# The Epiallelic Nature of Mouse Ribosomal DNA

Francisco Rodriguez-Algarra[1], Amy F Danson[1]†, Robert A E Seaborne[1]†, Selin Yildizoglu[1]†, Harunori Yoshikawa[2], Pui Pik Law[1,3], Zakaryya Ahmad[1], Victoria A Maudsley[1], Ama Brew[1], Nadine Holmes[4], Alan Hodgkinson[3], Sarah J Marzi[5,6], Madapura M Pradeepa[1], Matthew Loose[4], Michelle L Holland[3], Vardhman K Rakyan[1]*

[1]The Blizard Institute, School of Medicine and Dentistry, Queen Mary University of London; UK.
[2]Fujii Memorial Institute of Medical Sciences, Institute of Advanced Medical Sciences, Tokushima University; Tokushima, Japan
[3]Department of Medical and Molecular Genetics; School of Basic and Medical Biosciences, King's College London; UK.
[4]DeepSeq, School of Life Sciences, University of Nottingham; UK.
[5]UK Dementia Research Institute, Imperial College London; UK.
[6]Department of Brain Sciences, Imperial College London; UK.

*Corresponding author (v.rakyan@qmul.ac.uk) †Equal contributions.

**Summary**

Ribosomal DNA (rDNA) displays substantial inter-individual genetic variation in human and mouse. Here we report that 45S rDNA units in the C57BL/6J mouse strain are epiallelic, existing as distinct genetic haplotypes that influence the epigenetic state and transcriptional output of any given unit. Epigenetic dynamics at these haplotypes are dichotomous and lifestage specific: at one haplotype, the DNA methylation state is sensitive to the *in utero* environment, but refractory to post-weaning influences, whereas other haplotypes entropically gain DNA methylation during ageing only. rDNA epiallelism is influenced by total rDNA copy number, and also found in other inbred mouse strains and humans. In the future, it will be important to consider the impact of inter-individual rDNA (epi)genetic variation on mammalian phenotypes and diseases.

## Introduction

The ribosome is one of the fundamental macromolecular complexes in all living cells, enabling translation to occur in the cytoplasm. The mature mammalian 80S ribosome consists of small (40S) and large (60S) subunits, both comprised of a different complement of proteins and RNA. Despite the essential and highly conserved role played by the ribosome, it is now clear that it can be compositionally diverse even within a given individual organism[1]. It has been shown that the protein composition of the ribosome can vary developmentally[1], and that the ribosomal RNA (rRNA) components display inter- and intra-individual genetic variation[2,3]. Such variation is thought to ultimately influence preferential translation of some mRNAs i.e. the "ribosome filter hypothesis"[4].

With respect to rRNA variation in mammals, several studies have reported genetic variation within the 45S rDNA in human and mouse (**Fig. 1A)**. The 45S rDNA codes for the 18S rRNA that is incorporated into the 40S subunit, and the 5.8S and 28S rRNAs that are incorporated into the 60S subunit (which additionally contains the 5S rRNA that is coded by the 5S rDNA present at an unlinked genomic location). Because of the multi-copy and multi-allelic nature of 45S rDNA, functional genomic analyses of this source of mammalian genetic variation has been difficult. Indeed, rDNA clusters are typically excluded from genome assemblies, and none of the large-scale genomic analyses of recent years have yielded any insights into rDNA genetic variation. Nonetheless, several smaller studies have noted not only inter-individual genetic variation (in both sequence and copy number) in mouse and human rDNA, but also epigenetic variation in the context of whole organism stress responses[2,5-8]. The latter observation raises the intriguing possibility that the impact of mammalian rDNA sequence variation is also relevant at the DNA level, thereby impacting transcriptional outputs in different biological contexts. Our aim in this study was to establish a deeper understanding of the functional genomic impact of rDNA genetic variation in a mammalian model.

## Results

**Mouse rDNA exists as distinct genetic haplotypes.** Using the C57BL/6J mouse strain, we focussed on a ~15kb region of the rDNA termed the 'coding unit' (**Fig. 1A,** the intergenic sequence, IGS, contains high repeat sequence density and was not analysed in detail). Short-read whole genome sequencing (WGS) analysis of four different C57BL/6J mice identified 88 different coding unit intra- and inter-individual single nucleotide variants (SNVs) (**Supplementary Materials, table S1,** inclusion of indels did not affect SNV allele frequencies and were not explicitly considered further). To achieve a deeper understanding of rDNA genomic architecture, we sequenced a C57BL/6J mouse embryonic fibroblast (MEF) line using ultra-long read whole genome Nanopore technology[9], obtaining 932,683 reads (N50 ~72 kb), of which 1,760 contained one or more rDNA coding units

within a single read (**Supplementary Materials, table S2**). We confirmed 87/88 short-read rDNA coding unit SNVs in these reads (**table S1**). Previously, we reported a SNV at position -104 (relative to TSS, **Fig. 1A**) in the C57BL/6J rDNA that is associated with differential promoter methylation: 'A' variants at -104 are associated with 30-80% methylation, whereas 'C' variants display <25% methylation at the promoter[6,7] (**fig. S1**). Using -104 as a starting point to explore the possibility of larger haplotypes within the rDNA, analysis of the SNVs throughout the coding unit revealed 4 different rDNA haplotypes, that we term 'ATA', 'ATG', 'CCA', 'CTA', in approximately equal proportions (**Fig. 1B**). ATA and ATG are more genetically similar to each other compared to either CCA or CTA (**Fig. 1B, Supplementary Materials**). Independent support for these haplotypes was obtained by pairwise correlation analysis of the relevant haplotype-associated SNVs in the four different short-read kidney WGS datasets (**fig. S2, table S3**). Analysis of reads containing two or more complete rDNA units revealed that rDNA haplotypes are found adjacent to each other (**Fig. 1C**). Although further refinement of C57BL/6J rDNA haplotypes may be possible, the analyses below demonstrate that the rDNA haplotypes underlie bona fide molecular differences.

**The impact of rDNA haplotypes on functional genomic outcomes**. Nanopore sequencing permits direct assessment of DNA methylation in unamplified DNA. Strikingly, we found that the ATA haplotype displays significant DNA methylation (≳60%) across the length of the coding unit, CCA shows low methylation levels (≲20%), and the other two haplotypes are largely unmethylated (**Fig. 2A**). Analysis of individual reads revealed that individual coding units are either almost completely methylated or unmethylated (**fig. S3**), and haplotype-specific methylation differences do not extend into the adjoining IGS regions, which are generally hypermethylated (**fig. S3**). To provide further support for rDNA haplotype-specific methylation using an orthologous method, we analysed 4 different kidney whole-genome bisulfite sequencing (WGBS) and 7 different sperm reduced representation bisulfite sequencing (RRBS) C57BL/6J datasets (11 different mice) (**table S3**). Short-read methods cannot directly define long-range (epi)genetic patterns, and in our case the major ATA-defining SNV is an 'A' at position 6832. However, if the proposed (epi)genomic architecture of the haplotypes is correct, then if we combine the ATA and ATG haplotypes into a single 'A' haplogroup, that can be distinguished from the CCA and CTA haplotypes at multiple positions, then in short-read data the allele frequency and DNA methylation state at these positions should be predictable based on the combined individual allele frequencies and DNA methylation levels of ATA and ATG at position 6832. This is indeed what we observe (**Fig. 2B**, **fig. S4**). The WGBS and RRBS datasets also demonstrate that whilst relative allele frequencies and DNA methylation show inter-individual variation, only the ATA rDNA haplotype shows substantial methylation *in vivo*.

DNA methylation at the mouse rDNA promoter has a strong effect on rRNA expression[10]. If methylation is restricted largely to ATA haplotypes, then methylation observed at the rDNA promoter of 'A' haplogroup should derive from ATA haplotypes only, and correlate with the relative frequency of ATA in the rRNA. We therefore performed 'rRNA-seq', which omits the rRNA depletion step in the standard mRNA-seq protocol (**Supplementary Materials**). Also, because of the sheer abundance of cellular rRNA, even internal transcribed spacer (ITS) RNA is readily detected (ITS2 contains the ATA-defining SNV at 6832). We observed a strong negative correlation between 'A' haplogroup promoter methylation levels and the relative frequency of A's at position 6832 in muscle rRNA-seq (**Fig. 2C, table S4**). In fact, methylation even in the vicinity of position 6832 in the rDNA should be correlated with the variant ratios observed at position 6832 in rRNA-seq, which is what we observe using 4 different kidney WGBS datasets (**Fig. 2C**). Next, a CUT&Tag analysis of kidney samples showed that only ATA haplotypes display combined H3K9me3 enrichment/H3K27me3 depletion, consistent with previous reports of the relationship between DNA methylation and these histone modifications in other genomic regions[11] (**Fig. 2D, fig. S5, table S5**). Finally, it is known that the Upstream Binding Transcriptional factor (UBTF) binding to unmethylated rDNA is required for rRNA expression[12]. Re-analysis of a previously published dataset of UBTF binding in B-cells of C57BL/6J mice[13], showed that ATA is depleted for UBTF relative to the non-ATA haplotypes (**Fig. 2E**). Collectively, these analyses demonstrate key functional genomic features of ATA vs non-ATA rDNA haplotypes in a variety of tissue types.

**DNA methylation dynamics at rDNA are dichotomous and life-stage specific.** Epigenetic silencing of rDNA is an integral component of the stress response in all eukaryotic cells. Recent studies using the C57BL/6J strain have shown that *in utero* exposure to nutritional stress, e.g. maternal protein restriction, obesogenic or high-fat diet, induces DNA hypermethylation of rDNA that persists into adulthood[6,7,14]. We re-analysed our previous RRBS data from C57BL/6J individuals exposed to maternal protein restriction *in utero*[6], and found that the hypermethylation occurs specifically at the ATA variants (**Fig. 3A, fig. S6**). In the GEO database, RRBS data is available for one other pure C57BL/6J mouse model of *in utero* nutritional stress in which the offspring were exposed to a high fat diet during gestation+lactation, or during gestation+lactation+post-weaning[15]. In both experimental groups, relative to controls, rDNA hypermethylation is observed at ATA variants only (**Fig. 3A**). The *in utero* nutritional stress models demonstrate that certain stimuli methylate and silence ATA rDNA units. But, can ATA units also lose methylation and re-activate? We re-analysed Dahlet et al.'s RRBS data of C57BL/6J CRISPR-based Dnmt1 knock out 8.5dpc embryos[16], and found that Dnmt1 KO embryos show virtually no methylation at any of the rDNA

105 haplotypes (**Fig. 3B left panel, fig. S7**). In standard mRNA-seq protocols, rRNA is specifically
106 removed. However, this depletion is never completely efficient and we find millions of rRNA reads
107 in the Dahlet et al. mRNA-seq data (**Supplementary Materials**). Our own analysis of 22 matched
108 C57BL/6J muscle mRNA-seq and rRNA-seq revealed an excellent correlation of rRNA variant
109 frequencies ($R \geq 0.98$ in all cases, **fig. S8**). We then examined rRNA expression in Dahlet et al.'s
110 mRNA-Seq data and found that ATA variants are expressed at considerably higher relative levels in
111 the Dnmt1 KOs, proving that methylated ATA variants are not irreversibly silenced (**Fig. 3B right**
112 **panel**).
113
114 Interestingly, previous studies show that a low protein diet in C57BL/6J mice post-weaning shows
115 little, if any, rDNA hypermethylation[7,17]. The study by Cannon et al. also includes a fourth group in
116 which the mice were exposed to a high-fat diet post-weaning only, and in this group evidence for
117 diet-induced epigenetic differences at rDNA is weaker[15] **(fig. S6**). To further explore the idea that the
118 epigenetic state of ATA rDNA shows only limited dynamics in adulthood, we leveraged the RRBS
119 dataset of Petkovich et al., 2016, representing 193 different C57BL/6J mice spanning an age range
120 from 3-35 months[18]. Ageing-associated DNA methylation dynamics are observed genome-wide in a
121 large range of mammalian species[19-21], including at rDNA in humans and rodents[22-24]. In the
122 Petkovich et al. dataset, ATA showed no directional change with age (**Fig. 3C, fig. S9**). However,
123 methylation at non-ATA haplotypes, and in particular CCA and CTA, display a positive correlation
124 with ageing (**Fig. 3C**). By leveraging single-molecule level data, we were able to show that this was
125 primarily driven by an increase in DNA methylation entropy within individual DNA molecules at
126 non-ATA haplotypes (**Fig. 3D, fig. S10, S11**) [26-28]. Therefore, epigenetic dynamics at rDNA
127 haplotypes are life-stage specific and dichotomous: the ATA haplotype displays environmentally-
128 induced epigenetic dynamics during early development, but is less susceptible to further perturbations
129 in later life. On the other hand, non-ATA haplotypes entropically accumulate methylation during
130 ageing.
131
132 **rDNA epialleles are present in other mouse strains and human**. We then studied 5 additional
133 inbred mouse strains (**Fig. 4A**). For each strain, we generated kidney WGS, WGBS, rRNA-seq, and
134 droplet digital PCR (ddPCR) data from each of 6 different adult males (117 different datasets as 3
135 were discarded post-QC) (**table S3**). We confirmed that for any given sample, >82% of the rDNA
136 coding unit SNVs called in the WGS data were also found in the WGBS data (**table S6-10, fig. S12**).
137 We then asked which of these SNVs are associated with allelic methylation differences, i.e., are
138 'epivariants' (Wilcoxon rank sum test, FDR < 0.01), finding both strain-specific and multi-strain
139 epivariants (**Fig. 4A, table S6-10**). Analysis of matched rRNA-seq data showed that epivariant-

140 associated methylation differences throughout the coding unit impact variant frequencies in the
141 rRNA, with the exceptions being 129S1/SvImJ and C3H/HeJ (**Fig. 4B, fig. S13**). We noted that these
142 two strains also showed the lowest levels of rDNA methylation (**Fig. 4A**). Given the known positive
143 correlation between increasing total rDNA copy number and epigenetic silencing in lower
144 organisms[29], we considered the possibility of copy number (CN) effects and calculated CN from three
145 independent datasets - WGS, WGBS and ddPCR - for each individual mouse, as CN measurements
146 in mammals are known to be technically challenging[30] (see Methods for CN calculations, **fig. S14**).
147 Indeed, we found a significant positive correlation between total rDNA CN and DNA methylation
148 across the 6 different inbred strains (**Fig. 4C, fig. S14**, note: intra-strain CN variation was
149 considerably smaller and we did not have the sensitivity to examine CN vs methylation relationships
150 within a strain). Finally, we generated WGBS and rRNA-seq data for 24 different human
151 lymphoblastoid cell lines (LCLs) derived from 'Gambian in Western Division – Mandinka'
152 individuals that were sequenced in the 1000 genomes project[31]. We first confirmed that for any given
153 sample, >92% of the rDNA coding region SNVs called in the published WGS data were also called
154 in our WGBS data (**table S11**). Even in this relatively small number of samples for an outbred
155 population, we identified an epivariant at FDR <0.01. DNA methylation at this site (located within
156 the 28S), was negatively associated with variant representation in rRNA and positively associated
157 with total rDNA CN (**Fig. 4D**).
158

159 **Discussion**

160 Our data support the idea that mammalian rDNA is fundamentally epiallelic in nature. The genetic
161 identity of a rDNA unit has a strong probabilistic influence on the homeostatic epigenetic state, and
162 epigenetic dynamics at different life stages. The CN data suggests that the emergence of rDNA
163 epialleles could be linked to the need of the genome to silence extra copies of rDNA. However, the
164 silencing is not absolute, reminiscent of the epigenetic metastability associated with some
165 mammalian retroelements[32,33]. Could the loss/gain of specific rDNA variants underlie the CN
166 differences? It has been shown that the number of different chromosomes harbouring rDNA clusters
167 can vary from 3-5 amongst different inbred mouse strains[32]. The C57BL/6J genome contains rDNA
168 clusters on 4 separate chromosomes[33], leading to the possibility that each chromosome harbours a
169 different variant given that we find 4 different rDNA haplotypes in the C57BL/6J strain. Interestingly,
170 the C3H strain has only 3 rDNA clusters[34], and shows very few DNA methylation-based rDNA
171 epialleles. But could non-rDNA methylation based epialleles exist? A CUT&Tag analysis for
172 H3K9me3 and H3K27me3 in the 129S1/SvImJ mice did not yield any variant-specific clusters (**fig.
173 S15**). This will require further investigation, as will determining whether the human rDNA epialleles
174 we report also display similar haplotypic structures. In this regard it is worth noting that Parks et al

175 reported that a significant number of rRNA variants are evolutionarily conserved between human and
176 mouse[2].
177
178 With regards to the molecular consequences of rDNA (epi)genetic variation in models of nutritional
179 stress and ageing, a tantalizing possibility is the production of variant ribosomes that have subtle but
180 measurable effects on translational output[1]. It has been shown that endogenous rRNA sequence
181 variation regulates stress response gene expression in bacteria[35,36]. Previous studies have shown that
182 human and mouse coding subunit rDNA genetic variation is located in rRNA regions with known
183 roles in translation[1]. An initial Polysome-seq analysis shows that the ratio of 28S rRNA variants
184 incorporated into ribosomes is dependent on the ratio at which they are expressed (**fig. S16**). In
185 eukaryotes, there are examples of variation in either ribosomal proteins or rRNA modifications
186 influencing preferential translation of some mRNAs[1], yet the contribution of rRNA sequence
187 variation has remained largely unexplored and will require large-scale translatome analyses in the
188 future. In summary, our work highlights a source of mammalian genetic variation that is largely
189 overlooked, and it will be interesting to explore the possibility that rDNA genetic variation underlies
190 some apparently 'epigenetic' phenomena[37] and/or contributes, in trans, to the 'missing heritability'
191 in some human phenotypes and diseases.

**References:**


1. Genuth NR, Barna M. The Discovery of Ribosome Heterogeneity and Its Implications for Gene Regulation and Organismal Life. *Mol Cell.* **71**(3):364-374 (2018).
2. Parks MM, Kurylo CM, Dass RA, Bojmar L, Lyden D, Vincent CT, Blanchard SC. Variant ribosomal RNA alleles are conserved and exhibit tissue-specific expression. Sci Adv. 4(2):eaao0665 (2018).
3. Tseng H, Chou W, Wang J, Zhang X, Zhang S, Schultz RM. Mouse ribosomal RNA genes contain multiple differentially regulated variants. PLoS One. 3(3):e1843 (2008).
4. Mauro, VP, Matsuda D. Translation regulation by ribosomes: Increased complexity and expanded scope. *RNA Biol* **13**, 748-755 (2016).
5. Gibbons JG, Branco AT, Yu S, Lemos B. Ribosomal DNA Copy Number Is Coupled with Gene Expression Variation and Mitochondrial Abundance in Humans. *Nat Commun*. **11**(5):4850 (2011).
6. Holland ML, Lowe R, Caton PW, Gemma C, Carbajosa G, Danson AF, Carpenter AA, Loche E, Ozanne SE, Rakyan VK. Early-life nutrition modulates the epigenetic state of specific rDNA genetic variants in mice. *Science*. **353**(6298):495-8 (2016).
7. Danson AF, Marzi SJ, Lowe R, Holland ML, Rakyan VK. Early life diet conditions the molecular response to post-weaning protein restriction in the mouse. *BMC Biol.* **16**(1):51 (2018).
8. F. Shao, X. Liu, X. Zhang, Q. Wang, W. Wang Methylation of 45S Ribosomal DNA (rDNA) Is Associated with Cancer and Aging in Humans. *Int J Genomics* 2021, 8818007 (2021).
9. Jain M, Koren S, Miga KH, Quick J, Rand AC, Sasani TA, Tyson JR, Beggs AD, Dilthey AT, Fiddes IT, Malla S, Marriott H, Nieto T, O'Grady J, Olsen HE, Pedersen BS, Rhie A, Richardson H, Quinlan AR, Snutch TP, Tee L, Paten B, Phillippy AM, Simpson JT, Loman NJ, Loose M. Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nat Biotechnol.* **36**(4):338-45 (2018).
10. Santoro R, Grummt I. Molecular mechanisms mediating methylation-dependent silencing of ribosomal gene transcription. *Mol Cell.* **8**(3):719-25 (2001).
11. Rose NR, Klose RJ. Understanding the relationship between DNA methylation and histone lysine methylation. *Biochim Biophys Acta.* **1839**(12):1362-72 (2014).
12. O'Sullivan AC, Sullivan GJ, McStay B. UBF binding in vivo is not restricted to regulatory sequences within the vertebrate ribosomal DNA repeat. *Mol Cell Biol.* **22**(2):657-68 (2002).
13. Diesch J, Bywater MJ, Sanij E, Cameron DP, Schierding W, Brajanovski N, Son J, Sornkom J, Hein N, Evers M, Pearson RB, McArthur GA, Ganley ARD, et al. Changes in long-range rDNA-genomic interactions associate with altered RNA polymerase II gene programs during malignant transformation. *Commun Biol.* **2**:39 (2019).

*14.* Denisenko O, Lucas ES, Sun C, Watkins AJ, Mar D, Bomsztyk K, Fleming TP. Regulation of ribosomal RNA expression across the lifespan is fine-tuned by maternal diet before implantation. *Biochim Biophys Acta.* **1859**(7):906-13 (2016).

*15.* Cannon MV, Buchner DA, Hester J, Miller H, Sehayek E, Nadeau JH, Serre D. Maternal nutrition induces pervasive gene expression changes but no detectable DNA methylation differences in the liver of adult offspring. *PLoS One*. **9**(3):e90335 (2014).

*16.* Dahlet T, Argüeso Lleida A, Al Adhami H, Dumas M, Bender A, Ngondo RP, Tanguy M, Vallet J, Auclair G, Bardet AF, Weber M. Genome-wide analysis in the mouse embryo reveals the importance of DNA methylation for transcription integrity. *Nat Commun*. **11**(1):3153. (2020).

*17.* Shea JM, Serra RW, Carone BR, Shulha HP, Kucukural A, Ziller MJ, Vallaster MP, Gu H, Tapper AR, Gardner PD, Meissner A, Garber M, Rando OJ. Genetic and Epigenetic Variation, but Not Diet, Shape the Sperm Methylome. *Dev Cell.* **35**(6):750-8 (2015).

*18.* Petkovich DA, Podolskiy DI, Lobanov AV, Lee SG, Miller RA, Gladyshev VN. Using DNA Methylation Profiling to Evaluate Biological Age and Longevity Interventions. *Cell Metab.* **25**(4):954-960.e6. (2017).

*19.* Horvath S. DNA methylation age of human tissues and cell types. *Genome Biol*. **14**(10):R115 (2013).

*20.* Hannum G, Guinney J, Zhao L, Zhang L, Hughes G, Sadda S, Klotzle B, Bibikova M, Fan JB, Gao Y, Deconde R, Chen M, Rajapakse I, Friend S, Ideker T, Zhang K. Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol Cell.* **49**(2):359-67 (2013).

*21.* Mammalian Methylation Consortium. *www.biorxiv.org/content/10.1101/2021.01.18.426733v1.*

*22.* Swisshelm K, Disteche CM, Thorvaldsen J, Nelson A, Salk D. Age-related increase in methylation of ribosomal genes and inactivation of chromosome-specific rRNA gene clusters in mouse. *Mutat Res.* **237**(3-4):131-46 (1990).

*23.* Oakes CC, Smiraglia DJ, Plass C, Trasler JM, Robaire B. Aging results in hypermethylation of ribosomal DNA in sperm and liver of male rats. *Proc Natl Acad Sci U S A.* **100**(4):1775-80 (2003).

*24.* D'Aquila P, Montesanto A, Mandalà M, Garasto S, Mari V, Corsonello A, Bellizzi D, Passarino G. Methylation of the ribosomal RNA gene promoter is associated with aging and age-related decline. *Aging Cell*. **16**(5):966-75 (2017).

*25.* Wang M, Lemos B. Ribosomal DNA harbors an evolutionarily conserved clock of biological aging. *Genome Res.* **29**(3):325-33 (2019).

*26.* Landan G, Mendelson Cohen N, Mukamel Z, Bar A, Molchadsky A, Brosh R, Horn-Saban S, Amann Zalcenstein D, Goldfinger N, Zundelevich A et al. Epigenetic polymorphism and the stochastic formation of differentially methylated regions in normal and cancerous tissues. *Nat. Genet.* **44**:1207–1214 (2012).

*27.* Xie H, Wang M, de Andrade A, Bonaldo MDF, Galat V, Arndt K, Rajaram V, Goldman S, Tomita T, Soares MB. Genome-wide quantitative assessment of variation in DNA methylation patterns. *Nucleic Acids Res.* **39**:4099–4108 (2011).

*28.* Scherer M, Nebel A, Franke A, Walter J, Lengauer T, Bock C, Müller F, List M. Quantitative comparison of within-sample heterogeneity scores for DNA methylation data. *Nucleic Acids Res.* **48**(8):e46 (2020).

*29.* Jack CV, Cruz C, Hull RM, Keller MA, Ralser M, Houseley J. Regulation of ribosomal DNA amplification by the TOR pathway. *Proc Natl Acad Sci U S A*. **112**(31):9674-9 (2015).

*30.* Hall AN, Turner TN, Queitsch C. Thousands of high-quality sequencing samples fail to show meaningful correlation between 5S and 45S ribosomal DNA arrays in humans. *Sci Rep*. **11**(1):449 (2021).

*31.* 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, Abecasis GR. A global reference for human genetic variation. *Nature* **526**(7571):68-74 (2015).

*32.* Kazachenka A, Bertozzi TM, Sjoberg-Herrera MK, Walker N, Gardner J, Gunning R, Pahita E, Adams S, Adams D, Ferguson-Smith AC. Identification, Characterization, and Heritability of Murine Metastable Epialleles: Implications for Non-genetic Inheritance. *Cell*. **175**(6):1717 (2018).

*33.* Dominguez-Salas P, Moore SE, Baker MS, Bergen AW, Cox SE, Dyer RA, Fulford AJ, Guan Y, Laritsky E, Silver MJ, et al. Maternal nutrition at conception modulates DNA methylation of human metastable epialleles. *Nat Commun*. **5**:3746 (2014).

*34.* Dev VG, Tantravahi R, Miller DA, Miller OJ. Nucleolus Organizers in *Mus Musculus* Subspecies and in the Rag Mouse Cell Line. *Genetics*. **86**(2): 389-98 (1977).

*35.* Kurylo CM, Parks MM, Juette MF, Zinshteyn B, Altman RB, Thibado JK, Vincent CT, Blanchard SC. Endogenous rRNA Sequence Variation Can Regulate Stress Response Gene Expression and Phenotype. *Cell Rep.* **25**(1):236-248.e6 (2018).

*36.* Song W, Joo M, Yeom JH, Shin E, Lee M, Choi HK, Hwang J, Kim YI, Seo R, Lee JE, Moore CJ, Kim YH, Eyun SI, Hahn Y, Bae J, Lee K. Divergent rRNAs as regulators of gene expression at the ribosome level. Nat Microbiol. 2019 Mar;4(3):515-526.

*37.* Bughio F, Maggert KA. The peculiar genetics of the ribosomal DNA blurs the boundaries of transgenerational epigenetic inheritance. *Chromosome Res.* **27**(1-2):19-30 (2019).

**Acknowledgments**

We thank Profs Sue Ozanne and Oliver Rando for feedback on the work, Dr Miguel Branco for the H3K9me3 Ab, and Dr Pankaj Dubey for help in CUT&Tag-seq library preparation.

**Funding**

Biotechnology and Biological Sciences Research Council grant BB/G00711X/1 (VKR); Biotechnology and Biological Sciences Research Council grant BB/R00675X/1 (FRA, RAES, VKR); The Barts Charity Research grant MGU0390 (VKR); Rosetrees PhD studentship A1903 to SY (VKR); Biotechnology and Biological Sciences Research Council LIDO studentship BB/M009513/1 (ZA); Academy of Medical Sciences award SBF003\1026 (MLH); Edmond and Lily Safra Research Fellowship (SJM); MRC grant MR/T000783/1 (MMP) ; The Barts Charity Research grant MGU0475 (MMP); The Wellcome Trust grant 204843/Z/16/Z, (NH, ML)

**Author contributions**

Conceptualization: VKR; Methodology: AFD, RAES, HY, PPL, AH, SJM, MMP, MLH, ML, VKR, Software: FRA; Formal Analysis: FRA; Investigation: SY, AFD, RAES, HY, PPL, ZA, PD, NH, VAM, AB, MMP, MLH, ML, VKR; Resources: ML, MLH, VKR; Data Curation: FRA, SJM
Writing – original draft: FRA, VKR; Writing – review & editing: FRA, SY, AFD, RAES, HY, PPL, PD, SJM, MMP, AH, MLH, ML, VKR; Visualization: FRA, VKR; Supervision: ML, MLH, VKR; Project administration: VKR; Funding acquisition: MMP, MLH, ML, VKR

**Competing interests**

Authors declare that they have no competing interests.

**Data availability**

All data generated for this study is available from BioProject ID: PRJNA733656.

**Supplementary**

Star Methods

Figs. S1 to S17, Tables S1 to S13

**Figure 1. Long range haplotype characterization of rDNA in the C57BL/6J strain**. **(A)** Schematic representation of rDNA in the C57BL/6J mouse strain. Adapted from *Refs 2 and 34*. **(B)** Delineation of rDNA coding unit haplotypes in a C57BL/6J MEF line using ultra-long read Nanopore sequencing. The top track shows representative coverage distribution from C57BL/6J kidney short-read whole genome sequencing data, and locations of the validated SNVs used to derive haplotypes in the Nanopore data. The bottom track shows only those SNVs that distinguish amongst haplotypes. Bars with bold contour denote alleles unique to that haplotype. Bars with non-muted colours and no contour indicate positions associated with the A/C haplogroups defined by their allele at position -104. Bold text in the x-axis indicates coding subunit SNVs (18S, 5.8S, or 28S). In the three letter names given to each haplotype, the first and third characters refer to the alleles at -104 and its top-linked coding subunit SNV at 12736 (Fisher exact test, FDR-adjusted p-value < 1e-46), whereas the middle character refers to the allele at the 8063 position that maximally splits the variation of -104 "C" reads into two largely consistent haplotypes (**fig. S2**, Supplementary Materials). It should be noted that, although these three positions combined distinguish the haplotypes, each individual position is not necessarily haplotype-specific. **(C)** Co-localization analysis of rDNA haplotypes in MEF. For Nanopore reads spanning multiple rDNA units, each cell shows the average proportion of units assigned to the corresponding column haplotype in a read, given that the read includes at least one unit of the haplotype indicated in its row.

**Figure 2. The functional genomic outcomes of C57BL/6J rDNA haplotypes. (A)** Direct CpG methylation profiles of rDNA haplotypes from ultra-long read Nanopore MEF data, from -1000 bp upstream of the TSS up to the end of the 3' ETS. **(B)** Example of rDNA haplotype-specific methylation analysis on C57BL/6J kidney whole genome bisulfite sequencing (WGBS) data (left panel) showing positional information for the CpG sites associated with each haplotype in a single mouse, and aggregate sperm Reduced Representation Bisulfite Sequencing (RRBS) data for seven mice (right panel). See also **fig. S4**. **(C)** Relationship between methylation levels and rRNA-Seq expression of the ATA rDNA haplotype in C57BL/6J muscle from 22 mice with methylation obtained from Bis-PCR in the rDNA promoter (squares, $p = 7.6 \times 10^{-6}$), and kidney from 4 mice (circles) with methylation levels obtained from ATA-associated WGBS reads. **(D)** The ATA haplotype is enriched for H3K9me3 but depleted for H3K27me3 on CUT&Tag C57BL/6J kidney data (biological replicate - **fig S5**). **(E)** The ATA haplotype is depleted for UBTF relative to other haplotypes in ChIP-Seq data from wild-type C57BL/6J B cells from Diesch et al., 2019[13].

**Figure 3. Epigenetic dynamics at C57BL/6J rDNA haplotypes are life stage specific**. **(A)** Hypermethylation of the ATA haplotype (blue) is observed in different models of *in utero* stress

paradigms; the other three haplotypes (grey) are combined into a single category since they show similar effects. Methylation levels for ATA and non-ATA haplotypes for all panels are obtained on the CpG sites associated with the SNVs at 6007 and 6832. PR, LF, and HF stand for Protein-restricted, Low-fat and High-fat diet, respectively. **(B)** In Dnmt1 KO 8.5dpc embryos from Dahlet et al.[14], the ATA haplotypes lose methylation in RRBS data (left panel) and display de-repression of rRNA (right panel). Both ATA and ATG expression are obtained from ITS2 variants to avoid differential rRNA depletion effects. **(C)** In RRBS data from Petkovich et al. *(14)*, ageing-associated DNA hypermethylation is observed only at non-ATA haplotypes. The red line indicates the expected correlation coefficient for ATA haplotypes if they lost methylation at the same rate as the non-ATA haplotypes gain methylation. In this study, in the 3 month old mice the average methylation at ATA was ~75% and at non-ATA haplotypes was ~25%. **(D)** Only non-ATA haplotypes display increasingly disordered DNA methylation profiles with age (bottom track) in RRBS data from Petkovich et al. Disorder is estimated as the average of per-read Shannon entropy, exemplified here on two hypothetical mice (top track).

**Figure 4. The epiallelic nature of rDNA is a feature of other mouse inbred strains and humans**. **(A)** Identification of epivariants in kidney of 5 inbred strains besides C57BL/6J ('BL6J'): 129S1/SvImJ ('129S'), C3H/HeJ ('C3H'), C57BL/6N ('BL6N'), CAST/EiJ ('CAST') and A/J ('AJ'). For each strain, displayed positions show statistically significant allele-specific methylation differences (FDR-adjusted p-value < 0.01) for over half of the mice. **(B)** Comparison of alternative allele frequencies (AAF) at the epivariant positions between rDNA (WGS) and rRNA for one representative mouse of each strain. Taking into account variant-specific methylation levels ('Adjusted' values, circle), assuming only unmethylated units are expressed, improves correlation coefficients in 4/6 strains. **(C)** Association between total copy number from ddPCR and methylation level from WGBS of the different strains. Correlation coefficients displayed use all mice as data points (black) or a single average point per strain (grey). The analysis was repeated using the highly conserved 18S subunit only to exclude any possibility that the results are due to strain-specific differences in mapping efficiency (**fig. S13**). **(D)** Analysis of male human Mandinka samples shows allele-specific methylation levels (left panel) and increased correlation between DNA (WGBS) and rRNA allele frequencies when accounting for methylation (middle panel) at position 7980 of the KY962518.1 rDNA reference. Average methylation levels around 7980 also appear correlated with rDNA copy number estimated from WGS (right panel).

**Figure 1A**

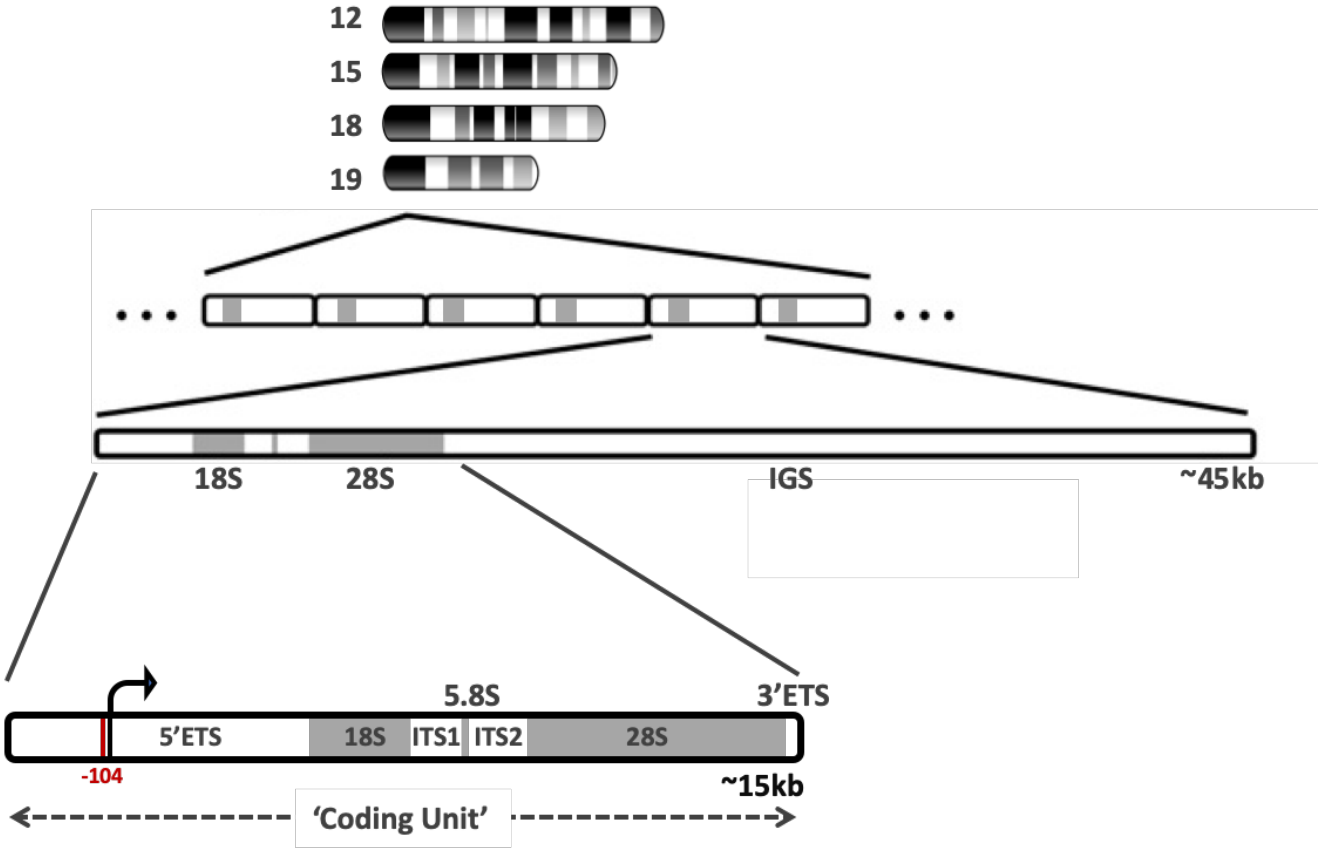

12
15
18
19
18S
28S
IGS
~45kb
5.8S
3'ETS
5'ETS
18S
ITS1
ITS2
28S
-104
~15kb
'Coding Unit'

**Figure 1B**

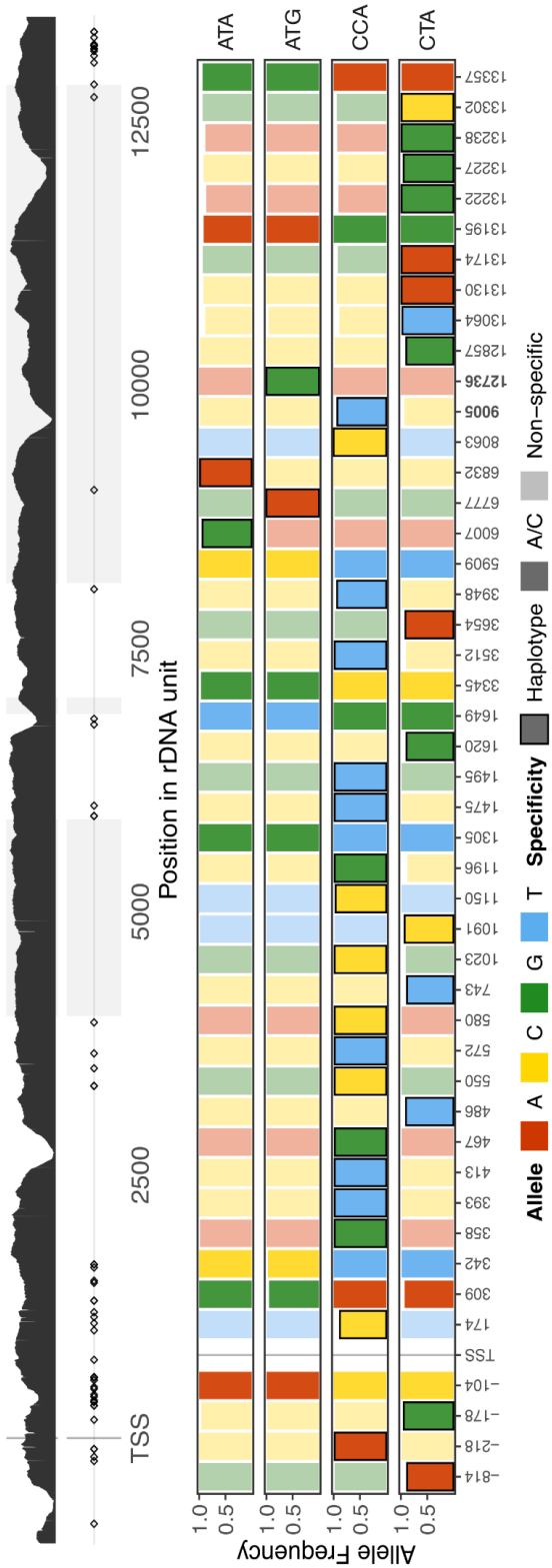

TSS
2500
5000
7500
10000
12500
Position in rDNA unit
ATA
ATG
CCA
CTA
Allele Frequency
1.0
0.5
Allele
A
C
G
T
Specificity
Haplotype
A/C
Non−specific

**Figure 1C**

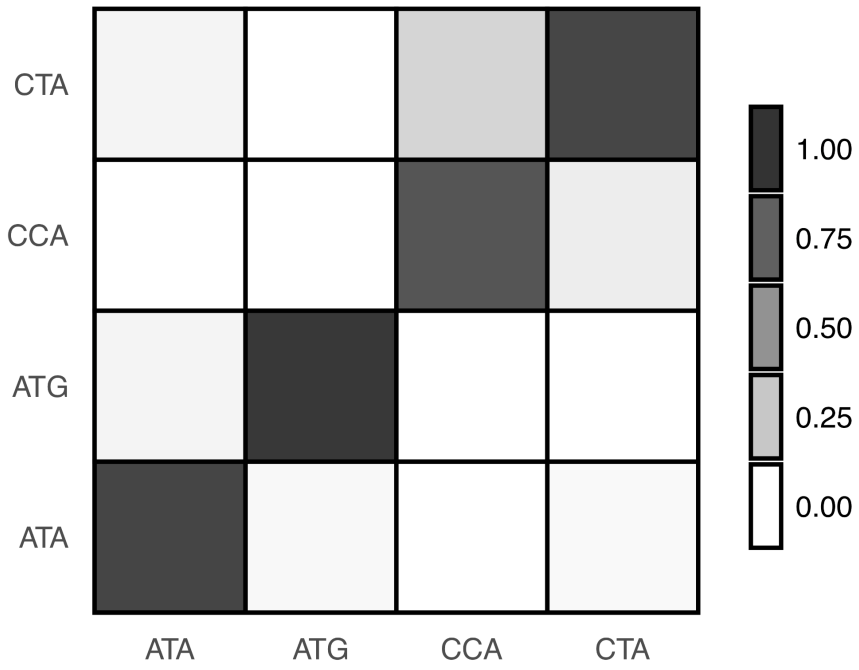


**Figure 2A**

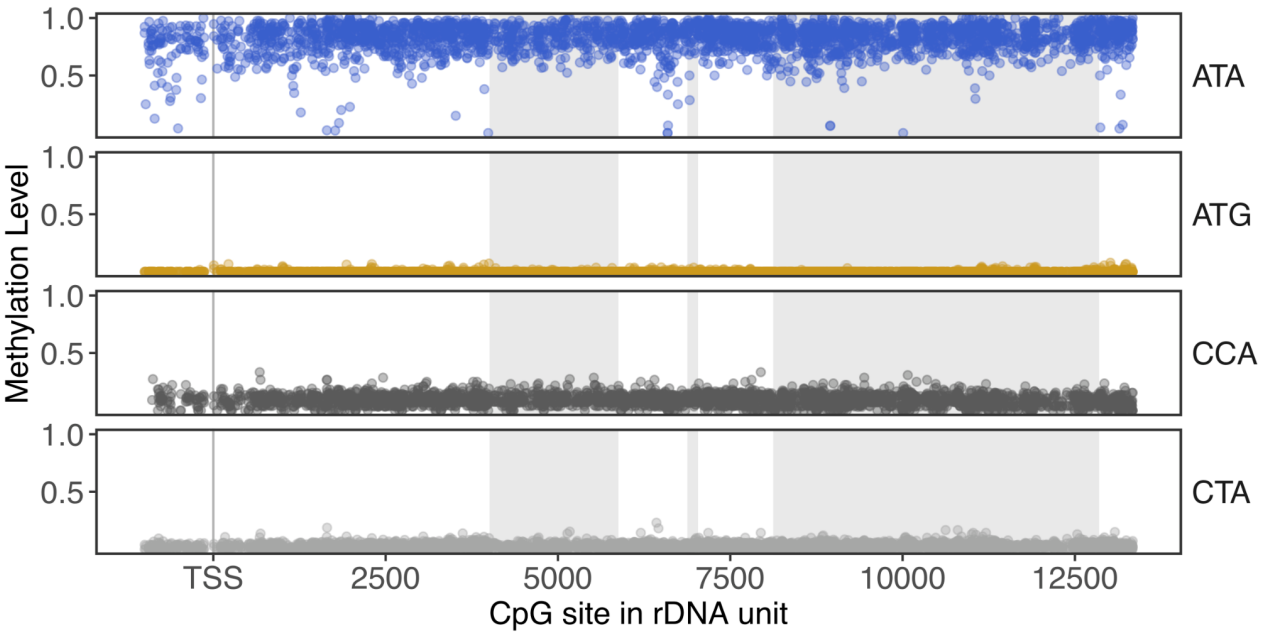

1.0
0.5
ATA
ATG
CCA
CTA
Methylation Level
TSS
2500
5000
7500
10000
12500
CpG site in rDNA unit

**Figure 2B**

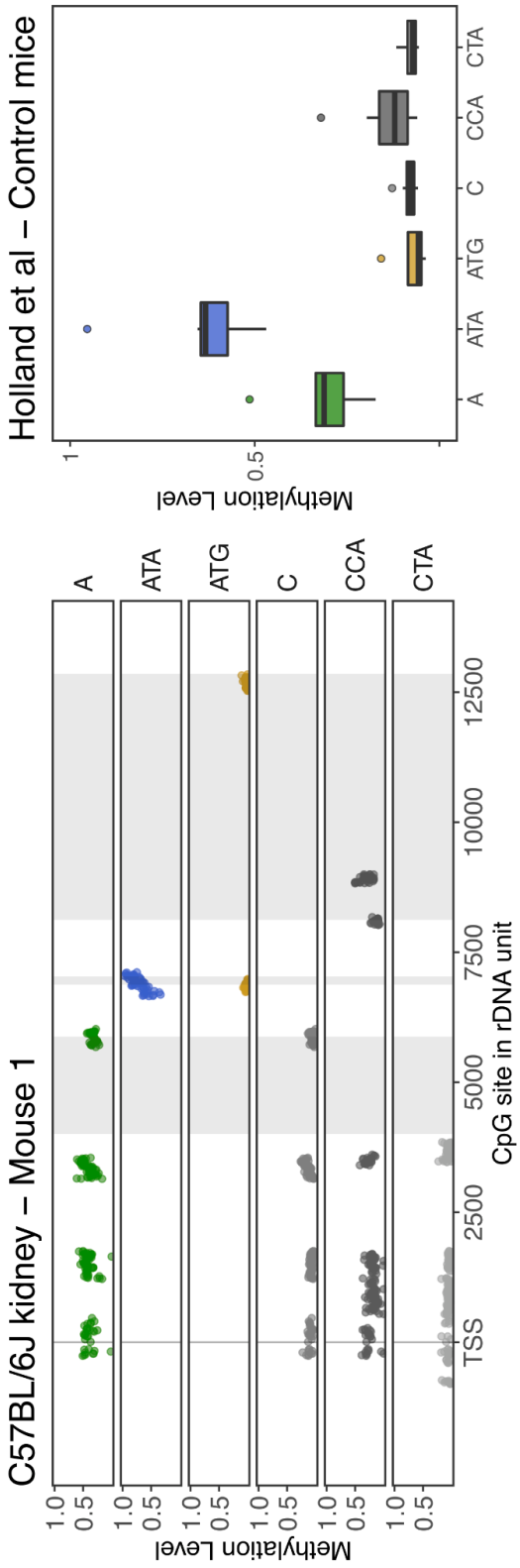

C57BL/6J kidney – Mouse 1
Methylation Level
1.0
0.5
A
ATA
ATG
C
CCA
CTA
TSS
2500
5000
7500
10000
12500
CpG site in rDNA unit
Holland et al – Control mice
Methylation Level
1
0.5
A
ATA
ATG
C
CCA
CTA

**Figure 2C**

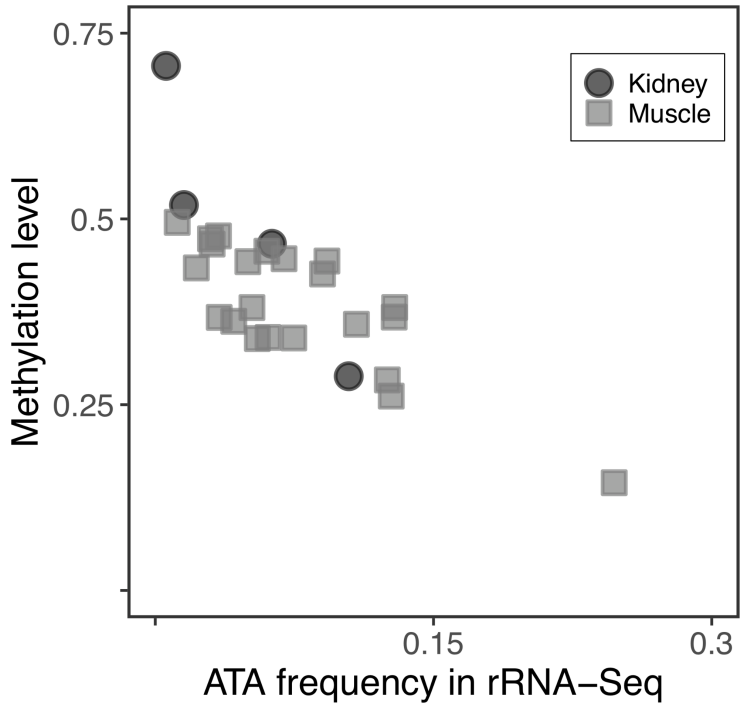

0.75
0.5
0.25
Methylation level
Kidney
Muscle
0.15
0.3
ATA frequency in rRNA−Seq

**Figure 2D**

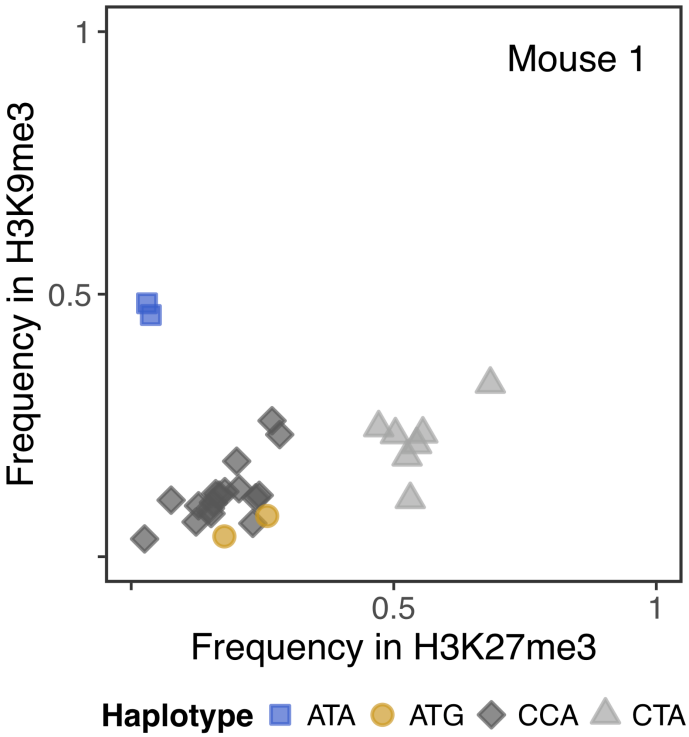

Mouse 1
Frequency in H3K9me3
1
0.5
0.5
1
Frequency in H3K27me3
Haplotype
ATA
ATG
CCA
CTA

**Figure 2E**

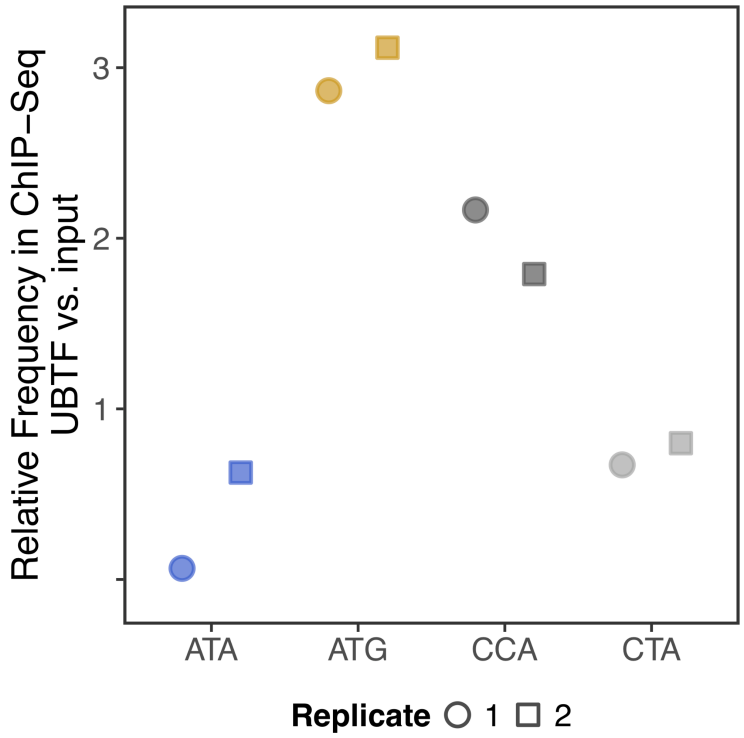

Relative Frequency in ChIP–Seq
UBTF vs. input
3
2
1
ATA
ATG
CCA
CTA
Replicate
1
2

**Figure 3A**

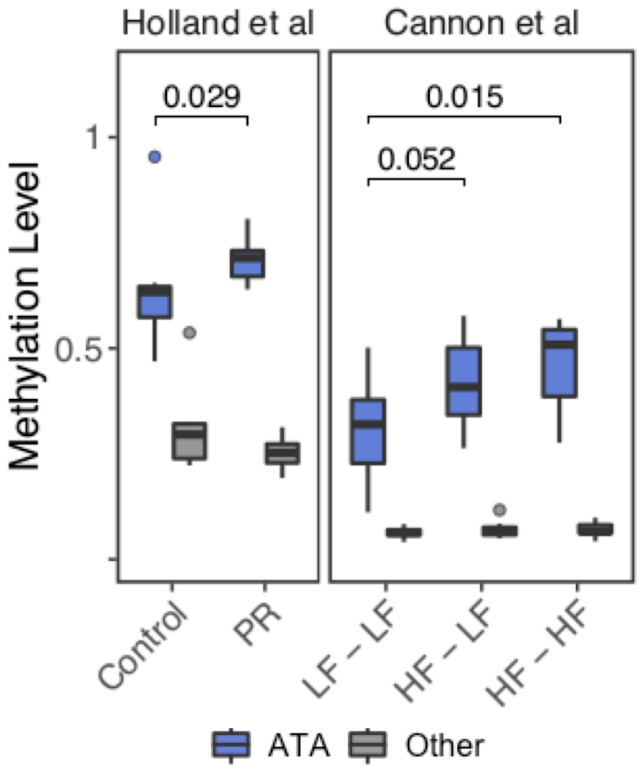

Holland et al
Cannon et al
0.029
0.015
0.052
Methylation Level
1
0.5
Control
PR
LF – LF
HF – LF
HF – HF
ATA
Other

**Figure 3B**

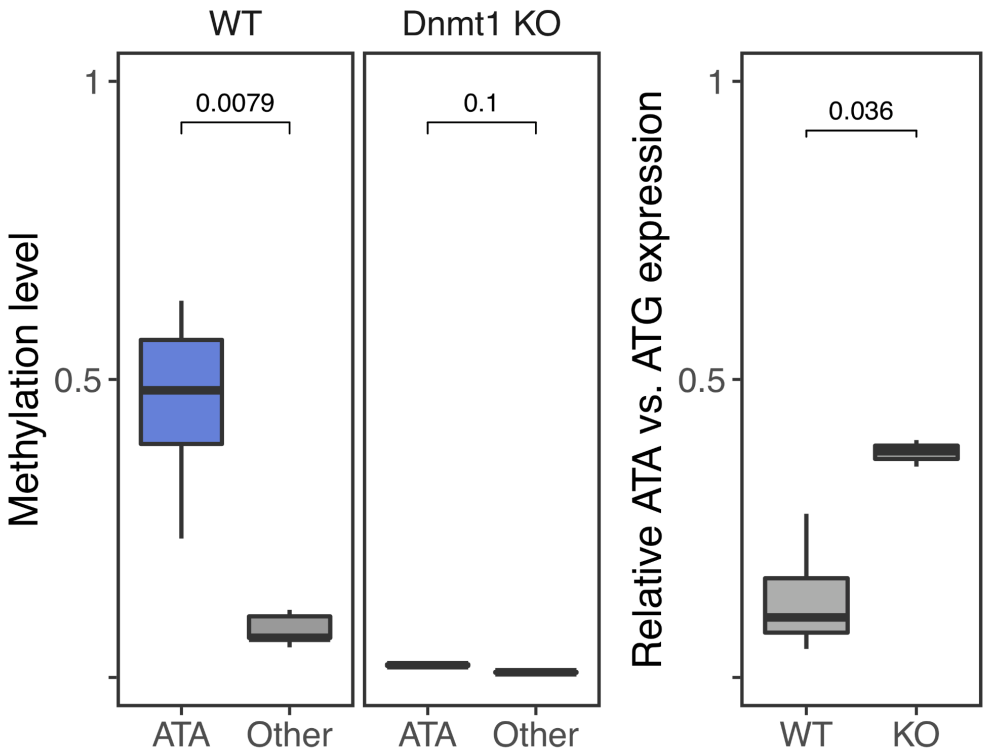

WT
Dnmt1 KO
Methylation level
1
0.5
0.0079
0.1
ATA
Other
ATA
Other
Relative ATA vs. ATG expression
1
0.5
0.036
WT
KO

**Figure 3C**

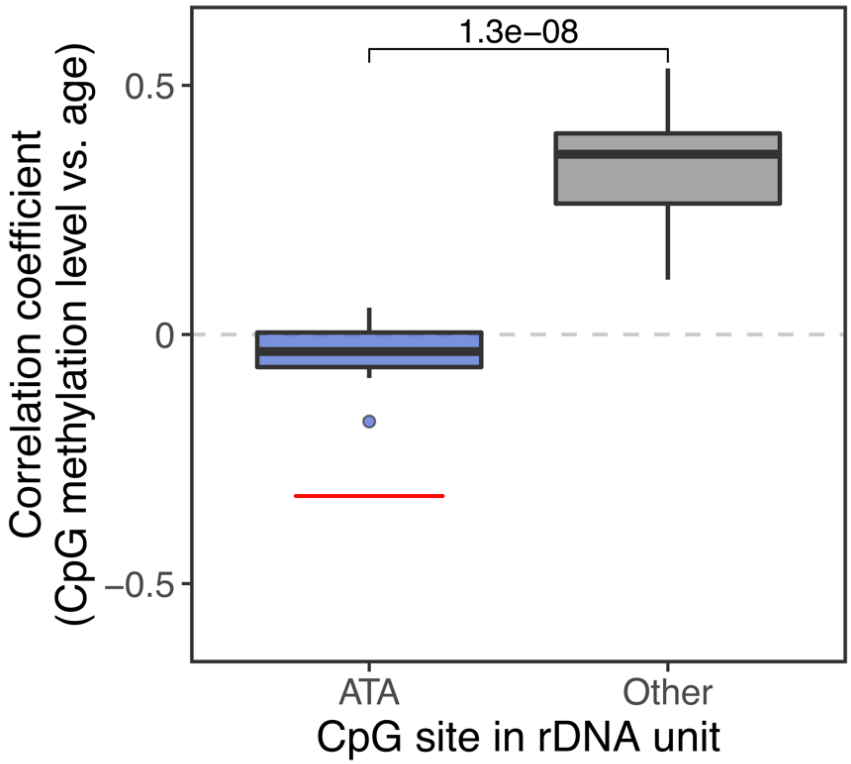

1.3e−08
0.5
0
−0.5
Correlation coefficient
(CpG methylation level vs. age)
ATA
Other
CpG site in rDNA unit

**Figure 3D**

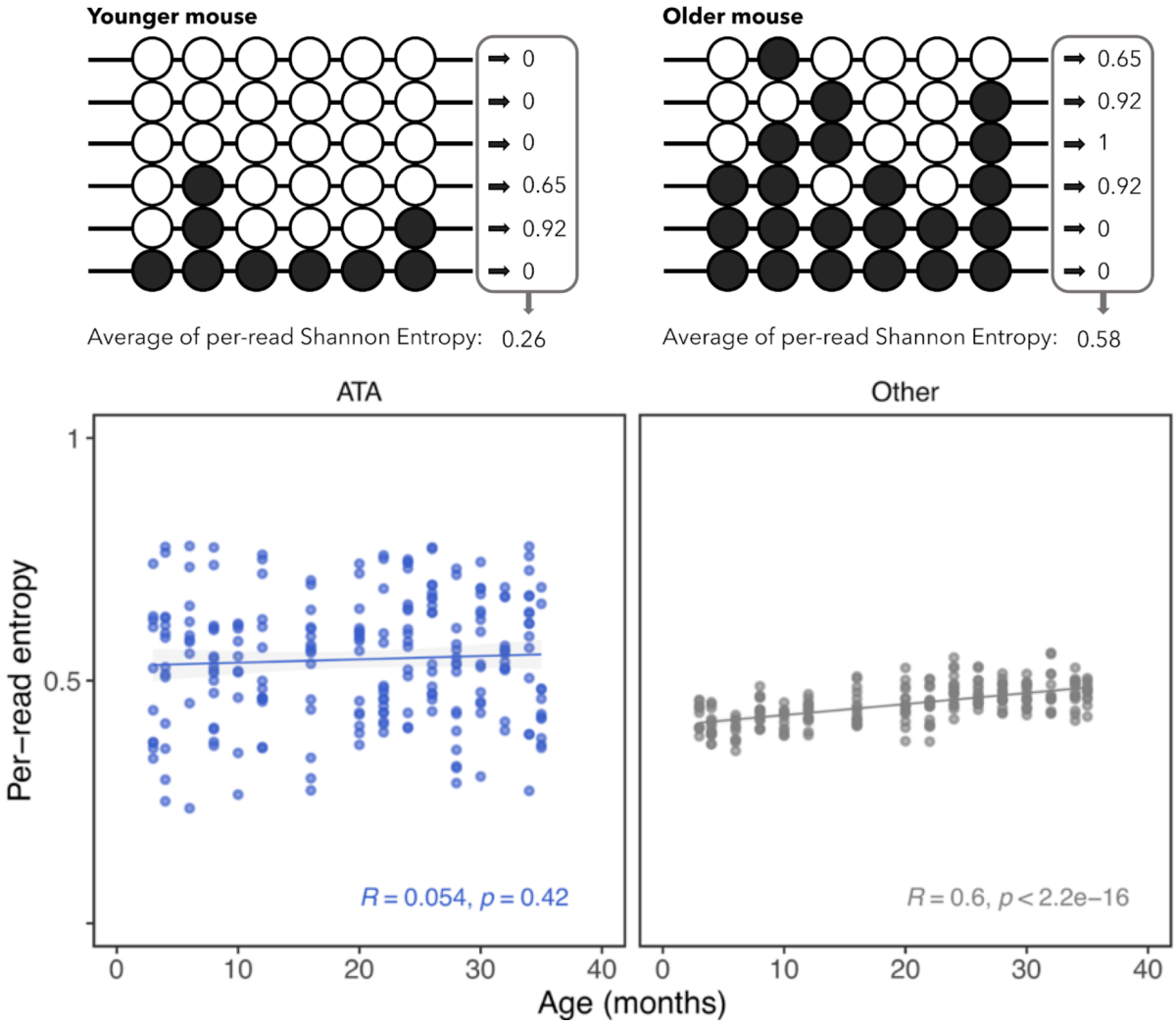

Younger mouse
0
0
0
0.65
0.92
0
Average of per-read Shannon Entropy: 0.26
Older mouse
0.65
0.92
1
0.92
0
0
Average of per-read Shannon Entropy: 0.58
ATA
Other
Per−read entropy
1
0.5
R = 0.054, p = 0.42
R = 0.6, p < 2.2e−16
0
10
20
30
40
Age (months)

**Figure 4A**

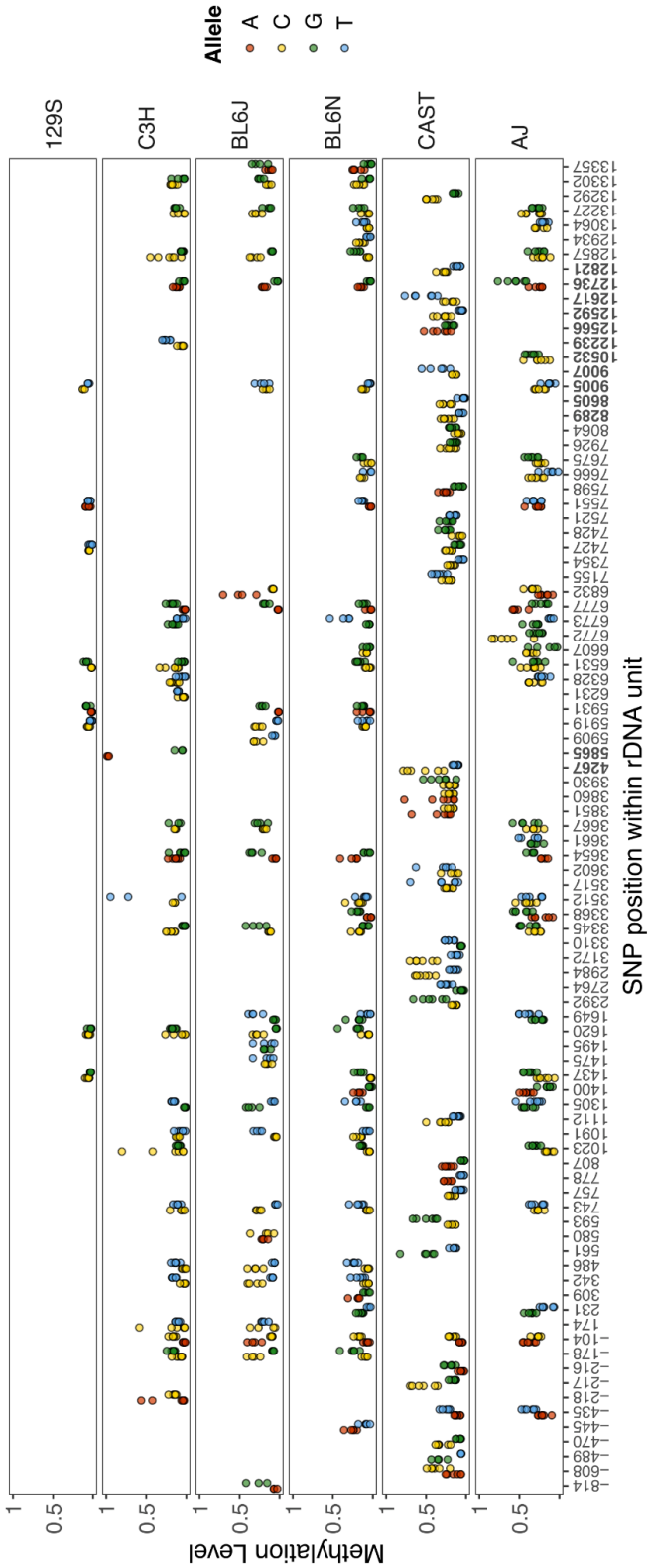



**Figure 4B**

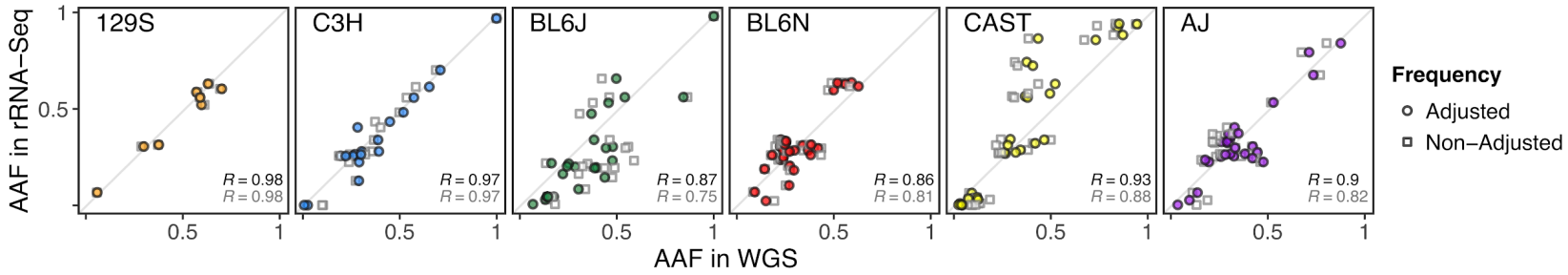

129S
C3H
BL6J
BL6N
CAST
AJ
R = 0.98
R = 0.98
R = 0.97
R = 0.97
R = 0.87
R = 0.75
R = 0.86
R = 0.81
R = 0.93
R = 0.88
R = 0.9
R = 0.82
AAF in rRNA−Seq
AAF in WGS
0.5
1
Frequency
Adjusted
Non−Adjusted

**Figure 4C**

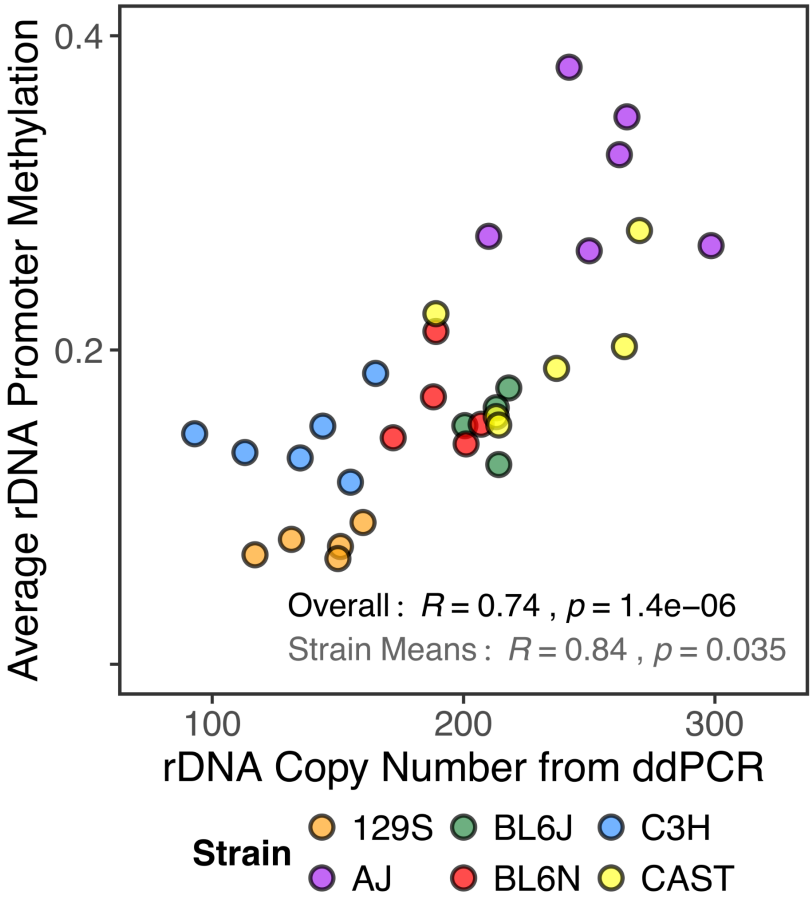
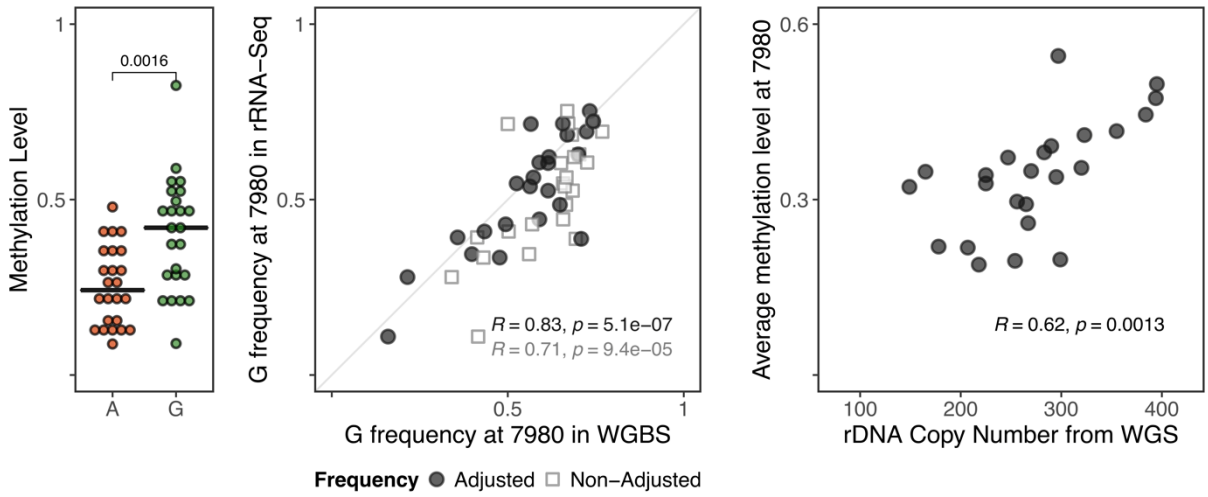

Average rDNA Promoter Methylation
0.4
0.2
Overall : R = 0.74 , p = 1.4e−06
Strain Means : R = 0.84 , p = 0.035
100
200
300
rDNA Copy Number from ddPCR
Strain
129S
BL6J
C3H
AJ
BL6N
CAST

**Figure 4D**


Methylation Level
0.0016
A
G
G frequency at 7980 in rRNA−Seq
G frequency at 7980 in WGBS
R = 0.83, p = 5.1e−07
R = 0.71, p = 9.4e−05
Frequency
Adjusted
Non−Adjusted
Average methylation level at 7980
rDNA Copy Number from WGS
R = 0.62, p = 0.0013