Profiling DNA Methylation Differences Between Inbred Mouse Strains on the Illumina Human Infinium MethylationEPIC Microarray

Short title: Cross-species utility of human DNA methylation microarray

Hemant Gujar[1], Jane W. Liang[1], Nicholas C. Wong[2], Khyobeni Mozhui[1,3*]

[1] Department of Preventive Medicine, University of Tennessee Health Science Center, Memphis, Tennessee 38163, USA

[2] Monash Bioinformatics Platform, Monash University, Clayton VIC, Australia

[3] Department of Genetics, Genomics and Informatics, University of Tennessee Health Science Centre, Memphis, Tennessee 38163, USA

*** Corresponding author**

Khyobeni Mozhui

Department of Preventive Medicine, College of Medicine

University of Tennessee Health Science Center

66 N. Pauline Street, #631

Memphis, TN 38163

Phone no. 901-448-7334

Email: kmozhui@uthsc.edu

Abstract

The Illumina Infinium MethylationEPIC provides an efficient platform for profiling DNA methylation in humans at over 850,000 CpGs. Model organisms such as mice do not currently benefit from an equivalent array. Here we used this array to measure DNA methylation in mice. We defined probes targeting conserved regions and performed a comparison between the array-based assay and affinity-based DNA sequencing of methyl-CpGs (MBD-seq). Mouse samples consisted of 11 liver DNA from two strains, C57BL/6J (B6) and DBA/2J (D2), that varied widely in age. Linear regression was applied to detect differential methylation. In total, 13,665 probes (1.6% of total probes) aligned to conserved CpGs. Beta-values (β-value) for these probes showed a distribution similar to that in humans. Overall, there was high concordance in methylation signal between the EPIC array and MBD-seq (Pearson correlation r = 0.70, p-value < 0.0001). However, the EPIC probes had higher quantitative sensitivity at CpGs that are hypo- (β-value < 0.3) or hypermethylated (β-value > 0.7). In terms of differential methylation, no EPIC probe detected significant difference between age groups at a Benjamini-Hochberg threshold of 10%, and the MBD-seq performed better at detecting age-dependent change in methylation. However, the top most significant probe for age (cg13269407; uncorrected p-value = 1.8×10^{-5}) is part of the clock CpGs used to estimate the human epigenetic age. For strain, 219 Infinium probes detected significant differential methylation (FDR cutoff 10%) with ~80% CpGs associated with higher methylation in D2. This higher methylation profile in D2 compared to B6 was also replicated by the MBD-seq data. To summarize, we found only a small subset of EPIC probes that target conserved sites. However, for this small subset the array provides a

42 reliable assay of DNA methylation and can be effectively used to measure differential

43 methylation in mice.

44 **Keywords:** DNA methylation, epigenetics, microarray, cross-species comparison

45

Introduction

There has been a surge in large-scale epigenetic studies in recent years. In particular, epigenome-wide association studies (EWAS) of DNA methylation have shown association with physiological traits [1,2], diseases [3-5], environmental exposures [6,7], aging [8], and even socioeconomic [9] and emotional experiences [10]. The development of robust and reliable methylation microarrays has been an important driving force. In particular, the Illumina Human Methylation BeadChips have made it both convenient and cost-effective to incorporate an epigenetic arm to large epidemiological studies [11,12]. The latest version, the Illumina Infinium MethylationEPIC BeadChip (EPIC), provides an efficient high throughput platform to quantify methylation at 866,836 CpG sites on the human genome [13,14]. A remarkable biological insight that has emerged from these array-based studies is the definition of the methylation-based “epigenetic clock,” a biomarker of human age and aging (i.e., the epigenetic clock) that is defined using specific probes represented on these arrays [8].

Currently there is no equivalent microarray platform for model organisms and work in experimental species have largely relied on high-throughput sequencing. For instance, while the human DNA methylation age can be calculated from a few hundred probes on the Illumina BeadChips, a similar effort in mice required a more extensive sequencing of the mouse methylome [15]. However, CpG islands (CGIs) are largely conserved between mice and humans and the two species share similar numbers of CGIs and similar proportions of CGIs in promoter regions of genes [16]. Considering that these CpGs and CGIs are highly conserved in gene regulatory regions, it is feasible that

probes on the human microarrays that target these sites may have some application in research using rodent models. This was previously evaluated for the two older versions of the Illumina HumanMethylation BeadChips [17]. The work by Wong et al. demonstrated that a subset of the probes targeting highly conserved sites provide reliable measures of DNA methylation in mice, and could be feasibly used to evaluate tissue specific methylation and in cancer related studies using the mouse as a model system.

In the present work, we extend the conservation analysis to the EPIC platform, and evaluate the capacity of these probes to detect differential methylation. We begin by defining the conserved probes and the key features of the corresponding CpG sites in the context of the larger mouse and human genome. We also compare the methylation signal detected by the conserved probes with affinity-based methyl-CpG enriched DNA sequence (MBD-seq) data from the same samples and evaluate if the conserved probes are informative of age and strain differences in mice.

Materials and Methods

Defining Conserved EPIC probes

Sequences for the 866,836 CpG probes were obtained from Illumina (http://www.illumina.com/). The probe sequences were aligned to the mouse genome (mm10) using bowtie2 (version 2.2.6) with standard default parameters. A total of 34,981 probes aligned to the mouse genome of varying alignment quality. Conserved probes were then defined based on quality of alignment. For this, we filtered out all

sequences with a low mapping quality (MAPQ) of less than 60 (15,717 excluded) and those that contain more than two non-matching base pairs (1,092). To retain only the high quality probes, we further filtered probes based on confidence in DNA methylation signal and based on this, 4,507 probes with detection p-values > 0.0001 were removed. This generated a list of 13,665 high quality probes that are conserved sequences and provide reliable methylation assays in mice (these are listed in **Supplementary Data S1**). CpG island annotations [18] for the respective genome were downloaded from UCSC Genome Browser (http://genome.ucsc.edu) and distribution of conserved probes and positions of CGIs were plotted to the human (GRCh37) and mouse (mm10) genomes using CIRCOS [19].

For conserved sequences, there is high correspondence in functional and genomic features between mouse and human genomes and we referred to the human probe annotations provided by Illumina to define the location of conserved probes with respect to gene features and CpG context (i.e., islands, shores, shelves) (**Supplementary Data S1**). To evaluate if the conserved set is enriched in specific features relative to the full background set, we performed a hypergeometric test using the phyper function in R.

Animals and sample preparation

Tissues samples were derived from mice that were part of an aging cohort maintained at the University of Tennessee Health Science Center (PI: Robert W. Williams). Details on animal rearing and sample collection are described in Mozhui and Pandey 2017 [20]. All animal procedures were approved by the UTHSC Animal Care and Use Committee.

Liver tissues were collected from mice aged at ~4 months (mos; young), ~12 mos (mid), and ~24 mos (old). The mice were of two different strains—C57BL/6J (B6) and DBA/2J (D2)—and as the colony was set up to study aging in females, the majority of the mice in this study are females (**Table 1**). Mice were euthanized by intraperitoneal injection of Avertin (250 to 500 mg/kg of a 20 mg/ml solution), followed by cardiac puncture and exsanguination. All sample collection procedures were done on the same day within a 3-hour timeframe. Liver samples were snap-frozen and stored at -80°C until use.

DNA was purified from the liver tissue using the Qiagen AllPrep kit (http://www.qiagen.com) on the QIAcube system. Nucleic acid quality was checked using a NanoDrop spectrophotometer (http://www.nanodrop.com). As reference, we also included two human samples. These are DNA derived from the buffy coats from two individuals.

Table 1: Sample details and average methylation signal intensity

Sample	Age	Age (months)	Strain [1]	Sex	Full set (850K)		Conserved set (13665)	
					Mean	Median	Mean	Median
Mouse1	young	4	D2	F	505	394	3206	1898
Mouse2	young	4	D2	F	926	524	10989	10278
Mouse7	young	4	B6	F	877	538	9866	8702
Mouse8	young	4	B6	F	766	397	10386	9975
Mouse3	mid	12	D2	F	852	483	10615	9880
Mouse4	mid	12	D2	F	818	430	10866	10542
Mouse5	mid	12	D2	M	845	456	11545	10982
Mouse9	mid	12	B6	M	852	444	11433	11187
Mouse6	old	24	D2	F	737	379	10206	9611
Mouse10	old	24	B6	F	845	448	10767	10436
Mouse11	old	24	B6	F	886	490	11302	10741
Human1					7568	7218	8710	8616
Human2					10668	10288	11761	11599

[1] D2: DBA/2J; B6: C57BL/6J

DNA methylation microarray and data processing

DNA methylation assays were performed as per the standard manufacturer's protocol (http://www.illumina.com/). In brief, 500 ng of DNA extracted from the mouse liver was treated with sodium bisulfite to convert cytosine to uracil. The 5-methyl cytosine remains unreactive to sodium bisulfite. The DNA is then hybridized to the EPIC BeadChip. After washing off unhybridized DNA, a single base extension was recorded to calculate the methylation level at the CpG probe site. DNA methylation assays were performed at the Genomic Services Lab at the HudsonAlpha Institute for Biotechnology (http://hudsonalpha.org). Raw intensity data files (idat files) for both mouse and human samples were processed using the R package, Minfi [21].

The intensity and β-values were used to evaluate the performance of the EPIC probes in mice and humans. Comparisons were based on the full set of 850K probes and the conserved set of 13,665 probes. We also used the β-values and signal intensity scores for the 13,665 probes to perform hierarchical clustering and principal component analysis for the mouse samples. From initial quality checks, we identified one outlier mouse sample (**Supplementary Fig. S1**) that had lower intensity and higher detection p-value compared to the other mouse samples. This sample was excluded from the statistical tests.

MBD-seq comparison

The mouse samples we report here were previously assayed for DNA methylation using MBD-seq [20]. This is an affinity-based enrichment of methylated CpGs using the methyl binding domain (MBD) of methyl-CpG-binding protein 2, followed by high throughput sequencing (MBD-seq) [22-24]. Sequencing was performed on Life Technologies' Ion Proton platform. Data have been deposited to the NCBI's Gene Expression Omnibus (https://www.ncbi.nlm.nih.gov/geo/; GEO accession ID GSE95361) and Sequence Repository Archive (https://www.ncbi.nlm.nih.gov/sra/; SRA accession ID SRP100703). To compare methylation signal detected by the conserved EPIC arrays, we extracted MBD-seq reads at the corresponding sites. MBD-seq does not provide single-base resolution as the resolution is limited to the fragment size, in this case ~300 bp. However, since methylation levels at neighboring CpGs are largely correlated [25], we derived quantitative data from the number of read fragments that map to a CpG region. For the sites in the mouse genome targeted by the conserved EPIC probes, we expanded the window to 300 bp bins, and extracted the MBD-seq fragment counts. The CpG density-normalized methylation level was then quantified using the MEDIPS R package [26]. We then used Pearson's correlation to compare the EPIC β-values and the relative methylation score (rms or the CpG density normalized methylation) detected by MBD-seq [27].

Analysis of differential methylation

Statistical analyses were done in R (https://www.r-project.org/) and JMP Statistics (JMP Pro 12). Mice were grouped into three age categories (young, mid, and old; additional sample details are in **Table 1**). To evaluate differential methylation detected by the

13,655 conserved probes, we applied a regression model with age, strain and sex as predictors (~ageGroups + strain + sex) for each probe using the R glm function and type III anova to calculate test statistics (equations are provided in **Supplementary Data S1**). For the MBD-seq reads, we performed differential methylation analysis of the read counts using the edgeR R package [28]. The same linear regression model was applied (~ageGroups + strain + sex) and equations are provided in **Supplementary Data S1.** We then cross-compared differential methylation detected by the two methods. Treating the EPIC data as a discovery set, we applied the Benjamini-Hochberg (BH) procedure to control the false discovery rate (FDR) [29,30]. We then defined differentially methylated CpGs (DMCpGs) and evaluated the corresponding region in the MBD-seq data to test replication at a lenient uncorrected p-value threshold of 0.05. Likewise, in the reverse comparison, we applied an FDR threshold to identify differentially methylated regions (DMRs) in the MBD-seq data, and tested replication of the corresponding CpG at an uncorrected p-value threshold of 0.05.

Results

Conserved Infinium MethylationEPIC probes

The human EPIC array contains 866,836 50-mer probes. Out of these, we defined a total of 13,665 probes that align to conserved sites in the mouse genome and provide high quality methylation signal (details on mapping quality scores and methylation signal confidence are provided in **Supplementary data S1)**. In the full set of EPIC probes, 71% are located within annotated gene features or within 200–1,500 bp upstream of transcription start sites (TSS). Compared to this background set, a higher percent of the

conserved probes (88%; 11,972 probes) target such functionally annotated regions. Probes that target CpGs located in exons, 5' UTR, and within 200 bp upstream of TSS (TSS200) are highly overrepresented among the conserved set (**Table 2**). This is expected, since sequences in these functional regions are conserved across species. The upstream regulatory regions and the first exon harbor a large percent of CGIs, and compared to the background set, there is close to a 2.5-fold higher enrichment in CGIs among the conserved probes (**Table 2**). In contrast, there is no enrichment in probes that target CpGs that are between 200–1,500 bp upstream of TSS (TSS1500), gene body (mostly intronic), 3' UTRs, and non-genic regions. Locations of the conserved probes and CGI densities in the human and mouse genomes are shown in **Fig. 1.**

Table 2: Genomic features of CpGs and enrichment in conserved sites

	Full set (850K)		Conserved set (13665)		
Feature	**Counts**	**Percent Total**	**Counts**	**Percent Total**	**Enrichment p[3]**
Gene features[1]					
TSS1500	107193	12	1195	9	ns
TSS200	65152	8	1940	14	<1.0E-15
5'UTR	73070	8	1269	9	1.8E-04
1stExon	26433	3	2028	15	<1.0E-15
Exon	5680	1	282	2	<1.0E-15
3'UTR	21594	2	340	2	ns
Body	318165	37	4918	36	ns
Non-Genic	249549	29	1693	12	ns
CpG islands and flanking regions[2]					
Islands	161598	19	6270	46	<1.0E-15
Shores	154735	18	2267	17	ns
Shelves	61811	7	664	5	ns
Open Sea	488692	56	4464	33	ns

[1] CpG position relative to gene features based on annotations from Illumina (UCSC_RefGene_Group).
TSS1500 and TSS200 are CpGs at 0–200 or 200–1500 upstream of are transcription start sites; Non-genic are CpG with no annotated gene features.

[2] Shores = 0–2 kb from islands; shelves = 2–4 kb from islands
[3] Enrichment of gene features and CpG regions in the conserved set compared to the full set based on hypergeometric test

Fig. 1. Location of conserved Illumina HumanMethylationEPIC probes and CpG densities in the human and mouse genomes.

The outer circle displays the chromosomes and circular karyotype of the human and mouse genomes. CpG island (CGI) density is shown in the second circle. The innermost circle displays the positions of CpGs targeted by the 13,665 conserved probes.

Comparison of probe performance in mouse and human samples

We used data generated from two human samples as reference. Using the full set of 850K probes, the mouse samples showed low overall signal intensity (**Fig. 2A**). The mean signal intensity for the two human samples was 9,118 ± 2,192 (**Table 1**). For the mouse samples, the mean signal intensity was 810 ± 114 (**Table 1**). The β-value distribution also showed poor performance for mice with a peak β-value at 0.4 that indicates failure for probes. The methylation β-values in human samples showed the expected bimodal distribution that characterizes the Illumina methylation arrays (**Fig. 2B**) [13,14].

The EPIC BeadChip clearly performed poorly in mice when we considered the full set of probes. However, when we considered only the 13,665 conserved probes, the methylation signal became comparable between the mouse and human samples. Total

mean signal intensity for the mouse samples ranged from 9,866 to 11,545 (Mouse1, which failed the initial QC, has very low signal intensity compared to the other mouse samples; this was excluded from differential methylation analysis) (**Table 1**). Mean signal intensity for the two human samples were 8,711 and 11,761 (**Table 1**). The bimodal β distribution was also observed for this set of conserved probes in mouse samples (**Fig. 2C, 2D**).

Fig. 2. Distribution of signal intensities and methylation β-values in mice and humans.

For the full set of 866,836 probes on the Illumina Infinium MethylationEPIC, the mouse samples have **(A)** low signal intensity compared to the two human samples, and **(B)** the β-values have a unimodal distribution that peaks at ~0.4. The two human samples have the expected bimodal distribution for β-values. For the conserved set of 13665 probes, both the **(C)** signal intensity, and **(D)** the β-value distribution in the mouse samples are comparable to the two human reference samples. The signal intensity for mouse1 is relatively low for the conserved set of probes and this sample plots as an outlier in the principal component analysis.

Comparison with MBD-seq

To determine if we could find concordant methylation signal, we compared the microarray β-values with the CpG density-normalized rms derived from MBD-seq data (average β-values and rms are provided in **Supplementary data S1**). Overall, there

was concordance between the two technologies, and the β-values and rms were significantly correlated (Pearson's correlation of 0.70, $p < 0.0001$). We grouped the EPIC probes into three categories based on β-values—hypomethylated for $\beta < 0.3$, hemimethylated for $0.3 \leq \beta \leq 0.7$, and hypermethylated for $\beta > 0.7$—and examined correlations with the rms within each category. Given the high representation of islands and CpGs in 5' regions of genes, which generally remain hypomethylated [16,31], the majority of the conserved probes fell into the hypomethylated category (**Table 3**). For the hypomethylated probes, 82% of the corresponding CpG regions also had rms < 0.3 (**Table 3**) and there was modest correlation between the rms and β-values (Pearson's $r = 0.18$; $p = 0.0001$; **Fig. 3A**). For many of the CpGs regions that correspond to the hypomethylated probes, the rms were close to 0 and appeared unmethylated in the MBD-seq data. For hemimethylated probes, 58% of the corresponding regions had $0.3 \leq$ rms ≤ 0.7 and 31% had rms < 0.3. This subset showed the highest correlation between the β-values and rms ($r = 0.46$; $p < 0.0001$; **Fig. 3B**). For hypermethylated probes, only 40% of corresponding regions were associated with rms > 0.7, and 54% had $0.3 \leq$ rms $\leq .7$. This subset showed lower correlation between the β-values and rms ($r = 0.04$, $p = 0.0392$; **Fig. 3C**). The corresponding CpG regions for this hypermethylated set tended to have rms close to 0.75. This clustered rms distribution for CpG regions at the lower and upper levels of methylation indicate that the MBD-seq has lower quantitative sensitivity at these regions.

Overall, the strong concordance with the MBD-seq data shows that the conserved EPIC probes provide a reliable quantification of methylation in mice. However, for CpGs that

are hypomethylated or hypermethylated, the EPIC technology may have an advantage and provide higher quantitative sensitivity compared to the MBD-seq.

Table 3. Counts of Illumina HumanMethylationEPIC probes by β-value and concordance with MBD-seq at corresponding CpG regions

CpG Category[1]	Probes counts[1]	Counts of CpG regions by rms value[2]		
		rms < 0.3	0.3 ≤ rms ≤ 0.7	rms > 0.7
Hypo β < 0.3	7548	6198	1000	350
Hemi 0.3 ≤ β ≤ 0.7	3159	973	1827	359
Hyper β > 0.7	2956	171	1599	1186

[1]Conserved probes on the HumanMethylationEPIC arrays were grouped by β-value. These are counts in each category.
[2]CpG For each category of probes, the corresponding CpG regions were counted and grouped by CpG density normalized relative methylation score (rms) to determine concordance between the array and MBD-seq

Fig. 3. Correlation between MethylationEPIC and MBD-seq data

The 13,665 conserved MethylationEPIC probes were classified into three categories based on average β-values: hypo for β < 0.3, hemi for 0.3 ≤ β ≤ 0.7, and hyper for β > 0.7. For each probe, the 300 bp window around the corresponding CpG was determined and the CpG density-normalized relative methylation score (rms) was estimated for that region from MBD-seq data. A comparison between the β-values and rms showed **(A)** modest correlation for the hypomethylated CpGs (Pearson r = 0.18; p = 0.0001), **(B)** strong correlation for hemimethylated CpGs (r = 0.46; p < 0.0001), and **(C)** low correlation for hypermethylated CpG (r = 0.04, p = 0.0392). For CpGs with low β-values, the corresponding regions showed rms that cluster close to 0, and for CpG with high β-values, the corresponding rms tended to cluster close to 0.75.

Differential Methylation Analysis

We applied linear regression to examine differential methylation by age group and strain, and cross-referenced the DM-CpGs detected by the EPIC array with DMRs detected by MBD-seq. For the effect of age, no conserved EPIC probe passed a 10% FDR threshold (full results and p-values are provided in **Supplementary data S1**). However, we note that the probe that detected the most significant effect of age, cg13269407, is among the 353 CpGs that are used to estimate the human epigenetic age [8]. This CpG is hemimethylated (average β-value of 0.55) and associated with a ~2.4-fold decline in methylation between young and old age (uncorrected p-value = 1.8 x 10^{-5}). In the MBD-seq, the corresponding region is classified as hypomethylated with rms = 0 for most of the samples and no reliable statistics could be carried out for this region due to small number of mapped reads. We then performed a reverse comparison to identify age-dependent DMRs (age-DMRs) in the MBD-seq data and evaluated replication by the EPIC probes. At the same FDR threshold of 10%, the MBD-seq detected seven age-DMRs. These strong age-DMRs have rms between 0.3 and 0.7 and are associated with an increase in methylation with age. Most occur in CGIs that have been reported previously [20]. Out of these seven age-DMRs, six corresponding EPIC probes replicated the age-dependent increase in methylation at a nominal p-value cutoff of 0.05 (**Table 4**).

313 **Table 4. Age-dependent differentially methylated CpGs/regions detected by**

314 **conserved probes and by MBD-seq**

				EPIC[1]		MBD-seq[1]	
ProbeID	**Gene[2]**	**Region[2]**	**Position (mm10)[3]**	**Coef.**	**Age (P)**	**logFC**	**Age (P)**
cg08949408	*C1QL3*	Body; Island	chr2:13.01	0.32	0.001	3.3	1.3E-10
cg10444382	*RFX4*	Body; Island	chr10:84.76	0.24	9.4E-04	2.9	2.5E-08
cg22384902	*LRRC4; SND1*	TSS1500; island	chr6:28.83	0.22	0.009	2.0	2.0E-06
cg06945399	*LRRC4; SND1*	TSS200; Island	chr6:28.83	0.18	0.057	1.5	2.2E-05
cg23398076	*MEIS1*	Body; Shelf	chr11:19.02	0.13	0.007	1.5	2.4E-05
cg05393688	*TSC22D1*	Body; Shore	chr14:76.51	0.17	0.005	1.5	2.8E-05
cg20563498	*USP35*	Body; Shelf	chr7:97.32	-0.02	0.27	1.1	3.2E-05

315 [1]These are age-dependent differentially methylated CpG regions discovered in the MBD-seq at an FDR of
316 10%; replicated for the corresponding CpG in the EPIC microarray at an uncorrected p-value cutoff of
317 0.05. Coef. is the linear regression coefficient (i.e., change in methylation β-value from young to old).
318 LogFC is $\log_2$ fold change in methylation from young to old.
319 [2]CpG location in relation to gene features and CpG region based in probe annotations for the human
320 methylation microarray; gene feature annotations are the same for the corresponding regions in the
321 mouse genome.
322 [3]Chromosome and Megabase coordinate based on mm10 mouse reference genome

323

324 For strain effect, 219 conserved EPIC probes detected significant difference in
325 methylation between B6 and D2 at an FDR threshold of 10% (strain-DMCpGs). Close to
326 80% of these CpGs (175 out of 219) are associated with higher methylation in D2
327 relative to B6. In the MBD-seq data, only 29 of the 219 corresponding regions replicated
328 strain effect at an uncorrected p-value cutoff 0.05 (**Table 5**). Of these, 9 were
329 associated with higher methylation in B6, and 20 were associated with higher
330 methylation in D2. In the reverse comparison, we identified only 37 strain-dependent
331 DMRs (strain-DMRs) at an FDR cutoff of 10%. Consistent with the EPIC data, the
332 majority of these regions (21 of the 37) showed higher methylation in D2 relative to B6.

Of these, 16 strain differences were replicated at the corresponding CpG in the EPIC data (6 with higher methylation in B6 and 10 with higher methylation in D2) (**Table 5**).

Table 5. Strain-dependent differentially methylated CpGs/regions detected by both EPIC probes and by MBD-seq

				EPIC[1]		MBD-seq[1]	
ProbeID	**Gene[3]**	**Region[4]**	**Position (mm10)[3]**	**Coef.[2]**	**Strain (P)**	**logFC[2]**	**Strain (P)**
Differentially methylated CpGs detected by EPIC probe at FDR 10%; replicated by MBD-seq							
cg21064315	*SZT2*	3'UTR; Shore	chr4:118.36	-0.82	5.5E-09	-2.0	1.7E-04
cg14945867	*CNIH*	1stExon; Island	chr14:46.79	0.27	1.3E-06	6.0	1.2E-09
cg04546815	*KANK4*	Body	chr4:98.78	0.34	1.6E-06	1.7	4.4E-04
cg10277781	*CNIH*	1stExon; Island	chr14:46.79	0.35	1.9E-06	6.0	1.2E-09
cg00049718	*CSDE1*	5'UTR	chr3:103.02	0.40	2.6E-06	6.8	9.8E-15
cg07211292	*C20orf160*	3'UTR; Island	chr2:153.08	-0.46	5.0E-06	-1.5	1.4E-05
cg24255125	*GRIK4*	Body; Island	chr9:42.52	-0.36	7.8E-06	-3.0	5.9E-09
cg03517030	*MTCH2*	1stExon; Island	chr2:90.85	0.35	1.6E-05	6.7	2.3E-14
cg05781968	*WNT5A*	Body; Island	chr14:28.51	0.31	4.4E-05	2.3	1.0E-05
cg04154281	*UBTF*	Body; Shore	chr11:102.31	0.17	6.5E-05	0.7	0.03
cg06861375	*ZNF697*	Body; Island	chr3:98.43	0.36	6.7E-05	4.5	2.8E-04
cg24959134	-	-	chr10:92.44	-0.33	9.4E-05	-2.4	0.01
cg06552810	-	-	chr2:106.19	0.26	1.1E-04	2.9	0.002
cg01663821		Shore	chr3:98.94	0.19	1.3E-04	0.9	0.02
cg00597112	-	-	chr11:109.01	0.21	1.4E-04	0.5	0.002
cg26857408	*UBTF*	Body; Shore	chr11:102.31	0.24	2.1E-04	0.7	0.03
cg15172734	*SLMAP*	5'UTR; Shore	chr14:26.53	-0.11	3.4E-04	-2.5	0.01
cg09990537	*WNT5A*	5'UTR; Shore	chr14:28.51	0.17	3.4E-04	1.0	0.004
cg12849734	-	-	chr2:157.71	0.14	4.4E-04	0.9	0.01
cg21746387	*NDUFA4L2*	TSS1500; Shore	chr10:127.51	-0.17	5.5E-04	-3.4	0.001
cg11382417	-	-	chr2:96.32	-0.21	6.0E-04	-4.7	1.3E-07
cg02865068	-	Shore	chr2:105.66	0.11	9.7E-04	2.9	0.04
cg14275842	*CHRNE*	Body; Island	chr11:70.62	0.18	0.001	1.0	0.005
cg02159996	*GABRR1*	5'UTR	chr4:33.13	0.13	0.001	1.2	2.5E-04
cg00920372	-	-	chr19:45.33	-0.08	0.001	-1.5	8.8E-04

cg03422015	*ERC1*	Body	chr6:119.69	0.04	0.001	1.1	0.02
cg04340318	-	-	chr4:86.04	0.16	0.001	2.3	0.001
cg14465355	*DYNC1H1*	Body; Shore	chr12:110.64	0.06	0.001	0.6	0.02
cg15002641	*SOX13*	Body	chr1:133.39	-0.10	0.001	-1.0	0.02
Differentially methylated regions detected by MBD-seq at FDR 10%; replicated by EPIC							
cg05362127	*WNT5A*	TSS200; Island	chr14:28.51	0.33	0.002	2.3	9.4E-06
cg24142850	-	-	chr8:92.55	-0.09	0.005	-2.9	9.4E-05
cg15585318	*WNT5A*	Body; Island	chr14:28.51	0.22	0.006	1.8	2.1E-06
cg09595163	*WNT5A*	Body; Island	chr14:28.51	0.18	0.006	2.3	1.2E-05
cg13868216	*BAIAP2L2*	Body; Island	chr15:79.26	0.11	0.01	1.6	1.8E-04
cg09972454	*PDXDC1*	Body; Shore	chr4:147.94	-0.06	0.01	-2.9	1.5E-06
cg18120446	-	Island	chr5:41.75	0.01	0.02	-2.2	2.7E-08

[1]These are strain-dependent differentially methylated CpGs (EPIC microarray) and CpG regions (MBD-seq) based on a "false discovery threshold" (FDR) cutoff of 10% and replication at an uncorrected p-value threshold of 0.05.
[2]Coef. is the linear regression coefficient (i.e., difference in methylation relative to C57BL/6J; negative is lower methylation in DBA/2J; and positive is higher methylation in DBA/2J compared to C57BL/6J). LogFC is $\log_2$ fold difference in methylation (i.e., difference in methylation relative to DBA/2J; negative is lower methylation in DBA/2J; and positive is higher methylation in DBA/2J compared to C57BL/6J).
[3]CpG location in relation to gene features and CpG region based in probe annotations for the human methylation microarray. For most conserved regions, mouse annotations are analogous to humans.
[4]Chromosome and Megabase coordinate based on mm10 mouse reference genome

Discussion

Given the high sequence conservation between the mouse and human genomes, we used the recently released Illumina MethylationEPIC microarray to assay DNA methylation at conserved CpGs in the mouse genome. We evaluated both the qualitative features as well as the quantitative performance and compared it with MBD-seq data that was generated on the same DNA samples from mice. Such a cross-species approach has been previously used to examine gene expression and perform comparative genomics studies [32-35]. The Illumina methylation array relies on bisulfite conversion and the probes are designed to target bisulfite-converted sequences. The

two older versions of this Illumina methylation microarrays, the Infinium HumanMethylation 27K (HM27) and HumanMethylation 450K (HM450), have been carefully evaluated for use in mice [17]. The number of probes that map to the mouse genome can vary somewhat depending on the alignment algorithm. In the work by Wong et al. [17], alignment to the bisulfite-converted mouse genome resulted in the highest number of conserved probes. Using a stringent parameter of 100% sequence identity to the bisulfite genome, Wong et al. identified a total of 1,308 (4.7% of total) uniquely aligned probes in the 27K array, and 13,715 (2.8% of total) uniquely aligned probes in the 450K array that can be used to interrogate conserved CpGs in the mouse. In our present work, we performed alignment in a non-bisulfite space. While we required unique alignment, we tolerated up to two non-matching base pairs and added detection confidence as another parameter to identify probes that we can use for reliably quantitative assays. With these parameters, we identified 1.6% of total probes (13,665 in the 850K MethylationEPIC array) that aligned uniquely to the mouse genome and associated with high confidence in signal detection. In this set of 13,665 conserved EPIC probes, 9,429 (69%) were CpG loci carried over from the HM450 array and 7,483 of these were also in Wong's list of conserved HM450 probes [17]. While alignment to the bisulfite-converted genome may have yielded a higher proportion of aligned probes, for our purposes the 13,665 probes provided a representative subset that we can use to assess quantitative performance in mouse samples and utility in detecting methylation variation.

The conserved probes mostly target CpGs located within annotated genes and in regulatory regions. In particular, exons, 5' UTRs, CGIs in proximal regulatory regions

(within 200 bp of TSS) are highly overrepresented among the set of 13,665 probes. This is expected since these coding and regulatory regions are the most conserved portions of the genome. Humans and mice have similar complements of CGIs and the genomic positions of these CGIs are also highly conserved, with 50% of CGIs located near annotated TSSs in both species [16,31,36]. In terms of quantitative variation in methylation, CGIs and promoter region CpGs show significant population variation [37]. However, compared to intergenic CpGs, the extent of inter-individual variability in methylation is reported to be much lower in these conserved sites [38,39]. Hence, an obvious limitation in using the conserved EPIC probes is that we attain only a narrow perspective of the mouse methylome and we may be sampling the portion of CpGs that shows the least quantitative variability in a population. Nonetheless, CpGs in regulatory regions and CGIs play crucial roles in development and cell differentiation, and are implicated in tumor development and aging [16,31,36,40,41]. While narrow in perspective, the conserved probes likely represent a subset of CpGs with high functional relevance and application in cross-species study of DNA methylation.

To evaluate the quantitative performance of the EPIC probes, we compared methylation levels measured by a complementary technology, MBD-seq. The type of methylation information measured by the microarray and MBD-seq are somewhat different. The EPIC probes, based on bisulfite conversion, measure the methylation status at a single CpG dinucleotide. MBD-seq, on the other hand, relies on affinity capture of DNA fragments by the methyl-CpG binding domain protein [22-24]. The affinity is directly proportional to the number of methylated CpGs in the DNA fragment and the methylation level is indirectly estimated based on the counts of sequenced reads that

map to that region. This means that the resolution is inherently limited by the sizes of the fragments (in this case ~300 bp). Since methylation of neighboring CpGs is generally correlated [42-44], MBD-seq provides information on the methylation level of CpGs in a region rather than one CpG site. For the 13,665 conserved EPIC probes, we extracted read counts from within 300 bp bins of the targeted CpGs and derived the CpG density-normalized read counts. Overall, there is strong concordance in methylation levels measured by the two technologies and the correlation between the β-values and rms was strongest for CpGs that are moderately methylated (we define these as β-values between 0.3 to 0.7 methylated). However, for CpGs that are hypomethylated and hypermethylated, the rms for the corresponding regions showed a more clustered distribution and indicated a limited quantitative sensitivity for MBD-seq and limited capacity in discerning quantitative variation at such CpG regions. Our observations agree with a previous study that compared HM450 and MBD-seq data generated using the same commercial kit we used [45].

For a direct comparison between the EPIC probes and MBD-seq, we applied the same regression model and crosschecked the DMCpGs and DMRs detected by the two technologies. While we expected a higher quantitative sensitivity for the EPIC probes as to age, the EPIC probes did not detect significant differential methylation at an FDR threshold of 10%. However, the topmost significant probe, cg13269407, is part of the 353 clock CpGs that are used to estimate human DNA methylation age [8]. Consistent with the negative correlation with age in humans, this age-informative CpG was associated with a ~2.4-fold reduction in methylation in the old mice relative to the young mice. Aside from cg13269407, only 10 other human clock CpG probes were in the

conserved set and none of these are associated with age in mice. Overall, the effect of age was weak when we considered individual CpGs. When we examined the corresponding CpG regions, the MBD-seq was more effective at detecting age-dependent methylation. At an FDR cutoff of 10%, we identified seven CpG regions that are classified as age-DMRs. These age-DMRs have been previously reported and show increases in methylation with age in mice [20]. For these age-DMRs identified by MBD-seq, we then checked whether the EPIC probes could verify the age effect. For this cross-verification, we used a less stringent statistical threshold of 0.05 for uncorrected p-values and found that six of the targeted CpGs are also associated with a significant age-dependent increases in β-values. Our observations suggest that age-dependent changes in methylation at these conserved sites may be more pronounced if we consider the correlated change of neighboring CpGs rather than methylation status of a single CpG. Despite the low overall quantitative sensitivity, the MBD-seq provides a complementary approach that may perform better for detecting methylation changes in regions harboring multiple correlated CpGs.

DNA methylation can vary substantially between mouse strains and a large fraction of this is likely due to underlying sequence differences between strains [20,46,47]. Strain variation in methylation has been shown to associate with complex phenotypes in mice such as insulin resistance, adiposity, and blood cell counts [48]. In our analysis, we detected 219 CpGs (i.e., 1.6% of the 13,365 interrogated CpGs) with a significant difference between strains at an FDR cutoff of 10%. A large majority (175 out of 219 CpGs) was associated with higher methylation in D2 compared to B6. While the overall lower methylation in B6 is intriguing, such variation between strains must be cautiously

449 interpreted. It is well known that SNPs in probe sequences can have a strong
450 confounding effect. This is particularly pernicious for mouse specific microarrays in
451 which probe sequences are usually based on the B6 mouse reference, and as a result,
452 there is more efficient hybridization for B6-derived samples, which results in a positive
453 bias for this canonical mouse strain [49-51]. In the present work, since the EPIC array is
454 based on the human sequence, we do not expect a systematic bias for one strain over
455 the other. For replication, we referred to the MBD-seq data and only 29 out of the 219
456 corresponding CpG regions had consistent differential methylation between B6 and D2
457 in the MBD-seq.

458 Unlike using a human array that should not bias one mouse strain over another, the
459 MBD-seq data is more vulnerable to technical artifacts caused by sequence differences.
460 As is the general practice, we performed the alignment of the MBD-seq reads to the
461 mouse reference genome. This means the alignment will be more efficient for
462 sequences from B6, while sequences from D2 will have more mismatches. Since
463 methylation quantification is estimated from the relative number of aligned reads, this
464 may result in a systematic negative bias for D2, and methylation levels in regions with
465 sequence differences will tend to have lower methylation due to poorer alignment. As a
466 result, a higher fraction of strain-DMR will have lower methylation in D2 compared to B6
467 [20]. In the case that these conserved CpGs have higher methylation in D2 compared to
468 B6, then the negative bias will lessen the quantitative difference between the strains.
469 This may explain why the effect of strain is less pronounced in the MBD-seq data. In the
470 MBD-seq, there were only 37 DMRs between B6 and D2 at an FDR threshold of 10%,
471 and the EPIC probes replicated 16 of these. Out of the 37 strain-DMRs, the majority (21

of the 37) was associated with higher methylation in D2. Both the EPIC and MBD-seq therefore show an overall lower methylation profile in B6 compared to D2 that warrants further investigation and verification. Such strain differences in overall methylation has been previously reported for A/J and WSB/EiJ, with the A/J strain exhibiting higher methylation of CGIs in normal liver tissue compared to WSB/EiJ. This difference in the methylome was suggested to contribute to differential susceptibility for nonalcoholic fatty liver disease that characterizes the two strains [46]. In the case of B6 and D2, the two strains are highly divergent in a number of complex phenotypes ranging from behavioral and physiological to aging traits. The panel of recombinant inbred progeny derived from B6 and D2 (the BXD panel) has been used extensively in genetic research [52-56]. If there is indeed a distinct profile in DNA methylation between B6 and D2, then it will be of interest to evaluate if it segregates in the BXDs and how the methylome contributes to some of the phenotypic differences. The BXD panel could be an extremely rich and as yet untapped resource for methylome-wide analysis of complex traits that can then be integrated with the extensive systems genetics work that has already been done with this mouse family [57,58]. No doubt, large-scale analysis of genome-wide DNA methylation in mouse genetic reference panels will be greatly accelerated with the development of a mouse version of the Infinium methylation arrays. And as is the case with other types of arrays, it will be crucial that the probes are designed against a more diverse panel of strains so that investigators can derive a more unbiased readout of methylation [59].

To conclude, we have catalogued a small subset of EPIC probes that target conserved CpGs in the mouse genome and that provide reliable quantification of DNA methylation

in mouse samples. While detection for age-dependent methylation was weaker for the EPIC probes compared to MBD-seq, we have identified significant strain variation in methylation at the conserved CpGs. Our results indicate lower methylation for B6 compared to D2 at sites that have significant strain effect. It is unclear how much of the strain variation results from underlying sequence differences between B6 and D2, and this strain-specific profile needs to be further evaluated and verified

Acknowledgments:

This work was supported by funds from UTHSC Faculty AwardUTCOM-2013KM. We are very grateful to Robert W. Williams for providing advice and access to the aging colony that is funded by NIA NIH grant R01AG043930.

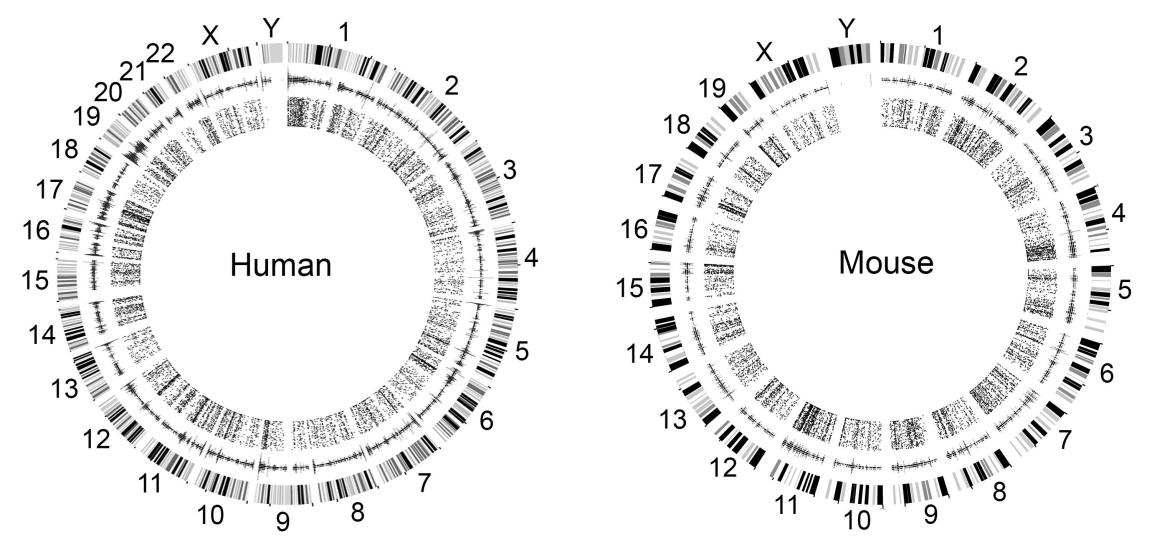
References:

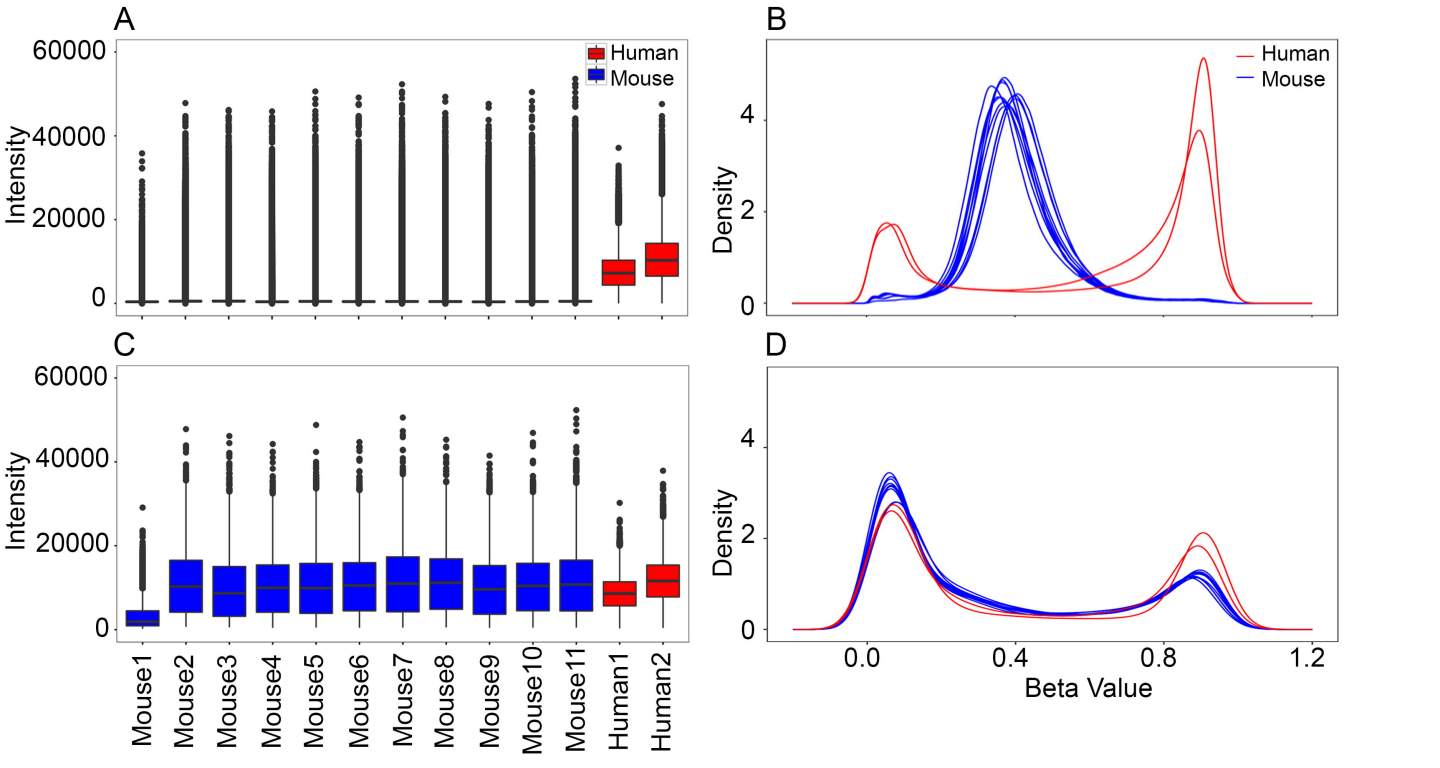
1. Laffita-Mesa JM, Bauer PO, Kouri V, Pena Serrano L, Roskams J, et al. (2012) Epigenetics DNA methylation in the core ataxin-2 gene promoter: novel physiological and pathological implications. Hum Genet 131: 625-638.
2. Simeone P, Alberti S (2014) Epigenetic heredity of human height. Physiol Rep 2.
3. Egger G, Liang G, Aparicio A, Jones PA (2004) Epigenetics in human disease and prospects for epigenetic therapy. Nature 429: 457-463.
4. Portela A, Esteller M (2010) Epigenetic modifications and human disease. Nat Biotechnol 28: 1057-1068.
5. Toperoff G, Aran D, Kark JD, Rosenberg M, Dubnikov T, et al. (2012) Genome-wide survey reveals predisposing diabetes type 2-related DNA methylation variations in human peripheral blood. Hum Mol Genet 21: 371-383.
6. Breton CV, Byun HM, Wenten M, Pan F, Yang A, et al. (2009) Prenatal tobacco smoke exposure affects global and gene-specific DNA methylation. Am J Respir Crit Care Med 180: 462-467.
7. Baccarelli A, Wright RO, Bollati V, Tarantini L, Litonjua AA, et al. (2009) Rapid DNA methylation changes after exposure to traffic particles. Am J Respir Crit Care Med 179: 572-578.
8. Horvath S (2013) DNA methylation age of human tissues and cell types. Genome Biol 14: R115.
9. Swartz JR, Hariri AR, Williamson DE (2017) An epigenetic mechanism links socioeconomic status to changes in depression-related brain function in high-risk adolescents. Mol Psychiatry 22: 209-214.
10. Roth TL, Lubin FD, Funk AJ, Sweatt JD (2009) Lasting epigenetic influence of early-life adversity on the BDNF gene. Biol Psychiatry 65: 760-769.
11. Bibikova M, Barnes B, Tsan C, Ho V, Klotzle B, et al. (2011) High density DNA methylation array with single CpG site resolution. Genomics 98: 288-295.
12. Bibikova M, Le J, Barnes B, Saedinia-Melnyk S, Zhou LX, et al. (2009) Genome-wide DNA methylation profiling using Infinium (R) assay. Epigenomics 1: 177-200.
13. Moran S, Arribas C, Esteller M (2016) Validation of a DNA methylation microarray for 850,000 CpG sites of the human genome enriched in enhancer sequences. Epigenomics 8: 389-399.
14. Pidsley R, Zotenko E, Peters TJ, Lawrence MG, Risbridger GP, et al. (2016) Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. Genome Biol 17: 208.
15. Stubbs TM, Bonder MJ, Stark AK, Krueger F, Team BIAC, et al. (2017) Multi-tissue DNA methylation age predictor in mouse. Genome Biol 18: 68.
16. Deaton AM, Bird A (2011) CpG islands and the regulation of transcription. Genes Dev 25: 1010-1022.
17. Wong NC, Ng J, Hall NE, Lunke S, Salmanidis M, et al. (2013) Exploring the utility of human DNA methylation arrays for profiling mouse genomic DNA. Genomics 102: 38-46.
18. Gardiner-Garden M, Frommer M (1987) CpG islands in vertebrate genomes. J Mol Biol 196: 261-282.

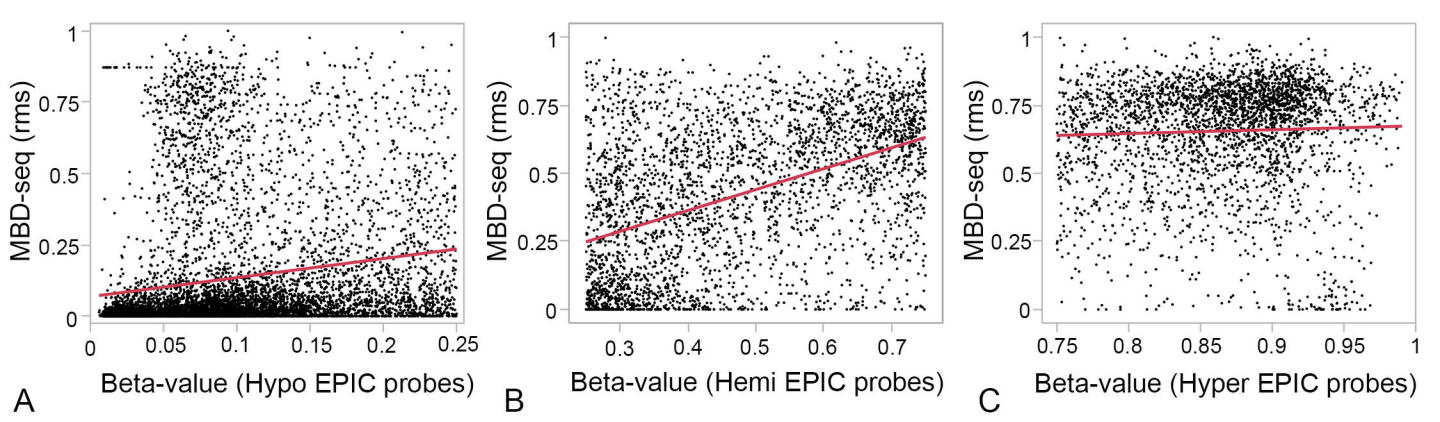
19. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, et al. (2009) Circos: an information aesthetic for comparative genomics. Genome Res 19: 1639-1645.
20. Mozhui K, Pandey AK (2017) Conserved effect of aging on DNA methylation and association with EZH2 polycomb protein in mice and humans. Mech Ageing Dev 162: 27-37.
21. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, et al. (2014) Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics 30: 1363-1369.
22. De Meyer T, Mampaey E, Vlemmix M, Denil S, Trooskens G, et al. (2013) Quality evaluation of methyl binding domain based kits for enrichment DNA-methylation sequencing. PLoS One 8: e59068.
23. Aberg KA, Xie L, Chan RF, Zhao M, Pandey AK, et al. (2015) Evaluation of Methyl-Binding Domain Based Enrichment Approaches Revisited. PLoS One 10: e0132205.
24. Aberg KA, McClay JL, Nerella S, Xie LY, Clark SL, et al. (2012) MBD-seq as a cost-effective approach for methylome-wide association studies: demonstration in 1500 case--control samples. Epigenomics 4: 605-621.
25. Bock C, Tomazou EM, Brinkman AB, Muller F, Simmer F, et al. (2010) Quantitative comparison of genome-wide DNA methylation mapping technologies. Nat Biotechnol 28: 1106-1114.
26. Lienhard M, Grimm C, Morkel M, Herwig R, Chavez L (2014) MEDIPS: genome-wide differential coverage analysis of sequencing data derived from DNA enrichment experiments. Bioinformatics 30: 284-286.
27. Chavez L, Jozefczuk J, Grimm C, Dietrich J, Timmermann B, et al. (2010) Computational analysis of genome-wide DNA methylation during the differentiation of human embryonic stem cells along the endodermal lineage. Genome Res 20: 1441-1450.
28. Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26: 139-140.
29. Benjamini Y, Hochberg Y (1995) Controlling the false Discovery Rate: A Practical and Powerful approach to Multiple Tesing. Journal of the Royal Statistical Society Series B (Methodological) 57: 289-300.
30. Benjamini Y, Drai D, Elmer G, Kafkafi N, Golani I (2001) Controlling the false discovery rate in behavior genetics research. Behav Brain Res 125: 279-284.
31. Illingworth RS, Gruenewald-Schneider U, Webb S, Kerr AR, James KD, et al. (2010) Orphan CpG islands identify numerous conserved promoters in the mammalian genome. PLoS Genet 6: e1001134.
32. Adjaye J, Herwig R, Herrmann D, Wruck W, Benkahla A, et al. (2004) Cross-species hybridisation of human and bovine orthologous genes on high density cDNA microarrays. BMC Genomics 5: 83.
33. Bar-Or C, Czosnek H, Koltai H (2007) Cross-species microarray hybridizations: a developing tool for studying species diversity. Trends Genet 23: 200-207.
34. Oshlack A, Chabot AE, Smyth GK, Gilad Y (2007) Using DNA microarrays to study gene expression in closely related species. Bioinformatics 23: 1235-1242.

35. Lu Y, Huggins P, Bar-Joseph Z (2009) Cross species analysis of microarray expression data. Bioinformatics 25: 1476-1483.
36. Illingworth RS, Bird AP (2009) CpG islands--'a rough guide'. FEBS Lett 583: 1713-1720.
37. Heyn H, Moran S, Hernando-Herraez I, Sayols S, Gomez A, et al. (2013) DNA methylation contributes to natural human variation. Genome Res 23: 1363-1372.
38. Wagner JR, Busche S, Ge B, Kwan T, Pastinen T, et al. (2014) The relationship between DNA methylation, genetic and expression inter-individual variation in untransformed human fibroblasts. Genome Biol 15: R37.
39. Gervin K, Hammero M, Akselsen HE, Moe R, Nygard H, et al. (2011) Extensive variation and low heritability of DNA methylation identified in a twin study. Genome Res 21: 1813-1821.
40. Teschendorff AE, Menon U, Gentry-Maharaj A, Ramus SJ, Weisenberger DJ, et al. (2010) Age-dependent DNA methylation of genes that are suppressed in stem cells is a hallmark of cancer. Genome Res 20: 440-446.
41. Heyn H, Li N, Ferreira HJ, Moran S, Pisano DG, et al. (2012) Distinct DNA methylomes of newborns and centenarians. Proc Natl Acad Sci U S A 109: 10522-10527.
42. Zhang W, Spector TD, Deloukas P, Bell JT, Engelhardt BE (2015) Predicting genome-wide DNA methylation using methylation marks, genomic position, and DNA regulatory elements. Genome Biol 16: 14.
43. Lovkvist C, Dodd IB, Sneppen K, Haerter JO (2016) DNA methylation in human epigenomes depends on local topology of CpG sites. Nucleic Acids Res 44: 5123-5132.
44. Li Y, Zhu J, Tian G, Li N, Li Q, et al. (2010) The DNA methylome of human peripheral blood mononuclear cells. PLoS Biol 8: e1000533.
45. De Meyer T, Bady P, Trooskens G, Kurscheid S, Bloch J, et al. (2015) Genome-wide DNA methylation detection by MethylCap-seq and Infinium HumanMethylation450 BeadChips: an independent large-scale comparison. Sci Rep 5: 15375.
46. Tryndyak VP, Han T, Fuscoe JC, Ross SA, Beland FA, et al. (2016) Status of hepatic DNA methylome predetermines and modulates the severity of non-alcoholic fatty liver injury in mice. BMC Genomics 17: 298.
47. Orozco LD, Rubbi L, Martin LJ, Fang F, Hormozdiari F, et al. (2014) Intergenerational genomic DNA methylation patterns in mouse hybrid strains. Genome Biol 15: R68.
48. Orozco LD, Morselli M, Rubbi L, Guo W, Go J, et al. (2015) Epigenome-wide association of liver methylation patterns and complex metabolic traits in mice. Cell Metab 21: 905-917.
49. Walter NA, McWeeney SK, Peters ST, Belknap JK, Hitzemann R, et al. (2007) SNPs matter: impact on detection of differential expression. Nat Methods 4: 679-680.
50. Ciobanu DC, Lu L, Mozhui K, Wang X, Jagalur M, et al. (2010) Detection, validation, and downstream analysis of allelic variation in gene expression. Genetics 184: 119-128.

51. Bottomly D, Walter NA, Hunter JE, Darakjian P, Kawane S, et al. (2011) Evaluating gene expression in C57BL/6J and DBA/2J mouse striatum using RNA-Seq and microarrays. PLoS One 6: e17820.
52. Sloane LB, Stout JT, Vandenbergh DJ, Vogler GP, Gerhard GS, et al. (2011) Quantitative trait loci analysis of tail tendon break time in mice of C57BL/6J and DBA/2J lineage. J Gerontol A Biol Sci Med Sci 66: 170-178.
53. Peirce JL, Lu L, Gu J, Silver LM, Williams RW (2004) A new set of BXD recombinant inbred lines from advanced intercross populations in mice. BMC Genet 5: 7.
54. Mozhui K, Ciobanu DC, Schikorski T, Wang X, Lu L, et al. (2008) Dissection of a QTL hotspot on mouse distal chromosome 1 that modulates neurobehavioral phenotypes and gene expression. PLoS Genet 4: e1000260.
55. Lang DH, Gerhard GS, Griffith JW, Vogler GP, Vandenbergh DJ, et al. (2010) Quantitative trait loci (QTL) analysis of longevity in C57BL/6J by DBA/2J (BXD) recombinant inbred mice. Aging Clin Exp Res 22: 8-19.
56. de Haan G, Van Zant G (1999) Dynamic changes in mouse hematopoietic stem cell numbers during aging. Blood 93: 3294-3301.
57. Wang X, Pandey AK, Mulligan MK, Williams EG, Mozhui K, et al. (2016) Joint mouse-human phenome-wide association to test gene function and disease risk. Nat Commun 7: 10464.
58. Mulligan MK, Mozhui K, Prins P, Williams RW (2017) GeneNetwork: A Toolbox for Systems Genetics. Methods Mol Biol 1488: 75-120.
59. Morgan AP, Fu CP, Kao CY, Welsh CE, Didion JP, et al. (2015) The Mouse Universal Genotyping Array: From Substrains to Subspecies. G3 (Bethesda) 6: 263-279.


Human
X
Y
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
Mouse
X
Y
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19


A
B
C
D
Intensity
Density
Human
Mouse
Beta Value
Mouse1
Mouse2
Mouse3
Mouse4
Mouse5
Mouse6
Mouse7
Mouse8
Mouse9
Mouse10
Mouse11
Human1
Human2


MBD-seq (rms)
1
0.75
0.5
0.25
0
0
0.05
0.1
0.15
0.2
0.25
Beta-value (Hypo EPIC probes)
A
MBD-seq (rms)
1
0.75
0.5
0.25
0
0.3
0.4
0.5
0.6
0.7
Beta-value (Hemi EPIC probes)
B
MBD-seq (rms)
1
0.75
0.5
0.25
0
0.75
0.8
0.85
0.9
0.95
1
Beta-value (Hyper EPIC probes)
C