Body weight at young adulthood and association with epigenetic aging and lifespan in the BXD murine family

Jose Vladimir Sandoval-Sierra[1], Alexandra H. B. Helbing [1], Evan G. Williams[3], David G. Ashbrook[2], Suheeta Roy[2], Robert W. Williams[2], Khyobeni Mozhui[1,2*]

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

[2]Department of Genetics, Genomics and Informatics, University of Tennessee Health Science Center, College of Medicine, Memphis, Tennessee, USA

[3] Department of Biology, Institute of Molecular Systems Biology, ETH Zurich, Zurich CH-8093, Switzerland

[*]Correspondence:
Khyobeni Mozhui
kmozhui@uthsc.edu

Short running title: Lifespan differences and epigenetic aging in mouse strains

Summary

DNA methylation (DNAm) is shaped by genetic and environmental factors and modulated by aging. Here, we examine interrelations between epigenetic aging, body weight (BW), and lifespan in 12 inbred mouse strains from the BXD panel that exhibit over two-fold variation in longevity. Genome-wide DNAm was assayed in 70 liver specimens from mice ranging in age from 6 to 25 months that were maintained on normal chow or high fat diet (HFD). We defined subsets of CpG regions associated with age, BW at young adulthood, and strain-by-diet dependent life expectancy. The age associated differentially methylated CpG regions (age-DMRs) featured distinct genomic characteristics, with DNAm gains over time occurring in sites with high CpG density and low average methylation. CpG regions associated with BW were enriched in introns and generally showed lower methylation in mice with higher BW, and inversely correlated with gene expression such that mRNA was higher in mice with higher BW. Lifespan-associated regions featured no distinct genomic characteristics but were linked to genes involved in lifespan regulation, including the telomerase reverse transcriptase gene, *Tert,* which showed lower methylation and higher gene expression in long-lived strains. An epigenetic clock defined from the age-DMRs conveyed accelerated aging in mice belonging to strains with shorter lifespans. Both higher BW at young adulthood and HFD were associated with accelerated epigenetic aging. Our results highlight the age-accelerating effect of heavier body weight. Furthermore, the study demonstrates that the measure of epigenetic aging derived from age-DMRs can predict strain and diet-induced differences in lifespan.

Keywords: epigenetics, lifespan, longevity, age acceleration

41 # Introduction

42 In the past few years, the "epigenetic clock" based on DNA methylation (DNAm) has emerged
43 as a robust and widely used biomarker of aging that perhaps surpasses telomere length assays
44 in its accuracy and utility [1]. Often referred to as DNA methylation age (DNAmAge), the CpG
45 based estimator of biological age comes in a few different versions for both humans and mice [2-
46 8]. All these clocks share a common feature—they rely on the methylation status of preselected
47 subsets of CpGs that are each assigned weights and are used collectively to estimate age. A
48 critical question has been: are these DNAmAge clocks detecting changes that are purely a
49 function of time, and therefore, correlates of chronological age? Or are they providing a
50 measure of the intrinsic pace of biological aging that can be related to health, fitness, and life
51 expectancy? Evidence from retrospective human epidemiological studies indicate that certain
52 versions of the clock perform better at measuring biological aging and at predicting life
53 expectancy. In general, a younger DNAmAge relative to chronological age conveys decelerated
54 biological aging, and is associated with a lower risk of disease and increased longevity [4,9-12].

55 In the mouse model, lifespan extending interventions such as calorie restriction (CR) and
56 treatment with rapamycin, or strong genetic mutations that drastically reduce body weight
57 (e.g., *Ames* and *Snell* dwarf mice), have been shown to significantly decelerate the epigenetic
58 clock [5,7,8,13,14]. However, to date, most DNAmAge estimations in the mouse have been modeled
59 on the canonical C57BL/6J (B6) background strain [5-8,13]. Similar to interindividual variation in
60 humans, aging trajectories vary considerably between different mouse strains, and common
61 DNA variants contribute to the pace of normal aging and longevity [15]. Whether the differential
62 rates of epigenetic aging can discern normative lifespan differences between mouse strains has
63 not yet been explored. Furthermore, if body weight reduction due to CR or dwarfing mutations
64 can slow down the clock, then an open question is whether subtle differences in normal body
65 weight can also have an impact on DNAmAge.

66 The BXD recombinant inbred family is the most deeply phenotyped mouse genetic reference
67 panel, and has a long history in aging and longevity research [16-18]. Notably, the different BXD
68 strains exhibit wide variation in natural life expectancy under normal laboratory conditions.
69 Some strains have a mean life expectancy of less than 15 months (e.g., BXD13, BXD5), while
70 other strains are expected to survive for well over 2 years (e.g., BXD19, BXD65) [16,18-20].
71 Longevity data in the BXDs has been collected since the 1980s, and lifespan data continues to
72 be collected from the current, expanded BXD panel that consists of 150 inbred strains [18,20,21].
73 The BXDs are derived by crossing the two parental strains, B6 and DBA/2J (D2), and then
74 inbreeding the F2s, or by advanced intercrosses to increase recombination [21]. Genomes of the
75 BXDs therefore represent random recombinants of the B6 and D2 genomes, where each strain
76 is a unique mosaic of homozygous B6 or D2 alleles. The D2 strain has been considered to have a
77 more accelerated aging profile, and has a significantly shorter lifespan than B6 [15,22]. Other age-
78 associated parameters include rapid thymic involution [23], quicker replicative senescence of
79 hematopoietic stem cells [17], and increased tail tendon breakage in D2 compared to B6 [24]. Due
80 to random assortment of independent gene variants, the progeny BXDs have a much greater

range of variation in lifespan and aging traits [17,20], and provide a unique population in which to dissect the interrelations between epigenetic aging and longevity.

Here, we leveraged the extensive longevity data generated for the BXD family [20] to evaluate associations between body weight, epigenetic aging, and lifespan. We used affinity-capture enrichment with the methyl-CpG-binding domain protein (MBD), followed by deep sequencing (MBD-seq) to profile the aging liver methylome in 12 members of the BXD family [25]. To examine the impact of a common metabolic stressor, we also quantified the methylome on a subset of BXD mice maintained on a high fat diet (HFD), which decreases longevity by as much as 12% [20]. Our goal was to chart differentially methylated regions related to aging, and to strain differences in body weight and lifespan, and to examine correlation with gene expression. We computed a simple DNAm clock using the age-dependent CpG regions, and this conveyed strain differences in rates of epigenetic aging. Overall, our results show that both higher body weight and HFD can have an accelerating effect on epigenetic aging, and this can then be related to lifespan.

Results

BXD strain selection and lifespan and body weight characteristics

The present work is based on data collected from two separate cohorts of BXD mice. Lifespan data were collected from a group of female BXDs, referred to as the "*longevity cohort*", that were allowed to age until mortality [20]. A parallel, identically treated and strain-matched group, the "*biospecimen cohort*", was maintained for tissue collection at different ages, and liver samples were obtained from this. The BXD strains have wide range in natural lifespan when kept on *ad libitum* standard chow (control diet or CD) [20]. HFD reduced longevity, but with marked differences among strains [20]. For DNAm analysis, we selected 12 members of the BXD panel (including F1 hybrids) that were representative of the wide variation in lifespan, and for five of these, we included mice maintained on HFD (**Table 1; Fig 1a**). Ages at natural death from the *longevity cohort* are plotted in **Fig 1a**. BXD65 showed the most significant decline in lifespan on HFD.

Based on lifespan averages, each strain-by-diet group was classified as short-lived (mean lifespan < 600 days), medium-lived (600–750 days), and long-lived (>800 days; no strain on HFD were in this group). Note that a strain classified as long-lived on CD may be classified as short-lived on HFD (**Table 1**). We then selected 70 liver specimens for the corresponding 17 strain-by-diet groups for DNAm assay. These specimens were chosen so that distribution of age at time of tissue collection was closely matched across the lifespan groups **(Fig. 1b**; individual level sample information in **Table S1)**. We note that aside from three male cases for BXD102, B6D2F1, and D2B6F1 (**Table 1**), all liver specimens were from females. While the samples were not chosen on the basis of body or organ weight, there was significant strain variation in body weight when the mice were initially weighed at young adulthood (mean age of 134 ± 81 days), before introduction to HFD (**Fig. 1c)**. We refer to this as baseline body weight or BW0. The long-lived F1s had higher body weights, and this hybrid vigor was apparent with or without the male

cases. The final weight of mice (i.e., weight on day of sample collection) continued to show significant strain variation (**Fig. 1c**). Liver weight appeared fairly consistent across strains (**Fig. 1c**). There was no group difference in BW0 between the two diets (CD = 25 ± 7 g vs. HFD = 23 ± 5 g; *n* = 70). By final weighing, the group on HFD had become significantly heavier when compared to the full set of mice on CD (HFD = 41 ± 12 g vs. CD = 29 ± 8 g, $p < 0.0001$; n = 70), or when compared only to the matched strains on CD (HFD = 41 ± 12 g vs. CD = 26 ± 6 g, $p < 0.0001$; n = 34). The weight of the liver on HFD was slightly heavier but the effect was not statistically significant, likely due to the modest sample size for HFD (HFD = 1.29 ± 0.23 g vs. CD = 1.22 ± 0.23 g; $p = 0.37$, n = 34). At young adulthood, age was a significant predictor of BW0 (r = 0.55, $p < 0.0001$, n = 70). By the age at which samples were collected, the age of mice was no longer correlated with final body weight ($r = 0.01$) or with weight of liver ($r = 0.12$). Instead, BW0 remained a significant predictor of final body (**Fig. 1d**), and liver weight (r = 0.53, $p < 0.0001$, n = 70). When restricted to only the few HFD mice, BW0 was still a significant correlate of final body weight ($r = 0.53$, $p = 0.05$, $n = 15$), but not liver weight ($r = 0.28$, $p = 0.30$).

We next examined correlations between the individual weight measures from the *biospecimen cohort*, and strain-level lifespan from the *longevity cohort*. The F1s exhibited vigor in both longevity and body weight and including the F1s resulted in no significant correlation between weight and lifespan. After excluding the F1s, both BW0 and final weight showed inverse correlations with all the strain-level indicators of lifespan (mean, median, maximum, and minimum lifespan) (**Table S2**). This indicates that mice with higher body weight, even at younger age, are more likely to belong to strains with shorter lifespan (**Fig 1e**). This observation is consistent with the strong inverse correlation between body weight and longevity that is seen in the larger BXD panel [20].

Table 1. Characteristics of selected strains from the BXD family

	Longevity trait					Biospecimen	
Strain/line	**Mean lifespan (days)** [1]	**Median lifespan (days)** [1]	**Range (days)** [1]	**Lifespan Group** [2]	**Diet** [3]	**N** [3]	**Age Range (days)** [3]
B6D2F1	933±86	896	856–1080	long	CD	5 [4]	216–726
BXD102	861±222	891	514–1131	long	CD	5 [4]	183–714
BXD40	585±239	577	172–949	short	CD	8	284–719
BXD48	695±124	684	516–948	med	CD	3	188–731
BXD48	523±152	517	167–821	short	HFD	3	189–595
BXD48a	617±196	670	161–881	med	CD	3	233–604
BXD48a	635±113	650	424–800	med	HFD	3	233–543
BXD65	824±199	896	275–938	long	CD	6	181–711
BXD65	534±128	551	236–736	short	HFD	3 [5]	230–541
BXD65b	726±91	751	536–854	med	CD	4	187–748
BXD73	702±116	687	553–854	med	CD	4	206–759
BXD73	699±112	715	480–889	med	HFD	3	206–694
BXD73b	820±129	807	656–1055	long	CD	3	237–743
BXD73b	742±193	790	154–951	med	HFD	3	237–729
BXD79	417±155	330	241–703	short	CD	7	217–570

BXD9	507±135	462	337–956	short	CD	3	245–548
D2B6F1	771±143	791	585–996	long	CD	4[4]	210–744

[1] Lifespan for strains under standard or high fat diet estimated from an aging cohort of mice co-housed with mice used for biospecimen collection and methylome assays
[2] Groups based on phenotypic lifespan: short = average lifespan < 600 days; med = average 640–750 days, and long = average lifespan > 800 days
[3] Diet, sample size, and age range of mice used to generate methylome data. CD = control diet; HFD = high fat diet
[4] One male case; see Table S1 for individual level data
[5] One case excluded due to uncertain identity

Strain-dependent patterns in global features of the methylome

Following genome-wide MBD-sequencing, we retained a set of 368,300 regions, each 150 bp in length, with sufficient coverage in the 70 samples. The majority of the CpG regions (83%) contained no sequence variants (SNPs or small insertions/deletions) segregating in the BXDs. For the 17% (62,422) with sequence variants, there was an average of 2 ± 1.6 variants within the 150 bp bin. Consistent with the DNA enrichment and filtering protocols, the 368,300 CpG regions were enriched in annotated gene features such as UTR, introns, exons and CpG islands, and also rRNA and LTRs compared to the background genome (**Table S3**).

We started with a principal component analysis (PCA). A plot of the top principal components, PC1 and PC2 (captured 19% and 13% of the variance, respectively), showed clustering of samples by strain identity, irrespective of diet (**Fig. 2a**). The one exception was a BXD65 on HFD that plotted away from the BXD65 cluster, and due to its questionable identity, it was excluded from downstream analyses, and remaining statistical tests were done in 69 samples. Sub-strains (e.g., BXD73/BXD73b; BXD65/BXD65b) also clustered in close proximity. Unsupervised hierarchical clustering confirmed the clustering of samples by strain identity rather than age or diet groups (QC plots in **Fig. S1)**

The top 5 PCs collectively explain 58% of the variance (**Table S1**), and we examined if these were associated with age, diet, body and organ weights, and strain lifespan. Age was not a significant correlate of any of the top 5 PCs. For strains with matched CD and HFD cases, the PCs did not differentiate between the two diets. For the weight measurements, PC1 was a significant negative correlate of BW0 (**Fig 2b**), and final body and liver weights (correlations with and without the F1s in **Table S4**). None of the other PCs were associated with the weight measurements. For the strain longevity data, PC4 (8% of variance) was the strongest correlate of mean, median, minimum, and maximum lifespan (**Fig 2c**; **Table S4**). PC2 and PC3 (11% of variance) were also significantly correlated with maximum lifespan (**Table S4**).

We computed the overall mean methylation for genic regions (i.e., CpG regions that overlap annotated gene features; 200,531 CpG bins in total), and intergenic regions (167,769 CpG bins), to examine whether these global features would explain the variance captured by the top PCs. Global methylation was highly strain specific (**Fig S2a, S2b**; **Table S1** for individual level data), and the methylation of intergenic CpGs was significantly correlated with PC1 and PC3 (**Fig S2c, S2d**). Mean methylation of genic CpGs was significantly correlated with PC4. However, mean

183 methylation values were not directly correlated with the longevity traits. For the weight
184 measurements, only the mean methylation of intergenic sites showed a significant positive
185 correlation with liver weight (**Table S4**).

186 Taken together, the methylome-wide analysis showed that the PCs capture strain-dependent
187 differences in overall mean methylation. These PCs in turn were associated with body weight,
188 and strain-level lifespan. Age and diet were however not associated with the large-scale
189 methylome features.

190 Characterizing differentially methylated CpG regions

191 We next performed epigenome-wide tests to identify CpG regions that were associated with
192 age, BW0, and strain median lifespan using a multiple linear mixed-model. Although aging did
193 not have a generalized impact on global DNAm features, regression tests showed that aging
194 was a significant predictor of site-specific DNAm at few CpG regions (**Fig 3a**). For the main effect
195 of age, 26 regions, covering 319 CpGs, were above the 10% Bonferroni threshold (unadjusted p
196 $\leq 2.7 \times 10^{-7}$). These strong age-dependent differentially methylation regions (age-DMRs) were
197 mostly located within genes, and 25 of the 26 bins were associated with increased methylation
198 with age (age-hypermethylation; **Table S5**). Although BW0 and the lifespan traits had strong
199 associations with global DNAm patterns, after partly accounting for the strain-dependent
200 effects with the mixed model, only a few CpG regions were significant at the 10% Bonferroni
201 threshold (**Fig 3b, 3c)**. For the main effect of BW0, only seven CpG regions were significant, and
202 all these BW0 associated differentially methylation regions (BW0-DMRs) had lower methylation
203 among mice with higher BW0 (negative regression estimates; **Table S6**). For the strain-level
204 median lifespan, only three CpG regions were significant at the same 10% Bonferroni threshold
205 **(Table S7**).

206 Given the non-independence of adjacent CpG regions, we then applied a lenient threshold of
207 uncorrected $p \leq 1.0 \times 10^{-4}$ to define the general characteristics of the sites associated with age,
208 BW0, and lifespan. In total, 306 CpG regions were age-DMRs at this suggestive threshold (**Fig**
209 **3a; Table S5**). A few of the age-DMRs were weakly associated with BW0 or lifespan; for
210 instance, the top 5 age-DMRs were all age-hypermethylated and also had weak negative
211 associations with median lifespan ($p \leq 0.05$; **Table S5**). However, for the most part, the age-
212 DMRs had only modest associations with BW0 and lifespan. Of the 306, 57% were age-
213 hypermethylated (**Table 2**). Compared to the background set of 368,300 CpG bins, the age-
214 DMRs were highly enriched in genic regions such as promoters and exons, and CpG island, and
215 depleted in intergenic regions (enrichment and depletion p-values in **Table S3**). For each age-
216 DMR, we computed the average methylation and CpG density, and compared these to the age
217 regression coefficients, which convey the change as a function of age. The regression estimates
218 appeared to be highly dependent on local genomic characteristics, and the most pronounced
219 changes involved age-hypermethylation (positive regression coefficients) in bins with high CpG
220 density (**Fig 3d**) and lower average methylation (**Fig 3e**). In contrast, age-DMRs that lost
221 methylation with age (age-hypomethylated; negative regression coefficients) featured lower
222 CpG density, and higher average methylation (**Fig 3d, e**). Functional annotations of these

223 regions identified significant enrichment in gene sets involved in establishment of cellular
224 polarity (e.g., *Wnt5a*, *Lrrd1*, *Ptk7*) (**Table S8**). We consulted the human GWAS catalogue to
225 identify genes represented by the age-DMRs that have been significantly associated with
226 human aging and longevity (**Table S5**), and this identified only one gene, *Cux2* [26].

227 At the suggestive threshold, 689 CpG regions were classified as BW0-DMRs, and the majority of
228 these (517 or 75%) were negatively correlated with BW0 (**Table 2**). Introns was the most
229 enriched gene feature (**Tables S3, S6**). Compared to the age-DMRs, the BW0-DMRs were in
230 regions with relatively lower CpG density and had only 3 regions within CpG islands (**Fig 3f**). The
231 regions that were negatively correlated with BW0 in particular, featured lower CpG densities
232 and also lower mean methylation levels (**3g**). Overrepresented gene ontologies (GO) included
233 regulation of protein homooligomerization, actin cytoskeleton reorganization, and GTPase
234 mediated signal transduction (**Table S8**). We also note that that several of these genes have
235 been associated with weight in human GWAS, including *Hmga2*, *Fto*, and *Ntrk2* (**Table S6**) [27-29].

236 For lifespan, there were 124 LS-DMRs (**Table S7)**, and a slight majority of these (59%, mostly
237 intergenic sites) were positively correlated with higher median lifespan (**Table 2**). The LS-DMRs
238 did not show any enrichment or depletion in gene features compared to the background set,
239 and the LS-DMRs featured lower CpG density (**Fig 3h**). There was no difference in methylation
240 levels between the DMRs that had positive or negative regression estimates (**Fig 3i**). The LS-
241 DMRs were not enriched in any particular biological GO category. However, there was a
242 significant enrichment in genes that result in premature death in single knockout mice, and this
243 included an intronic DMR in the telomerase reverse transcriptase gene (*Tert*) (**Table S8**). Also
244 enriched were genes related to abnormal eating behavior and a decrease in body mass index
245 (e.g., *Igfbp3*, *Mc4r*, *Lpar1*), and genes related to abnormal mineral levels (*Calcr*, *Sptb*, *Wwox*).

246 In terms of the potential effect of underlying sequence differences, we examined what fraction
247 of the DMRs were sites that overlapped DNA variants segregating in the BXDs (**Table 2**). Since
248 the background set has 62,422 bins with variants, the 51 variant containing age-DMRs did not
249 represent a biased enrichment of such regions (hypergeometric enrichment $p = 0.58$). The
250 BW0-DMRs and LS-DMRs were only slightly enriched in sequence variants (enrichment $p = 0.02$
251 and $p = 0.09$, respectively) (**Table 2**).

252 **Table 2. Tally of differentially methylated regions**

	Positive coefficients			Negative coefficients			
	N	Genic	Intergenic	N	Genic	Intergenic	Variants
Age-DMR	175	146 (48%)	29 (9%)	131	94 (31%)	37 (12%)	51 of 306 (17%)
BW0-DMR	172	93 (14%)	79 (11%)	517	368 (53%)	149 (22%)	137 of 689 (20%)
LS-DMR	73	25 (20%)	48 (39%)	51	31 (25%)	20 (16%)	27 of 124 (21.77%)

253

254 **Relating differentially methylated regions to gene expression**

255 Fifty-two of the samples with liver MBD-seq data also had matched liver RNA-seq data, and we
256 used this group to examine whether the DMRs were associated with gene expression. We
257 linked the age-, BW0-, and LS-DMRs to the corresponding transcript. Of the 306 age-DMRs, 279
258 were paired with the corresponding mRNA. For these pairs, we tested how many of the DMRs
259 were *cis*-correlated with the gene expression, and how many of the transcripts were also
260 correlated with age of mice (**Table S9**). At a nominal p of 0.05 ($|r| = 0.27$), 79 age-DMRs (28% of
261 age-DMRs linked to transcripts) were correlated with the expression of cognate genes (**Fig 4a**).
262 The age-hypermethylated DMRs mostly showed positive correlations with gene expression such
263 that the transcripts also showed increased expression with age (e.g., *Jak3*, *Amn*, *Tradd*). The
264 age-hypomethylated DMRs, on the other hand, were more likely to be negatively correlated
265 with gene expression such that the transcripts showed an increase with age (*Nfkbia*, *Slit3*).

266 For the BW0-DMRs, 614 of the CpG regions were paired to corresponding transcripts. At $|r| =$
267 0.27 ($p \leq 0.05$), 121 (~20%) BW0-DMRs were correlated with gene expression (**Table S10**). The
268 overall pattern indicated that the DMRs that had lower methylation in mice with higher BW0
269 (i.e., negative regression estimates) tended to be negatively correlated with gene expression
270 such that there was higher expression of these genes in mice with higher BW0 (**Fig 4b**). This
271 included *Mc5r*, *Nfkb1*, and *Tcf4*. Few of the DMRs that had positive associations with BW0 were
272 negatively correlated with gene expression, and the corresponding transcripts were also
273 upregulated in mice with higher BW0 (e.g., *Aldh18a1*, *Madd*, *Rap2a*).

274 For the LS-DMR, 111 paired to corresponding transcripts, and of these, only 19 CpG regions
275 were *cis*-correlated with gene expression (**Fig 4c; Table S11**). The strongest *cis*-correlations
276 were between the LS-DMRs linked to *Hpse* and *Dapp1* ($r = -0.70$ for both). Both CpG regions
277 had positive regression values for lifespan, and the corresponding transcripts had lower
278 expression in mice belonging to strains with longer lifespan. The LS-DMR in *Tert* was also
279 negatively correlated with gene expression (**Fig 4c**, $r = -0.32$, $p = 0.02$), and the expression of
280 *Tert* had a modest positive correlation with lifespan ($r = 0.26$, $p = 0.07$). Comparison of *Tert*
281 methylation and gene expression among the three lifespan groups showed that the longer lived
282 strains had significantly lower methylation levels compared to both the medium- and short-
283 lived groups ($F_{2,66} = 7.67$, $p = 0.001$), and the long-lived mice had significantly higher gene
284 expression compared to the short-lived mice ($F_{2,49} = 3{,}31$, $p = 0.04$; **Fig 4d**).

285 Measure of age-acceleration from age associated CpG regions

286 We next evaluated whether we could derive a measure of differential rate of aging from the
287 age-DMRs that could be predictive of lifespan differences. We first summarized the age-
288 dependent changes by computing the weighted averages for the set of 306 age-DMRs. The
289 weighted averages, as expected, had a strong correlation with the chronological age of mice,
290 and for a more direct comparison, the values were scaled to the age range for the 69 samples
291 (**Table S1**). We refer to this DMR-based age estimate as DMRmAge, and this showed a nearly
292 linear correlation with chronological age at $r = 0.88$ ($p < 0.0001$, n =69; **Fig 4a**). Age acceleration
293 was derived as the residuals from the regression of DMRmAge on chronological age [7,9], with
294 positive values indicating an older or accelerated biological age, and negative values indicating

295 a decelerated biological age, and we refer to this as DMRmAge-acc (**Table S1**). The DMRmAge-
296 acc showed a significant negative correlation with strain maximum lifespan (r = –0.35, *p* =
297 0.003, n = 69; **Fig 4b**). It was also negatively correlated with mean and median lifespan, but this
298 was not statistically significant (r = –0.13, *p* = 0.28 for both mean and median lifespan). The
299 correlation between DMRmAge-acc and maximum lifespan remained significant when the
300 analysis was limited to only the CD mice (r = –0.30, *p* = 0.03, n = 55). There were notable strain
301 differences in average DMRmAge-acc (**Fig 4c**). These indicate that mice belonging to strain-diet
302 groups with lower maximum lifespan have an accelerated rate of aging. We note that BXD65,
303 which was the strain with the largest reduction in lifespan on HFD (**Fig 1a**) also showed a much
304 higher age acceleration in the HFD group (mean DMRmAge-acc = 54.94) compared to the CD
305 group (mean DMRmAge-acc = –0.17).

306 The DMRmAge-acc showed a significant positive correlation with both BW0 (r = 0.23, *p* = 0.05;
307 **Fig 4c**) and final body weight of mice (r = 0.31, *p* = 0.009) that suggests more accelerated aging
308 with higher body weight. Limiting to only the CD mice (n = 55), the correlation with BW0
309 became slightly stronger (r = 0.29, *p* = 0.03), but the correlation with final weight became
310 weaker (r = 0.23, *p* = 0.09). This suggests that higher body weight at baseline is associated with
311 accelerated aging. We tested the effect of diet on the DMRmAge-acc in the BXDs that had
312 matched samples from both diets (n = 33), and this showed a significantly higher age
313 acceleration in the HFD group (36.13 ± 27.32 in HFD, –3.22 ± 36.99 in CD, *p* = 0.002; **Fig 4d**).

314 Taken together, the analysis demonstrates three points. First that the age-DMR based
315 estimates of age acceleration is predictive of strain dependent differences in lifespan; second,
316 that higher body weight at younger age is associated with more accelerated aging; and third,
317 HFD significantly accelerates aging, as measured by DMRmAge-acc.

318 Discussion

319 Genomic features of differentially methylated regions

320 In the present study, we parsed the variance in the methylome and defined site-specific
321 methylation differences that may be attributed to a strain-level phenotype, median lifespan,
322 and two individual-level variables—age and BW0. For the age-DMRs, the time-dependent
323 patterns were consistent with previous reports [14,30]. As in Sziráki et al. [14], methylation loss over
324 time occurred mostly in regions with higher average DNAm, and methylation gains occurred
325 mostly in regions with lower average DNAm. Since DNAm was quantified over 150 bp non-
326 overlapping bins, we were also able to relate the differential methylation patterns to the local
327 CpG density. In particular, for the age-hypermethylated DMRs, the increase in DNAm with aging
328 had a strong positive correlation with CpG density. This too, is consistent with reports that CpG
329 dense regions—a feature of CpG islands, which typically remain unmethylated—are the sites
330 that tend to gain methylation with age [31]. The genes represented by the age-DMRs included a
331 few notable members such as *Cyp46a1* and *Abca7*, which are involved in cholesterol
332 metabolism and implicated in Alzheimer's disease [32], and few members of the WNT-signaling
333 and mesenchyme developmental pathways (e.g., *Fzd1*, *Fzd8*, *Wnt5a*, *Jak3, Ptk7*, *Nrp2*) . The

334 current data replicated the CpG islands in *C1ql3* and *Ptk7*, which we previously reported as age-
335 hypermethylated sites in the BXD parental strains, B6 and D2 [33]. A region-based functional
336 annotation analysis revealed a significant enrichment in genes involved in cell polarity, an
337 aspect of cells that is established during development through interaction with the WNT
338 polarity signaling pathway, and which becomes dysregulated during aging [34,35]. In terms of the
339 potential impact on gene expression, the general pattern was for the cognate genes to increase
340 in expression with age regardless of whether the corresponding CpG regions were age-
341 hypomethylated or age-hypermethylated.

342 Body weight, even at young adulthood and before introduction to HFD, was significantly
343 associated with the global methylome patterns, and was also predictive of strain longevity. We
344 therefore included BW0 as a predictor variable in the multiple regression model. BW0 was
345 associated with 689 CpG regions, most at the suggestive level, and these BW0-DMRs were
346 significantly enriched in GO categories related to protein assembly, rho and GTPase mediated
347 signaling, and oligosaccharide metabolism. While genes related to fat metabolism or adiposity
348 were not enriched, we note that there were a few genes that have been previously associated
349 with body weight by human GWAS at $p < 5e\text{-}08$ [26]. This included an intron DMR in the fat mass
350 and obesity associated *Fto* gene, which plays a key role in energy homeostasis, and consistently
351 shown to influence body weight in humans that may be mediated by differences in DNAm and
352 gene expression [36]. The BW0-DMRs were mostly located within genes, particularly introns, and
353 although the differential methylation patterns showed some relation with the local CpG counts,
354 this was not as pronounced as in the case of the age-DMRs. The more striking aspect of the
355 BW0-DMRs was that 75% were negatively associated with BW0. For the BW0-DMRs that were
356 linked to the corresponding transcript, the overall pattern indicated a negative correlation
357 between DNAm and gene expression. This means that while DNAm at these sites were, in
358 general, inversely correlated with body weight, the transcript levels of the corresponding genes
359 generally had higher expression in mice that were heavier.

360 For the longevity trait, only 124 CpG regions were significant at the suggestive threshold. These
361 sites were not enriched in any particular genomic feature or GO category, and did not show a
362 strong dependence on CpG density or mean methylation. A point of distinction for the LS-DMRs
363 is that, while the age- and BW0-DMRs were related to individual-level variables, the LS-DMRs
364 were related to life expectancy based on the strain and the diet. This indirect association with
365 the phenotype may explain why only few DMRs were uncovered by the current mixed model.
366 Given this small number of genes, it is particularly striking that the region-based annotation
367 revealed that genes that cause premature death in single knockout mice are the most enriched
368 gene set among the LS-DMRs (phenotype ID MP:0002083, defined as "death after weaning, but
369 before the normal lifespan" in http://www.informatics.jax.org). For these functionally critical
370 genes that result in premature death when deleted, our results imply that quantitative
371 differences in DNAm at these sites can modulate strain variation in normal lifespan. Among the
372 LS-DMRs, both *Tert* and *Igfbp2* have also been linked to aging and longevity by human GWAS
373 [37,38].
374

375 **Building clocks from age dependent CpG regions**

376 Currently, there are several different versions of the DNAmAge estimator available for both
377 mice and humans, some optimized for a specific tissue, and others with multi-tissue application
378 [2,3,5-8]. The standard protocol for developing DNAmAge clocks starts by applying a regression
379 algorithm in a training dataset, followed by age estimation in validation cohorts to gauge the
380 accuracy of the clock. Our goal in this study was not to develop another DNAmAge clock. Given
381 the sample size of the present study, that would not have been a feasible pursuit. Instead, our
382 goal was to test whether the age-dependent CpG regions could discern lifespan differences
383 between the mouse strains, and how the rate of epigenetic aging relates to body weight and
384 diet. For this, we simply summarized the age-DMRs by computing the weighted averages for
385 each sample. The weighted averages, unsurprisingly, correlated strongly with the chronological
386 age of mice. More importantly, the age acceleration derived from the age-DMR based clock was
387 (a) inversely correlated with the strain lifespan phenotype, (b) was significantly more
388 accelerated in the HFD group, and (c) was positively correlated with body weight. This conveys
389 that the age-DMRs can estimate genetically modulated differences in rates of biological aging
390 and lifespan, and is a modifiable outcome that is accelerated by HFD. Furthermore, the results
391 highlight the interdependence between body weight, diet, and health and aging, and our
392 observations agree with the well-known influence of body mass on longevity, and the more
393 favorable health profile associated with lower body mass [39].

394 The strain with the most decelerated clock, and presumably slowest rate of biological aging,
395 was BXD102 on CD, which is also the longest-lived BXD strain we had in the study (**Table 1**). For
396 these mice, the DMRmAge-acc for the females ranged from –78 to –196 days. Interestingly, the
397 one male BXD102 had the least decelerated clock among the BXD102 samples at –66 days. The
398 most accelerated epigenetic aging was seen for the short-lived BXD9 on CD, and BXD65 on HFD.
399 However, we note that there were a few mismatches between strain lifespan classification and
400 the DMRmAge-acc. BXD40 on CD, although classified as short-lived, was the only short-lived
401 strain with a mean negative DMRmAge-acc. The F1 hybrids, on the other hand, although
402 classified as long-lived, were the only long-lived groups with mean positive DMRmAge-acc.
403 Unlike the inbred BXDs, the heterozygous F1s have hybrid vigor in both body weight and
404 lifespan, and this likely explains the inconsistency.

405 Technical considerations and caveats

406 Before concluding, we should address a few caveats. The sequence alignment was done to the
407 mm10 B6 reference genome, which means that for regions with genetic variants segregating in
408 the BXDs, the sequence differences could compromise alignment. The strong "population
409 stratification" in the PC plot, reminiscent of human genetic data, is likely the result of true
410 quantitative variation in methylation, and also due to a portion of the CpG regions serving as
411 surrogates for underlying genotype [40]. Aging is independent of genetic background and the age-
412 DMRs are expected to be less susceptible to the confounding effect of DNA sequence variants.
413 For body weight and lifespan, the interpretation is complicated by the fact that both
414 phenotypes are closely linked to genotype, and the DMRs may reflect true differences in DNAm
415 levels, or differential quantification due to sequence effects. To partly control for this, we used
416 a mixed model that fitted each strain-diet group as a random intercept. For the age-DMRs, 17%

417 of the CpG regions contained sequence variants, and this is similar to the 17% of variant
418 containing bins in the background set of 368,300 regions. The BW0-DMRs and LS-DMRs were
419 only slightly enriched in variant containing bins, and for the most part, the CpG regions were
420 devoid of sequence differences.

421 To our knowledge, almost all existing DNAm clocks have relied on bisulfite-based assays, which
422 have the advantage of providing single CpG resolution [2-8] . For the mouse model, the reduced
423 representation data (RRBS) generated by Petkovich et al. [5] from ~141 B6 mice, has been used in
424 different studies to define tens of thousands of age dependent CpG sites (over 43,000 CpGs in
425 Lowe et al. [41], and over 146,000 in Szirki et al. [14]). Compared to these, the present work
426 identified only 306 age-DMRs at the suggestive threshold. These CpG regions cover 2,691 CpG
427 sites, and this is still a relatively modest number of age-dependent CpGs. This is likely a result of
428 the smaller sample size, and the rigorous statistical model due to the different genotypes that
429 constituted the present cohort. Another contributing factor may be the methylome assay we
430 used, as MBD-seq provides lower quantitative sensitivity than the bisulfite based assays [42].
431 Nonetheless, MBD-sequencing still delivers highly sensitive and replicable quantification of
432 genome-wide methylation [25]. Using alternate assay methods also demonstrates that the DNAm
433 age estimators are robust to the techniques used to quantify DNAm.

434 Conclusion

435 In conclusion, our results demonstrate that the epigenetic clock defined from age-DMRs is
436 sensitive to subtle differences in natural lifespan that arise from common genetic variants, and
437 is modifiable by environmental interventions, such as diet. The intercorrelations between
438 epigenetic aging, body weight, and longevity also provide evidence that the methylome could
439 provide a mechanistic link between the well-known effect of body mass on aging and lifespan.

440 Experimental Procedures

441 Statistical analysis

442 Descriptions of sample processing for MBD-seq, sequence alignment, initial bioinformatics, and
443 data filtering are provided in **Appendix S1**. In brief, DNAm quantification was done by counting
444 sequenced reads at every 150 bp non-overlapping bins. After removing bins with low coverage,
445 we retained 368,300 CpG regions for further analysis. PCA and hierarchical clustering was done
446 in R for this full set of bins (**Fig S1a**). These detected no outlier samples and the DNAm profiles
447 were consistent across the samples (**Fig S1b**), and averaged at 3.8 ± 0.72 logRPKM. The CpG
448 regions were then divided into bins that occurred within annotated genes (genic set, 200,531
449 bins), and those that were in intergenic regions (167,769 bins). For each sample, the overall
450 average methylation and variance for these genic and intergenic sets were computed. The
451 intercorrelations between the large-scale methylome features, body weight measures, and
452 strain-level lifespan phenotype were examined using Pearson correlations.

453 To detect DMRs, we applied the following model using the lme4 R package (v4_1.1-21) [43]:
454 lmer(logRPKM ~ age + BW0 + medianLifeSpan + (1|StrainDiet)). Manhattan plots were
455 generated using the qqman R package [44]. For the DMRs, we performed two types of enrichment
456 analyses. First, we evaluated relative enrichment in genomic features (i.e., introns, exons, CpG
457 islands, etc.) compared to the background set using the hypergeometric test in R (R codes in
458 **Table S3**). Following that, we carried out functional annotation and enrichment analysis against
459 the whole background genome using the GREAT application (version 4.0.4) [45]. We followed the
460 GREAT recommendation of calculating enrichment using both a binomial test and a
461 hypergeometric test, and reporting only categories that are significant by both methods (FDR <
462 0.05). To search for human GWAS hits, we referred to the GWAS catalogue [26], and searched for
463 the terms "body weight", "aging", and "longevity". For variants with reported association $p \leq$
464 5.0e-08, the mapped human gene symbols were then matched to the corresponding mouse
465 genes associated with the DMR.

466 **Estimating epigenetic age and age acceleration**

467 To estimate epigenetic age, we computed the weighted averages with each of the 306 age-
468 DMRs weighed by the respective age regression coefficient. The weighted averages were then
469 scaled to the age range in the 69 samples using the following formula: DMRmAge =
470 {(((weighted.average – min.weighted.average) x age.range) / weighted.average.range) +
471 min.age}, where min.weighted.average and weighted.average.range are the minimum value
472 and range for the weighted averages in the 69 BXD samples, and age.range = 578 days is the
473 range of chronological age in the 69 BXDs, and min.age = 181 days is the minimum age for the
474 69 BXDs. As recommended in Thompson et al. 2018 [7], the "age acceleration" was computed as
475 the residuals after fitting the predicted age to chronological ages: residuals(lm(DMRmAge ~
476 Age)).

477 **Transcriptomes analyses**

478 We used liver gene expression data that are available from GeneNetwork
479 (http://genenetwork.org; data accession ID GN877, UTHSC BXD Harvested Liver RNA-Seq
480 (Oct19)). While this is a bigger cohort, we limited the analysis to only 52 mice that had matched
481 MBD-seq in the present study. Gene expression was $\log_2$ transformed transcripts per million
482 (TPM). For the age-, BW0-, and LS-DMRs, the cognate genes were defined as the gene in which
483 the DMR is located for genic sites, or the nearest gene promoter for intergenic regions. These
484 were then matched by gene symbol to the corresponding transcript. For *cis*-correlations, we
485 performed a Pearson correlation between the DNAm levels and transcripts levels. In cases
486 where a DMR paired to multiple transcript variants, we retained only the DMR-mRNA pair that
487 had the strongest *cis*-correlation such that each DMR was paired to only a unique mRNA.
488 Following this, we used Pearson correlations to relate the expression levels with the
489 corresponding variable, i.e., age, BW0, or median lifespan.

490 **Acknowledgements**

The present work relied on resources from the BXD Aging Colony, and we thank the entire team, particularly Dr. Lu Lu, Jesse Ingels, Casey J Chapman, Melinda S McCarty, Arthur Centeno, and Dr. Megan K Mulligan. We thank Dr. Saunak Sen for his invaluable comments on the statistics. We thank the UTHSC-Rhodes College Population Health Researcher Program and Dr. Teresa Waters for support and for bringing an excellent summer student to the PI's lab. We thank Dr. Karolina A. Aberg and her team at the Virginia Commonwealth University for their advice on the MBD-seq technique. This study was funded by the NIH NIA grants R21AG055841 and R01AG043930.

Conflict of interest. No competing interests

Author contribution. KM supervised the study, contributed to conception of study design and analysis, and wrote the manuscript. RWW is the principal investigator of the BXD Aging Colony and contributed to the conception of the aging study, and provided access to the biorepository resource. JVSS, AHBH, and EGW contributed to the lab work, and JVSS performed the primary bioinformatics and initial data processing. DA and SR contributed to analysis and data acquisition. All authors contributed to and approved the final version of the manuscript.

Data availability. The normalized MBD-seq data for the 368,300 CpG bins (including chromosomal coordinates, region annotations, CpG density, and variant counts) are available from the NCBI NIH Gene Expression Omnibus. The full raw fastq files are available from the NCBI NIH Sequence Repository Archive. GEO accession ID is GSE137277, and the linked SRA accession ID is SRP221380. (currently set to private but will be made available upon official publication)

Ethics approval. All animal procedures were in accordance to protocol approved by the Institutional Animal Care and Use Committee (IACUC) at the University of Tennessee Health Science Center.

515 **References.**

516 1. Marioni RE, Harris SE, Shah S, et al. The epigenetic clock and telomere length are
517 independently associated with chronological age and mortality. *International journal of*
518 *epidemiology.* 2018;47(1):356.
519 2. Hannum G, Guinney J, Zhao L, et al. Genome-wide methylation profiles reveal
520 quantitative views of human aging rates. *Molecular cell.* 2013;49(2):359-367.
521 3. Horvath S. DNA methylation age of human tissues and cell types. *Genome biology.*
522 2013;14(10):R115.
523 4. Levine ME, Lu AT, Quach A, et al. An epigenetic biomarker of aging for lifespan and
524 healthspan. *Aging.* 2018;10(4):573-591.
525 5. Petkovich DA, Podolskiy DI, Lobanov AV, Lee SG, Miller RA, Gladyshev VN. Using DNA
526 Methylation Profiling to Evaluate Biological Age and Longevity Interventions. *Cell*
527 *metabolism.* 2017;25(4):954-960 e956.
528 6. Stubbs TM, Bonder MJ, Stark AK, et al. Multi-tissue DNA methylation age predictor in
529 mouse. *Genome biology.* 2017;18(1):68.
530 7. Thompson MJ, Chwialkowska K, Rubbi L, et al. A multi-tissue full lifespan epigenetic
531 clock for mice. *Aging.* 2018;10(10):2832-2854.
532 8. Wang T, Tsui B, Kreisberg JF, et al. Epigenetic aging signatures in mice livers are slowed
533 by dwarfism, calorie restriction and rapamycin treatment. *Genome biology.*
534 2017;18(1):57.
535 9. Marioni RE, Shah S, McRae AF, et al. DNA methylation age of blood predicts all-cause
536 mortality in later life. *Genome biology.* 2015;16(1):25.
537 10. Chen BH, Marioni RE, Colicino E, et al. DNA methylation-based measures of biological
538 age: meta-analysis predicting time to death. *Aging.* 2016;8(9):1844-1865.
539 11. Marioni RE, Shah S, McRae AF, et al. The epigenetic clock is correlated with physical and
540 cognitive fitness in the Lothian Birth Cohort 1936. *International journal of epidemiology.*
541 2015;44(4):1388-1396.
542 12. Lu AT, Quach A, Wilson JG, et al. DNA methylation GrimAge strongly predicts lifespan
543 and healthspan. *Aging.* 2019;11(2):303-327.
544 13. Cole JJ, Robertson NA, Rather MI, et al. Diverse interventions that extend mouse
545 lifespan suppress shared age-associated epigenetic changes at critical gene regulatory
546 regions. *Genome biology.* 2017;18.
547 14. Sziraki A, Tyshkovskiy A, Gladyshev VN. Global remodeling of the mouse DNA
548 methylome during aging and in response to calorie restriction. *Aging cell.*
549 2018;17(3):e12738.
550 15. Yuan R, Tsaih SW, Petkova SB, et al. Aging in inbred strains of mice: study design and
551 interim report on median lifespans and circulating IGF1 levels. *Aging cell.* 2009;8(3):277-
552 287.
553 16. de Haan G, Gelman R, Watson A, Yunis E, Van Zant G. A putative gene causes variability
554 in lifespan among genotypically identical mice. *Nature genetics.* 1998;19(2):114-116.
555 17. De Haan G, Van Zant G. Genetic analysis of hemopoietic cell cycling in mice suggests its
556 involvement in organismal life span. *FASEB journal : official publication of the Federation*
557 *of American Societies for Experimental Biology.* 1999;13(6):707-713.

558 18. Gelman R, Watson A, Bronson R, Yunis E. Murine chromosomal regions correlated with
559 longevity. *Genetics.* 1988;118(4):693-704.
560 19. Lang DH, Gerhard GS, Griffith JW, et al. Quantitative trait loci (QTL) analysis of longevity
561 in C57BL/6J by DBA/2J (BXD) recombinant inbred mice. *Aging clinical and experimental*
562 *research.* 2010;22(1):8-19.
563 20. Roy S, Sleiman MB, Jha P, et al. Gene-by-environmental modulation of longevity and
564 weight gain in the murine BXD family. *bioRxiv.* 2019:776559.
565 21. Ashbrook DG, Arends D, Prins P, et al. The expanded BXD family of mice: A cohort for
566 experimental systems genetics and precision medicine. *bioRxiv.* 2019:672097.
567 22. Yuan R, Peters LL, Paigen B. Mice as a mammalian model for research on the genetics of
568 aging. *ILAR journal / National Research Council, Institute of Laboratory Animal*
569 *Resources.* 2011;52(1):4-15.
570 23. Hsu HC, Li L, Zhang HG, Mountz JD. Genetic regulation of thymic involution. *Mechanisms*
571 *of ageing and development.* 2005;126(1):87-97.
572 24. Sloane LB, Stout JT, Vandenbergh DJ, Vogler GP, Gerhard GS, McClearn GE. Quantitative
573 trait loci analysis of tail tendon break time in mice of C57BL/6J and DBA/2J lineage. *The*
574 *journals of gerontology Series A, Biological sciences and medical sciences.*
575 2011;66(2):170-178.
576 25. Aberg KA, Chan RF, Xie L, Shabalin AA, van den Oord E. Methyl-CpG-Binding Domain
577 Sequencing: MBD-seq. *Methods in molecular biology.* 2018;1708:171-189.
578 26. Buniello A, MacArthur JAL, Cerezo M, et al. The NHGRI-EBI GWAS Catalog of published
579 genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic*
580 *acids research.* 2019;47(D1):D1005-D1012.
581 27. Weedon MN, Lettre G, Freathy RM, et al. A common variant of HMGA2 is associated
582 with adult and childhood height in the general population. *Nature genetics.*
583 2007;39(10):1245-1250.
584 28. Thorleifsson G, Walters GB, Gudbjartsson DF, et al. Genome-wide association yields new
585 sequence variants at seven loci that associate with measures of obesity. *Nature*
586 *genetics.* 2009;41(1):18-24.
587 29. Metrustry SJ, Edwards MH, Medland SE, et al. Variants close to NTRK2 gene are
588 associated with birth weight in female twins. *Twin Res Hum Genet.* 2014;17(4):254-261.
589 30. Ciccarone F, Tagliatesta S, Caiafa P, Zampieri M. DNA methylation dynamics in aging:
590 how far are we from understanding the mechanisms? *Mechanisms of ageing and*
591 *development.* 2018;174:3-17.
592 31. Rakyan VK, Down TA, Maslau S, et al. Human aging-associated DNA hypermethylation
593 occurs preferentially at bivalent chromatin domains. *Genome research.* 2010;20(4):434-
594 439.
595 32. Carter CJ. Convergence of genes implicated in Alzheimer's disease on the cerebral
596 cholesterol shuttle: APP, cholesterol, lipoproteins, and atherosclerosis. *Neurochem Int.*
597 2007;50(1):12-38.
598 33. Mozhui K, Pandey AK. Conserved effect of aging on DNA methylation and association
599 with EZH2 polycomb protein in mice and humans. *Mechanisms of ageing and*
600 *development.* 2017;162:27-37.

601 34. Berger H, Wodarz A, Borchers A. PTK7 Faces the Wnt in Development and Disease. *Front*
602 *Cell Dev Biol.* 2017;5:31.
603 35. Budovsky A, Fraifeld VE, Aronov S. Linking cell polarity, aging and rejuvenation.
604 *Biogerontology.* 2011;12(2):167-175.
605 36. Zhou Y, Simmons D, Lai D, Hambly BD, McLachlan CS. rs9939609 FTO genotype
606 associations with FTO methylation level influences body mass and telomere length in an
607 Australian rural population. *International journal of obesity.* 2017;41(9):1427-1433.
608 37. Lu AT, Xue L, Salfati EL, et al. GWAS of epigenetic aging rates in blood reveals a critical
609 role for TERT. *Nature communications.* 2018;9(1):387.
610 38. Teumer A, Qi Q, Nethander M, et al. Genomewide meta-analysis identifies loci
611 associated with IGF-I and IGFBP-3 levels with impact on age-related traits. *Aging cell.*
612 2016;15(5):811-824.
613 39. Bartke A. Healthy aging: is smaller better? - a mini-review. *Gerontology.* 2012;58(4):337-
614 343.
615 40. Barfield RT, Almli LM, Kilaru V, et al. Accounting for population stratification in DNA
616 methylation studies. *Genet Epidemiol.* 2014;38(3):231-241.
617 41. Lowe R, Barton C, Jenkins CA, et al. Ageing-associated DNA methylation dynamics are a
618 molecular readout of lifespan variation among mammalian species. *Genome biology.*
619 2018;19(1):22.
620 42. Gujar H, Liang JW, Wong NC, Mozhui K. Profiling DNA methylation differences between
621 inbred mouse strains on the Illumina Human Infinium MethylationEPIC microarray. *PloS*
622 *one.* 2018;13(3):e0193496.
623 43. Bates D, Maechler M, Bolker B, Walker S. lme4: Linear mixed-effects models using Eigen
624 and S4. . *R package version 11-7, http://CRANR-projectorg/package=lme4.*
625 44. Turner S. qqman: an R package for visualizing GWAS results using Q-Q and manhattan
626 plots. *Journal of Open Source Software.* 2018;3(25):731.
627 45. McLean CY, Bristor D, Hiller M, et al. GREAT improves functional interpretation of cis-
628 regulatory regions. *Nature biotechnology.* 2010;28(5):495-501.
629

Figure legends

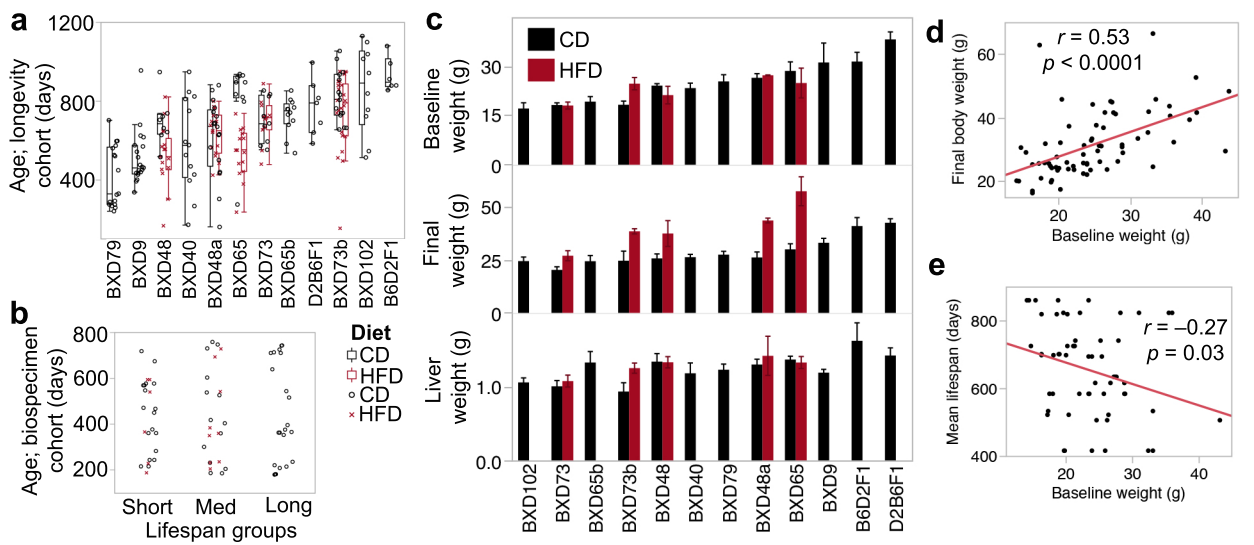
Fig. 1. Age distribution and body weight characteristics.
(a) Each point depicts a mouse that was allowed to age till mortality. There were a total of 225 mice for the 17 strain-by-diet groups on control diet (CD; black circles) or high fat diet (HFD; red crosses).
(b) Each point depicts a mouse used for tissue collection and methylome assay. Age distribution (y-axis)is uniform across the three lifespan groups.
(c) The bar plots show average body weight at young adulthood (baseline weight), and body and liver weight at final weighing (error bars are standard error).
The baseline body weight (x-axis) was **(d)** a significant predictor of individual body weight at older age (y-axis), and **(e)** was negatively correlated with mean lifespan (y-axis) for the strain-by-diet group.

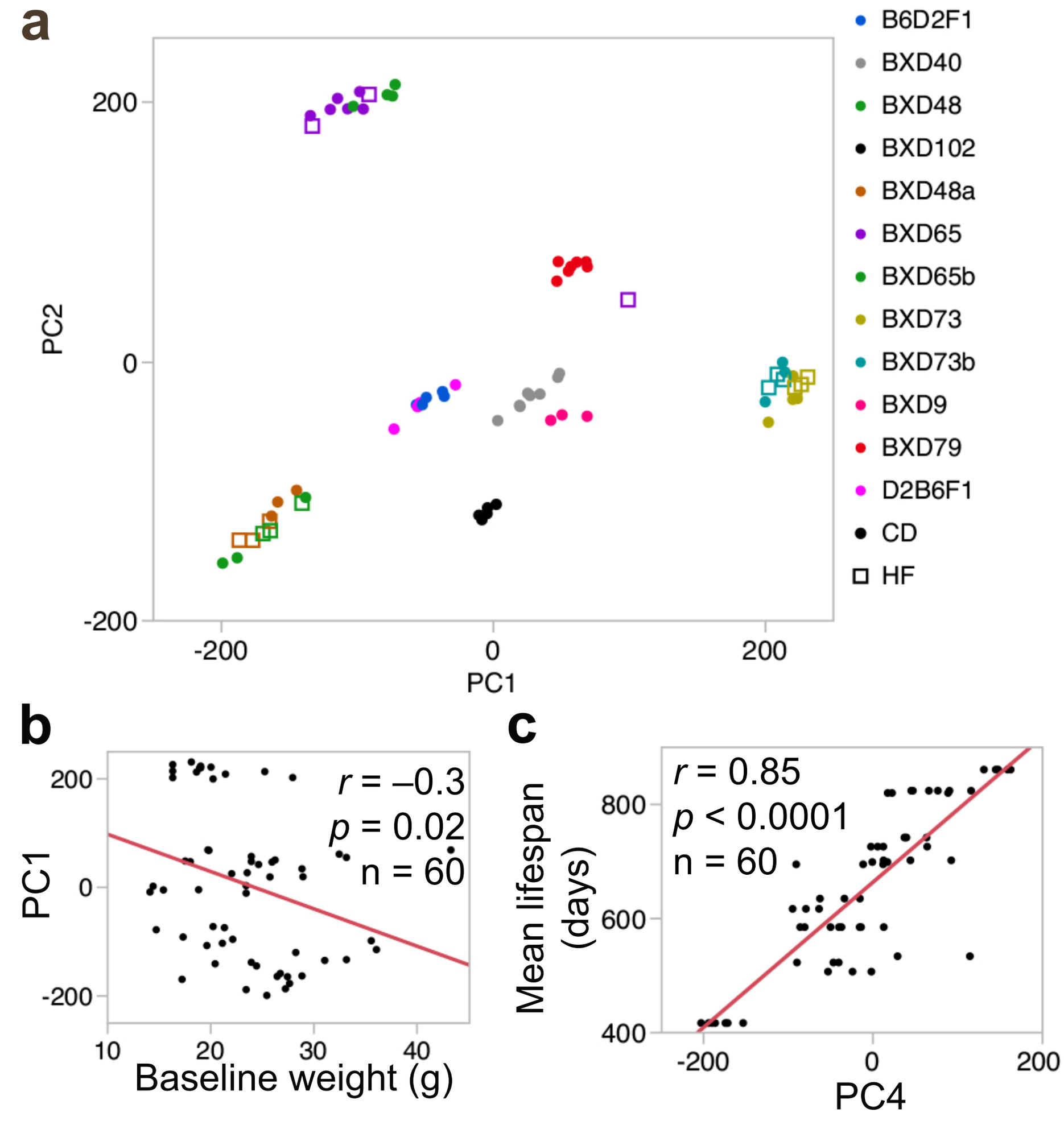
Fig. 2. Global features of the methylome
(a) Scatter plot between the top 2 principal components—PC1 (19% of variance) and PC2 (13% of variance)—shows a strong population structure with mice clustering by strain identity (color coded). For strains with animals from both standard chow (CD; solid circles) and high fat diet (HFD; squares), mice on HFD co-cluster with the CD mice.
(b) Body weight at young adulthood has a significant negative correlation with PC1.
(c) The methylome in turn may be predictive of lifespan, and PC4 is strongly correlated with lifespan.

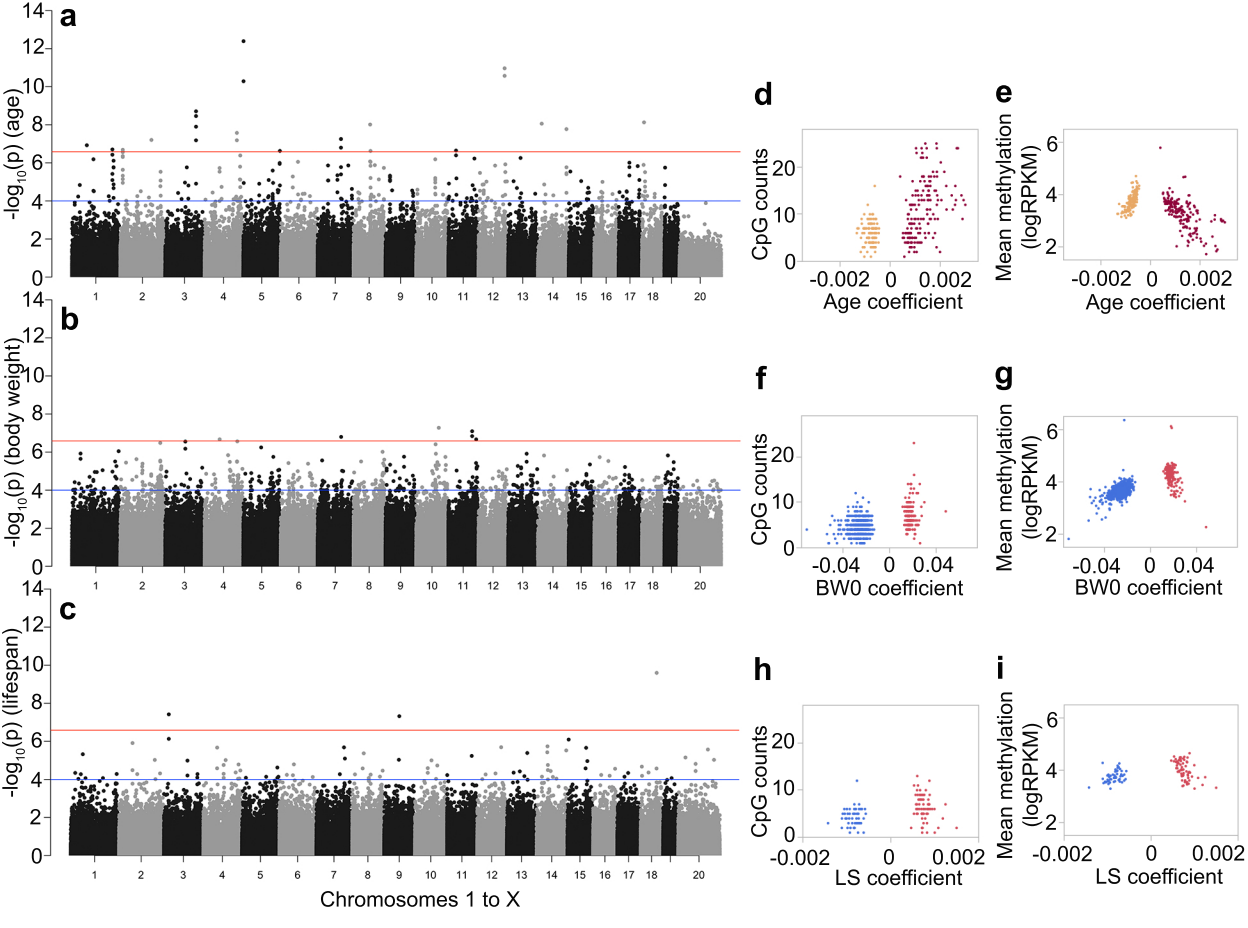
Fig. 3. Features of differentially methylated CpGs regions (DMRs)
Each point in the Manhattan plot represents the location of a CpG region (x-axis: autosomal chromosomes 1 to 19, and chromosome X as 20), and the association $-\log_{10}p$ (y-axis) for **(a)** age effect, **(b)** body weight at young adulthood (BW0), and **(c)** median lifespan (LS). The genome-wide significant threshold is set at $-\log_{10}(2.7e\text{-}7)$ (red line; 10% Bonferroni threshold for 368,300 tests) and the suggestive threshold at $-\log_{10}(1.0e\text{-}4)$ (blue line). For the age-DMRs, the regression coefficients for age (i.e., change in DNA methylation per unit change in age in days, $\log_{10}$ scale), and whether a site gained (positive coefficients, burgundy) or lost methylation (negative coefficients; sandy brown) was highly dependent on **(d)** the CpG density, and **(e)** mean methylation. For BW0-DMRs, the body weight associated coefficients also showed correlations with the CpG counts within the bin **(f)**, and sites that were negatively associated with BW0 (blue) had lower mean methylation than the sites that were positively associated with BW0 (red) **(g)**. For the LS-DMRs, whether a site was positively (red) or negatively (blue) associated with strain median lifespan was only modestly dependent on **(h)** the CpG counts, and not dependent on **(i)** the mean methylation levels.

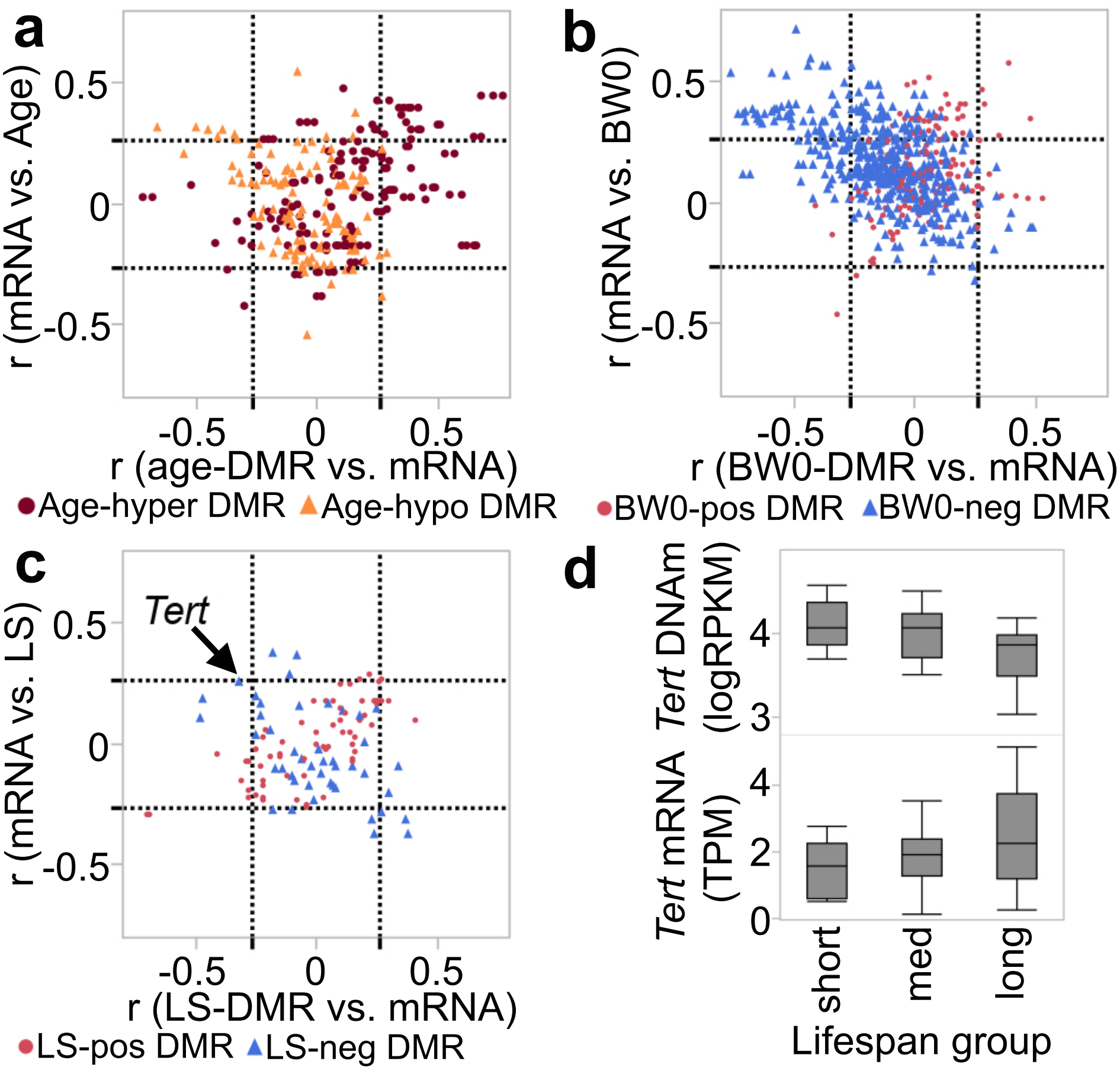
Fig. 4. Comparison between DMRs and expression of cognate genes
(a) The plot shows correlations between the age-DMRs and expression of corresponding transcripts (Pearson r on x-axis), and the correlation between the transcript and age of mice (y-axis). The dashed lines demarcate the nominally significant $|r| = 0.27$ threshold ($p \leq 0.05$). Most

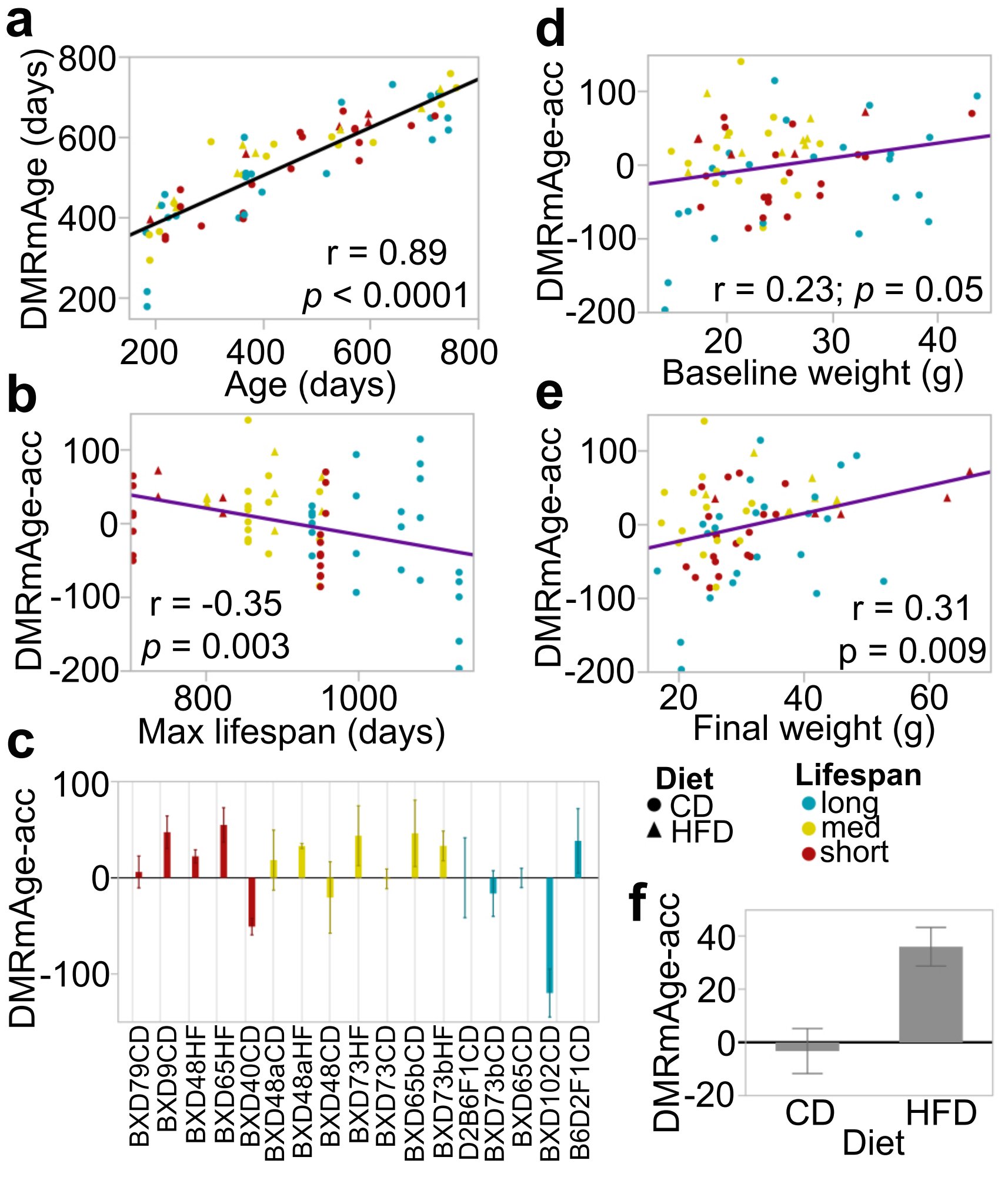
673 of the age-hypermethylated DMRs (burgundy circles) are positively correlated with mRNA, and
674 these mRNAs also tend to be positively correlated with age (top right square of graph). A few of
675 the age-hypomethylated-DMRs (sandy brown triangles) have negative correlation with gene
676 expression, and the corresponding transcripts are positively correlated with age (top left
677 square).
678 **(b)** Majority of the BW0-DMRs are negatively associated with BW0 (blue triangles), and most of
679 the DMR-mRNA pairs are located in the top-left square of the plot, i.e., gene expression is
680 negatively correlated with DNA methylation, and positively correlated with BW0. A few of the
681 BW0-DMRs that have positive associations with BW0 (red circles) also have positive
682 correlations with gene expression (top right square).
683 **(c)** The few LS-DMRs have modest correlations with gene expression. The DMR for *Tert* is
684 negatively correlated with transcript (arrow), and the transcript has modest correlation with
685 median lifespan at r = 0.26 (p = 0.07; n = 51 samples with matched MBD-seq and RNA-seq).
686 **(d)** Comparison of DNA methylation levels (top) in the three lifespan groups shows significantly
687 lower methylation in the long-lived group ($F_{2,66}$ = 7.67, p = 0.001). A similar comparison for gene
688 expression shows a significantly higher gene expression in the long-lived group ($F_{2,49}$ = 3,31, p =
689 0.04).
690
691 **Fig. 5. Age-DMR based measure of epigenetic aging**
692 **(a)** The epigenetic age of mice was estimated by taking the weighted averages of the 306 age-
693 DMRs. These age estimates, referred to as DMRmAge, has a strong positive correlation with the
694 chronological age of mice.
695 **(b)** The age acceleration residuals (DMRmAge-acc) derived from this clock has a significant
696 negative correlation with the maximum lifespan data for the 17 strain-by-diet groups.
697 **(c)** The bar plots show the mean DMRmAge-acc values for each strain-by-diet group (error bars
698 are standard error) with the graph ordered by increasing mean lifespan (x-axis).
699 The DMRmAge-acc is positively correlated with both body weight at young adulthood **(d)**, and
700 final body weight **(e)**.
701 **(f)** In the BXD strains with matched samples from both control diet (CD) and high fat diet (HFD),
702 the DMRmAge-acc is significantly higher in the HFD group compared to the CD group (–3.22 ±
703 36.99 in CD, 36.13 ± 27.32 in HFD, p = 0.002, n = 33).


a
Age; longevity cohort (days)
1200
800
400
BXD79
BXD9
BXD48
BXD40
BXD48a
BXD65
BXD73
BXD65b
D2B6F1
BXD73b
BXD102
B6D2F1
b
Age; biospecimen cohort (days)
800
600
400
200
Short
Med
Long
Lifespan groups
Diet
CD
HFD
CD
HFD
c
CD
HFD
Baseline weight (g)
30
15
0
Final weight (g)
50
25
0
Liver weight (g)
1.0
0.0
BXD102
BXD73
BXD65b
BXD73b
BXD48
BXD40
BXD79
BXD48a
BXD65
BXD9
B6D2F1
D2B6F1
d
r = 0.53
p < 0.0001
Final body weight (g)
60
40
20
20
30
40
Baseline weight (g)
e
r = –0.27
p = 0.03
Mean lifespan (days)
800
600
400
20
30
40
Baseline weight (g)


a
B6D2F1
BXD40
BXD48
BXD102
BXD48a
BXD65
BXD65b
BXD73
BXD73b
BXD9
BXD79
D2B6F1
CD
HF
PC2
PC1
200
0
-200
b
r = –0.3
p = 0.02
n = 60
PC1
Baseline weight (g)
10
20
30
40
c
r = 0.85
p < 0.0001
n = 60
Mean lifespan (days)
PC4
800
600
400


a
b
c
d
e
f
g
h
i
-log10(p) (age)
-log10(p) (body weight)
-log10(p) (lifespan)
Chromosomes 1 to X
CpG counts
Age coefficient
BW0 coefficient
LS coefficient
Mean methylation (logRPKM)


a
r (mRNA vs. Age)
0.5
0
-0.5
-0.5
0
0.5
r (age-DMR vs. mRNA)
Age-hyper DMR
Age-hypo DMR
b
r (mRNA vs. BW0)
0.5
0
-0.5
-0.5
0
0.5
r (BW0-DMR vs. mRNA)
BW0-pos DMR
BW0-neg DMR
c
r (mRNA vs. LS)
Tert
0.5
0
-0.5
-0.5
0
0.5
r (LS-DMR vs. mRNA)
LS-pos DMR
LS-neg DMR
d
Tert DNAm (logRPKM)
4
3
Tert mRNA (TPM)
4
2
0
short
med
long
Lifespan group


a
DMRmAge (days)
800
600
400
200
200
400
600
800
Age (days)
r = 0.89
p < 0.0001
b
DMRmAge-acc
100
0
-100
-200
800
1000
Max lifespan (days)
r = -0.35
p = 0.003
c
DMRmAge-acc
100
0
-100
BXD79CD
BXD9CD
BXD48HF
BXD65HF
BXD40CD
BXD48aCD
BXD48aHF
BXD48CD
BXD73HF
BXD73CD
BXD65bCD
BXD73bHF
D2B6F1CD
BXD73bCD
BXD65CD
BXD102CD
B6D2F1CD
d
DMRmAge-acc
100
0
-100
-200
20
30
40
Baseline weight (g)
r = 0.23; p = 0.05
e
DMRmAge-acc
100
0
-100
-200
20
40
60
Final weight (g)
r = 0.31
p = 0.009
Diet
CD
HFD
Lifespan
long
med
short
f
DMRmAge-acc
40
20
0
-20
CD
HFD
Diet