuses was assessed. Expression data is plotted as mean ± sd, and relative to C57Bl/6J levels (normalized to 1). Independent t tests or one-way ANOVA with Dunnett's multiple comparisons tests were performed. *p<0.05, **p<0.01, ***p<0.001.



Supplementary Figure 8 **Epigenetic signature of the intragenic *Cwc27* DMR in wildtype mouse ESCs and TSCs.** Enrichment of DNA binding proteins (CTCF) and histone modifications (H3K27ac, H3K27me3, H3K4me1, H3K4me3, H3K9me3) in the *Cwc27* locus on chromosome (Chr) 13 (~37,000 kb downstream of *Mtrr* gene) using published ChIP-seq data sets[4] in wildtype embryonic stem cells (ESCs) and trophoblast stem cells (TSCs). Grey rectangles indicate enriched peaks for each histone mark. Red rectangle and grey shading indicate *Cwc27* differentially methylated region (DMR) identified in sperm of *Mtrr*$^{+/gt}$ males. Green rectangles indicate enhancer regions. Partial schematic of protein encoding *Cwc27* transcript is shown.



Supplementary Figure 9 **Epigenetic signature of the intragenic *Tshz3* DMR in wildtype ESCs and TSCs.** Enrichment of DNA binding proteins (CTCF) and histone modifications (H3K27ac, H3K27me3, H3K4me1, H3K4me3, H3K9me3) in the *Tshz3* locus on chromosome (Chr) 7 using published ChIP-seq data sets[4] in wildtype embryonic stem cells (ESCs) and trophoblast stem cells (TSCs). Grey rectangles indicate enriched peaks for each histone mark. Red rectangle and grey shading indicate *Tshz3* differentially methylated region (DMR) identified in sperm of $Mtrr^{+/gt}$ males. Blue rectangles indicate CpG islands. Green rectangle indicates enhancer regions. Schematic of protein encoding *Tshz3* transcript is shown.



Supplementary Figure 10 **Analysis of *Hira* DMR methylation and HIRA protein expression. a-c** Bisulfite pyrosequencing analysis of average percentage methylation at individual CpG sites in the *Hira* DMR (see Fig. 5) in C57Bl/6J embryos (N=6-8 embryos; black circles) at E10.5 compared to the following: **a** F3 wildtype ($Mtrr^{+/+}$) embryos (em) at E10.5 derived from F0 $Mtrr^{+/gt}$ males (N=4 phenotypically normal (n) embryos teal circles; N=3 severely affected (sa) embryos, turquoise circles), **b** F1 wildtype embryos at E10.5 derived from F0 $Mtrr^{+/gt}$ females (N=7 phenotypically normal (n) embryos, orange circles; N=3 severely affected (sa) embryos, green circles). **c** F2 wildtype embryos at E10.5 derived from F0 $Mtrr^{+/gt}$ females (N=5 phenotypically normal (n) embryos, purple circles; N=4 severely affected (sa) embryos, pink circles). Data is presented as average methylation (± standard deviation (sd)) per CpG site. Two-way ANOVA, with Sidak's multiple comparisons test, performed on mean methylation per CpG site. **d** RT-qPCR analysis of *Hira* lncRNA expression (striped bars, primer set 2) in embryos (N=6-8) and placentas (N=5-8) from C57Bl/6J (black bars) and $Mtrr^{gt/gt}$ conceptuses (blue bars). Independent t test, **p<0.01. **e, f** Western blot analysis of HIRA protein in C57Bl/6 (black bars) and $Mtrr^{gt/gt}$ (blue bars) embryos

(N=4-8) and placentas (N=5-10) at E10.5. Blots are shown in **f**. HIRA protein levels were normalised to β-actin loading control. Independent t test, *p<0.05, **p<0.01.

SUPPLEMENTARY REFERENCES

1. Watkins-Chow, D.E. & Pavan, W.J. Genomic copy number and expression variation within the C57BL/6J inbred mouse strain. *Genome Res* **18**, 60-6 (2008).
2. Hackett, J.A. *et al.* Germline DNA demethylation dynamics and imprint erasure through 5-hydroxymethylcytosine. *Science* **339**, 448-52 (2013).
3. Kobayashi, H. *et al.* Contribution of intragenic DNA methylation in mouse gametic DNA methylomes to establish oocyte-specific heritable marks. *PLoS Genet* **8**, e1002440 (2012).
4. Schoenfelder, S. *et al.* Divergent wiring of repressive and active chromatin interactions between mouse embryonic and trophoblast lineages. *Nat Commun* **9**, 4189 (2018).
5. Jung, Y.H. *et al.* Chromatin States in Mouse Sperm Correlate with Embryonic and Adult Regulatory Landscapes. *Cell Rep* **18**, 1366-1382 (2017).
6. Smith, Z.D. *et al.* Epigenetic restriction of extraembryonic lineages mirrors the somatic transition to cancer. *Nature* **549**, 543-547 (2017).
7. Gillich, A. *et al.* Epiblast stem cell-based system reveals reprogramming synergy of germline factors. *Cell Stem Cell* **10**, 425-39 (2012).
8. Rameix-Welti, M.A. *et al.* Visualizing the replication of respiratory syncytial virus in cells and in living mice. *Nat Commun* **5**, 5104 (2014).
9. Kim, S. *et al.* PRMT5 protects genomic integrity during global DNA demethylation in primordial germ cells and preimplantation embryos. *Mol Cell* **56**, 564-79 (2014).
10. Tuorto, F. *et al.* RNA cytosine methylation by Dnmt2 and NSun2 promotes tRNA stability and protein synthesis. *Nat Struct Mol Biol* **19**, 900-5 (2012).
11. Padmanabhan, N. *et al.* Mutation in folate metabolism causes epigenetic instability and transgenerational effects on development. *Cell* **155**, 81-93 (2013).