Genetic Loci and Metabolic States Associated With Murine Epigenetic Aging

Khyobeni Mozhui [1, 2*], Ake T. Lu[3], Caesar Z Li [3], Amin Haghani [4], Jose Vladimir Sandoval-Sierra [1], Yibo Wu[5,6], Robert W. Williams [2], Steve Horvath[3, 4]

[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 Human Genetics, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA

[4] Department of Biostatistics, Fielding School of Public Health, University of California Los Angeles, Los Angeles, CA, USA

[5]YCI Laboratory for Next-Generation Proteomics, RIKEN Center for Integrative Medical Sciences, Yokohama City, Kanagawa, Japan

[6]University of Geneva, Geneva, Switzerland

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

Abstract

Changes in DNA methylation (DNAm) are linked to aging. Here, we profile highly conserved CpGs in 339 predominantly female mice belonging to the BXD family for which we have deep longevity and genomic data. We use a 'pan-mammalian' microarray that provides a common platform for assaying the methylome across mammalian clades. We computed epigenetic clocks and tested associations with DNAm entropy, diet, weight, metabolic traits, and genetic variation. We describe the multifactorial variance of methylation at these CpGs, and show that high fat diet augments the age-associated changes. Entropy increases with age. The progression to disorder, particularly at CpGs that gain methylation over time, was predictive of genotype-dependent life expectancy. The longer-lived BXD strains had comparatively lower entropy at a given age. We identified two genetic loci that modulate rates of epigenetic age acceleration (EAA): one on chromosome (Chr) 11 that encompasses the *Erbb2/Her2* oncogenic region, and a second on Chr19 that contains a cytochrome P450 cluster. Both loci harbor genes associated with EAA in humans including *STXBP4*, *NKX2*-3, and *CUTC*. Transcriptome and proteome analyses revealed associations with oxidation-reduction, metabolic, and immune response pathways. Our results highlight concordant loci for EAA in humans and mice, and demonstrate a tight coupling between the metabolic state and epigenetic aging.

Keywords: epigenetic clock, lifespan, entropy, DNA methylation, genetic mapping, QTL, weight, diet

Introduction

Epigenetic clocks are widely used molecular biomarkers of aging.[1] These DNA methylation (DNAm) age predictors are based on the methylation levels of select CpGs that are distributed across the genome. Each CpG that is used in a clock model is assigned a specific weight, typically derived from supervised training algorithms,[2-4] and collectively, the methylation status across this ensemble of "clock CpGs" are used to estimate the epigenetic age (DNAmAge). This estimate tracks closely, but not perfectly, with an individual's chronological age. How much the DNAmAge deviates from the known chronological age can be a measure of the rate of biological aging. Denoted as "epigenetic age acceleration" (or EAA), a more accelerated clock (positive EAA) suggests an older biological age, and a decelerated clock (negative EAA) suggests a younger biological age. While DNAmAge predicts age, its age-adjusted counterpart, EAA, is associated with variation in health, fitness, exposure to stressors, body mass index (BMI), and even life expectancy.[5-9]

These DNAm clocks were initially reported for humans.[10-12] Since then, many different models of human DNAm clock have been develop, and this rapid expansion was made possible by reliable DNAm microarrays that provide a fixed CpG content—starting with the Illumina Infinium 27K to the current 850K EPIC array.[11,13-15] These clock variants differ in the subset of CpGs that go into the age estimation model. Some clock models are specific to cells or tissues, others are multi-tissue. Some clocks perform better at predicting chronological age, others better capture biological aging and predict health and life expectancy. [8,16-18] The performance of these clocks depend heavily on the training models, and the size and tissue types of the training set.[13]

This age biomarker has also been extended to model organisms, and this has opened up the possibility of directly testing the effects of different interventions such as calorie restriction, rapamycin, and genetic manipulation.[3,19-23] However, one point to note is that model organisms have not benefitted from a microarray platform comparable to that of the human methylation Infinium arrays. Most rodent studies have used enrichment-based DNAm sequencing, and this limits the transferability and reproducibility of clocks between datasets since the same CpGs are not always covered.[21] Moreover, these studies are usually performed in a single inbred strain (for mouse, the canonical C57BL/6), or at most, a few genetic backgrounds, and this makes it impossible to carry out genetic mapping studies that can complement the human genome-wide association studies (GWAS) of epigenetic aging.[24-28]

A new microarray was recently developed to profile CpGs that have high conservation in mammals. This pan-mammalian DNAm array (HorvathMammalMethylChip40) surveys over 37K CpGs and provides a unifying platform to study epigenetic aging in mammals.[29] This array has been used to build multi-tissue universal clocks and lifespan predictors that are applicable to a variety of mammalian species.[30,31] Here, we use this array to examine the dynamism and variability of the conserved CpGs in a genetically diverse cohort of mice belonging to the BXD family.[32,33]

The BXDs are one of the pre-eminent murine genetic reference panels used as the experimental paradigm of precision medicine.[34] They are a large family of recombinant inbred (RI) strains

made by crossing the C57BL/6J (B6) and DBA/2J (D2) parental strains. The family has been expanded to 150 fully sequenced progeny strains.[34,35] The individual members of the BXD family (e.g., BXD1, BXD27, BXD102), each represents a replicable isogenic cohort. The family segregates for a high level of genetic variation, and likewise, family members have high variation in their metabolic profiles, responses to diet, aging rates, and life expectancies.[32-34,36-38] The availability of deep sequence data, and unrivaled multi-omic and phenomic data make the BXDs a powerful tool with which to evaluate the causal linkage between genome, epigenome, and aging rates.

In our previous work, we used an enrichment-based sequencing to assay the methylome in a modest number of BXD mice, and reported rapid age-dependent methylation changes in mice on high fat diet, and in mice with higher body weight.[39] In the present work, we start by testing the performance of new pan-tissue and liver-specific epigenetic mouse clocks, and evaluate how these relate to metabolic states, genotype-dependent life expectancy, and methylome entropy. We also apply a multi-factor analysis of site-specific CpG methylation to understand association among four key variables—chronological age, diet, weight, and lifespan—and the liver methylome. We perform quantitative trait locus (QTL) mapping, along with multi-omic gene expression analyses, and identify upstream gene loci that modulate the DNAm clocks in mice.

Our results are consistent with a faster clock for cases on HFD, and with higher body weight. This may be partly because exposure to HFD augmented the age-dependent gains in methylation at specific CpGs. We also observed that BXD genotypes with longer life expectancy tend to have lower methylation at CpGs that undergo age-dependent methylation gains, and the entropy computed from this set of CpGs have a significant inverse correlation with strain lifespan. QTL mapping uncovered loci on chromosomes (Chrs) 11 and 19 that are associated with EAA. A strong candidate gene in the chromosome (Chr) 11 interval (referred to as *Eaa11*) is *Stxbp4*, a gene that has been consistently associated with EAA by human genome-wide association studies (GWAS).[24,26,27] The Chr19 QTL (*Eaa19*) also harbors strong contenders including *Cyp26a1*, *Myof*, *Cutc*, and *Nkx2–3*, and the conserved genes in humans have been associated with longevity and EAA.[27,40,41] We performed gene expression analyses using transcriptomic and proteomic data to clarify the molecular pathways associated with epigenetic aging, and this highlighted metabolic networks, and also apolipoproteins (including APOE) as strong expression correlates.

Results

Description of samples

Liver DNAm data was from 321 female and 18 male belonging to 45 members of the BXD family, including both parental strains and F1 hybrids. Age ranged from 5.6 to 33.4 months. Mice were all weaned onto a normal chow (control diet; CD) and a balanced subset of cases were then randomly assigned to HFD (see Roy et al. for details[33]). Tissues were collected at approximately six months intervals (see Williams et al.[32]). Individual-level data are in **Data S1**.

DNAm clocks, entropy, and chronological age prediction

124 We built three different mouse clocks, and each was developed as a pair depending on whether
125 the training set used all tissues (pan-tissue) or a specific tissue (in this case, liver). These are: (1)
126 a general DNAm clock (referred to simply as DNAmAge): clock trained without pre-selecting for
127 any specific CpGs; (2) developmental clock (dev.DNAmAge): built from CpGs that change during
128 development (defined as the period from prenatal to 1.6 months); and (3) interventional clock
129 (int.DNAmAge): built from CpGs that change in response to aging related interventions (calorie
130 restriction and growth hormone receptor knockout). The clocks we report here were trained in
131 a larger mouse dataset that excluded the BXDs and are therefore unbiased to the
132 characteristics of the BXD Family.[30,31,42] The specific clock CpGs and coefficients for DNAmAge
133 computation are in **Data S2**. All the mouse clocks performed well in age estimation and had an
134 average r of 0.89 with chronological age. However, the interventional clocks had higher
135 deviation from chronological age and higher median predictive error (**Table 1; Fig 1a**). The age-
136 adjusted EAA derived from these clocks showed wide individual variation (**Fig 1b**).

137 We next estimated the methylome-wide entropy as a measure of randomness and information
138 loss. This was computed from 27966 probes that provide high-quality data and have been
139 validated to perform well in mice.[29] Consistent with previous reports,[10,43] this property
140 increased with chronological age, and age accounted for about 6% (in CD) to 28% (in HFD) of
141 the variance in entropy (**Fig 1c**). As direct correlates of chronological age, all the DNAmAge
142 estimates also had significant positive correlations with entropy (**Table 1**). We hypothesized
143 that higher entropy levels will be associated with higher EAA, and based on this bivariate
144 comparison, most of the EAA showed a significant positive correlation with entropy (**Data S3;**
145 **Fig 1d**).

146 **Table 1. Chronological age prediction and correlation with methylome-wide entropy**

Clock type	DNAmAge name	Tissue	r with age (n=339) [1]	Age prediction median error	r with entropy (n=339)[1, 2]
Standard clocks	DNAmAge	pan	0.89	0.12	0.43
		liver	0.92	0.10	0.40
Developmental clocks	dev.DNAmAge	pan	0.87	0.14	0.39
		liver	0.91	0.12	0.37
Interventional clocks	int.DNAmAge	pan	0.85	0.17	0.29
		liver	0.86	0.15	0.47

147 [1]$p < .0001$; [2]$p < .0001$ Methylome-wide entropy calculated from ~28K CpGs

148 How the epigenetic readouts relate to diet, sex, and metabolic traits

149 ***Diet.*** HFD was associated with higher EAA for four of the clocks (**Table 2**). For instance, the
150 liver-specific interventional clock diverged between the diets (**Fig 1a**), and CD mice had an
151 average of –0.04 years of age deceleration, and HFD mice had an average of +0.11 years of age
152 acceleration **(Table 2**). The two clocks that were not affected by diet were the liver general and
153 developmental clocks. Methylome-wide entropy was not different between the diets.

154 ***Body weight.*** Body weight was first measured when mice were at an average age of 4.5 ± 2.7
155 months. We refer to this initial weight as baseline body weight (BW0). For mice on HFD, this

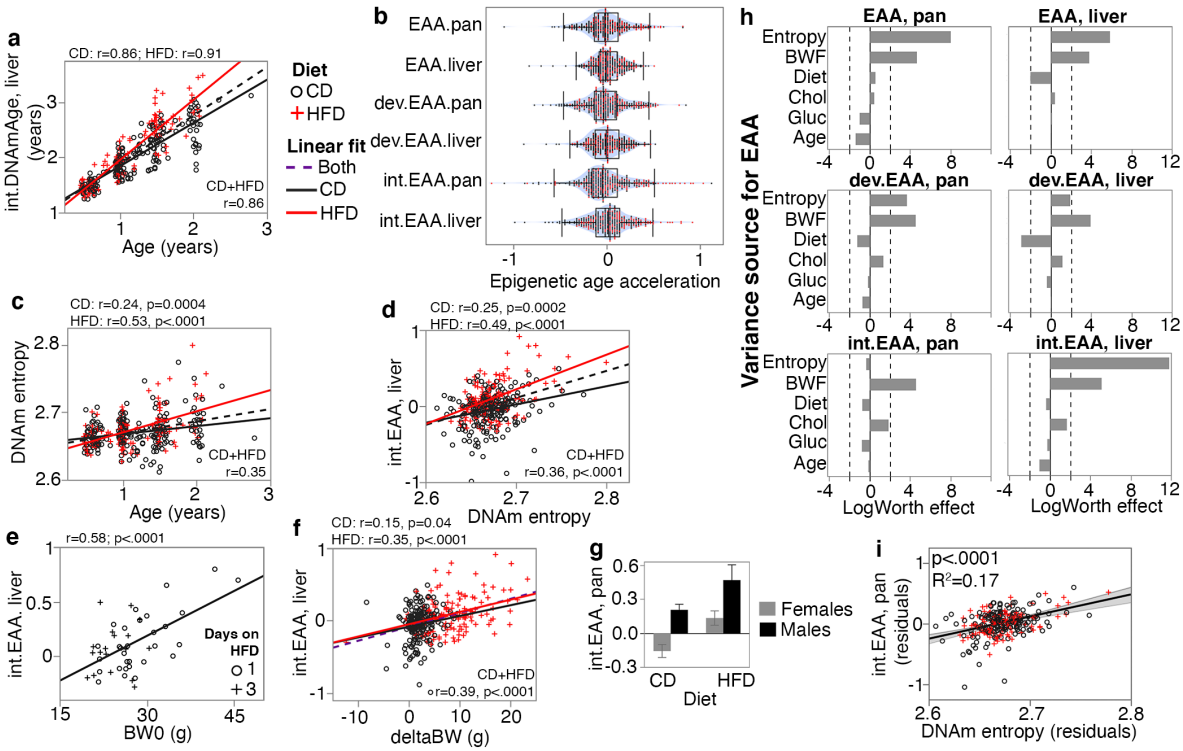


Fig 1. Correlates and modifiers of epigenetic clocks and methylome-wide entropy
(a) Correlation between chronological age and predicted age (shown for the liver intervention clock or int.DNAmAge). Black circles are control diet (CD, n = 210); red crosses are high fat diet mice (HFD, n = 128) **(b)** Violin plots of age-adjusted epigenetic age acceleration (EAA) ("int" stands for interventional, "dev" stands for developmental). **(c)** Shannon entropy, calculated from the full set of high quality CpGs, increases with age. **(d)** Methylome entropy has a direct correlation with EAA (shown for the liver int.EAA). **(e)** For 48 mice, initial body weight (BW0) was measured 1 or 3 days after introduction to HFD, and these showed significant correlation with EAA. **(f)** Weight was first measured at mean age of 4.5 ± 2.7 months (BW0), and then at 6.3 ± 2.8 months (BW1). Weight gains during this interval (deltaBW = BW1 – BW0) is a direct correlate of EAA. **(g)** For BXD genotypes with males and females samples, males have higher age acceleration. Bars represent mean ± standard error; 40 females (26 CD, 14 HFD) and 18 males (10 CD, 8 HFD). **(h)** Relative effects of different predictor variables on EAA shown as logworth scores ($-\log_{10}p$). The dashed lines correspond to p = 0.01. Positive logworth values indicate positive regression estimates (for diet, positive means higher in high fat diet compared to control diet). BWF is final weight; Chol is serum total cholesterol; Gluc is fasted glucose levels. **(g)** The residual plot display association between methylome-wide entropy and the pan-tissue int.EAA after adjustment for diet, age, weight, glucose, cholesterol, and batch.

156

157 **Table 2. Association with diet and weight, and heritability of the epigenetic readouts**

Type	EAA	Diet	Mean (SD)	Diet (p)	r BW0 [a]	p BW0	r BWF [a]	p BWF	H^2	Strain r [b]
Mouse clocks	EAA, pan	CD	-0.05 ± 0.21	<.0001	0.19	0.006	0.29	<.0001	0.49	0.54
		HFD	0.07 ± 0.21		0.21	0.01	0.42	<.0001	0.50	
	EAA, liver	CD	0 ± 0.17	ns	0.09	ns	0.20	0.003	0.40	0.73
		HFD	0.03 ± 0.14		0.22	0.01	0.49	<.0001	0.52	
	dev.EAA, pan	CD	-0.04 ± 0.23	0.004	0.09	ns	0.22	0.001	0.53	0.76
		HFD	0.03 ± 0.22		0.27	0.002	0.45	<.0001	0.61	
	dev.EAA, liver	CD	0 ± 0.2	ns	0.19	0.002	0.29	<.0001	0.46	0.78
		HFD	0 ± 0.16		0.29	0.0007	0.47	<.0001	0.60	
	int.EAA, pan	CD	-0.05 ± 0.25	0.0003	0.03	ns	0.21	0.002	0.27	0.66
		HFD	0.06 ± 0.33		0.22	0.01	0.46	<.0001	0.45	
	int.EAA, liver	CD	-0.04 ± 0.22	<.0001	0.05	ns	0.18	0.01	0.59	0.80
		HFD	0.11 ± 0.25		0.27	0.002	0.58	<.0001	0.54	
Entropy	-	CD	2.67 ± 0.02	ns	0.09	ns	0.05	ns	0.31	0.24 (ns)
		HFD	2.67 ± 0.02		0.15	0.09	0.15	0.09	0.32	

158 [a] BW0 is body weight at about 4.5 months of age (n = 339; 210 CD and 129 HFD); BWF is final weight at tissue collection (1 HFD case missing data; n = 338; 210
159 CD and 128 HFD)
160 [b] Pearson correlation between strain means for n = 29 BXD genotypes kept on CD and HFD

161 was usually before introduction to the diet, except for 48 cases that were first weighed 1 or 3
162 days after HFD (**Data S1**). In the CD group, only the EAA from the pan-tissue standard and liver
163 developmental clocks showed modest but significant positive correlations with BW0 (**Table 2**).
164 In the HFD group, the positive correlation with BW0 was more robust and consistent across all
165 the clocks, and this may have been due to the inclusion of the 48 cases that had been on HFD
166 for 1 or 3 days. Taking only these 48 cases, we found that higher weight even after 1 day of HFD
167 had an age-accelerating effect on all the clocks (**Data S3**), and this was particularly strong for
168 the interventional clocks ($r = 0.45$, $p = 0.001$ for the pan-tissue int.EAA; $r = 0.58$, $p < 0.0001$ for
169 the liver int.EAA) (**Fig 1e**). Second weight was measured 7.4 ± 5.2 weeks after BW0 (mean age
170 6.3 ± 2.8 months). We refer to this as BW1 and we estimated the weight change as deltaBW =
171 BW1 – BW0. DeltaBW was a positive correlate of EAA on both diets, albeit more pronounce in
172 the HFD group (**Fig 1f**; **Data S3**). The final body weight (BWF) was measured at the time of
173 tissue harvest, and EAA from all the mouse clocks were significant correlates of BWF on both
174 diets **(Table 2**). In contrast, entropy did not show an association with either BW0 or BWF. We
175 do note that when stratified by diet, the entropy level had a slight positive correlation with
176 BW1 in the HFD group ($r = 0.23$, $p = 0.008$), but not in the CD group (**Data S3)**.

177 ***Sex.*** Four BXD genotypes (B6D2F1, D2B6F1, BXD102, B6) had cases from both males and
178 females. We used these to test for sex effects. All the clocks showed significant age acceleration
179 in male mice, and this effect was particularly strong for the both dev.EAA, and the pan-tissue
180 int.EAA (**Fig 1g; Data S3**). This effect was independent of the higher BWF of males, and the
181 higher age-acceleration in males was detected after adjusting for BWF (**Table S1**). There was no
182 significant difference in entropy between the sexes.

183 ***Metabolic measures.*** 278 cases with DNAm data also had fasted serum glucose and total
184 cholesterol,[32,33] and we examined whether these metabolic traits were associated with either
185 the EAA measures or methylome entropy. Since these are highly dependent on diet, weight,
186 and age, we applied a multivariable model to jointly examine how the different metabolic
187 variables (cholesterol and glucose, as well as diet and weight) and entropy relate to EAA after
188 adjusting for age. To test the robustness of associations, we also include the methylation array
189 batch as another covariate (**Data S1** has batch information; **Data S4** has the full statistics). **Fig**
190 **1h** shows the relative strengths and directions of associations between these variables and the
191 EAA traits. Except for the pan-tissue interventional clock, entropy had a strong positive
192 association with EAA. For example, a plot of residuals between entropy and the liver int.EAA
193 indicates that after adjusting for all the other covariates, the methylome-wide entropy explains
194 17% of the variance in int.EAA (**Fig 1i**). Since diet strongly influences BWF, the inclusion of BWF
195 in the regression diminished the effect of diet. For the two clocks that were not influenced by
196 diet (the liver EAA and liver dev.EAA), adjusting for the effect of BWF resulted in an inverse
197 association with diet (i.e., the residual EAA values after accounting for BWF were slightly lower
198 in the HFD group). Fasted glucose did not have a significant effect on EAA. Cholesterol had a
199 positive association with the interventional clocks but the effects were modest (residual R^2 =
200 0.02 and $p = 0.02$ for cholesterol and the pan-tissue int.EAA) (**Data S4**).

201 We also performed a similar analysis with BW0 instead of BWF (**Fig S1**), and here, HFD
202 remained as an accelerator of the pan-tissue EAA and liver int.EAA. Cholesterol also became a
203 significant positive correlate of EAA for the interventional clocks (**Fig S1**). This would suggest

204 that the effect of diet on EAA is mostly mediated by its impact on physiological and metabolic
205 traits, and BWF becomes a prominent predictor of EAA.

206 To summarize, our results indicate that the degree of disorder in the methylome increases with
207 age, and may partly contribute to the epigenetic clocks as higher entropy is associated with
208 higher EAA. The EAA traits were also associated with biological variables (i.e., body weight, diet,
209 and sex). Of these, BWF was the strongest modulator of EAA.

210 How the epigenetic readouts relate to strain longevity

211 We next obtained longevity data from a parallel cohort of female BXD mice that were allowed
212 to age on CD or HFD.[33] Since the strain lifespan was determined from female BXDs, we
213 restricted this to only the female cases. For strains with natural death data from n ≥ 5, we
214 computed the minimum (minLS), 25th quartile (25Q-LS), mean, median lifespan, 75th quartile
215 (75Q-LS), and maximum lifespan (maxLS) (**Data S1**). Specifically, we postulated an accelerated
216 clock for strains with shorter lifespan (i.e., inverse correlation). Overall, the EAA measures
217 showed the expected inverse correlation trend with the lifespan statistics. However, these
218 correlations were weak. The correlations were significant only for the pan-tissue general clock
219 (**Fig S2a**) and the liver intervention clock, with explained variance in lifespan of only ~3% (**Table**
220 **S2; Fig S2b, S2c**). When separated by diet, these correlations became weaker indicating that
221 while we see the expected inverse relationship, the EAA is only weakly predictive of strain
222 longevity. Entropy estimated from the full set of CpGs was unrelated to strain longevity.

223 Multifactor variance of the conserved CpGs

224 Both the entropy and clock readouts capture the overall variation across multiple CpGs, and to
225 gain insights into the underlying variance patterns, we performed a multivariable epigenome-
226 wide association study (EWAS). For this, we applied a site-by-site regression on the 27966
227 validated CpGs,[29] and tested for association with age, BWF, diet, and genotype-dependent
228 strain median lifespan (full set of probes, annotations, and EWAS results in **Data S5**).

229 Age was clearly the most influential variable, and this is apparent from the volcano plots (**Fig**
230 **2a–d**). We used a cutoff of Bonferroni p ≤ 0.05 to define differentially methylated CpGs (DMCs),
231 and 6553 CpGs were associated with age (referred to as age-DMCs), 733 with weight (weight-
232 DMCs), 321 with diet (diet-DMCs), and 236 with genotype-dependent lifespan (LS-DMCs). We
233 note extensive overlap among the lists of DMCs that shows that variation at these CpGs are
234 multifactorial in nature (**1e**). Majority of the age-DMCs (77%) gained methylation (or age-gain),
235 and consistent with previous observations, age-gain CpGs tended to be in regions with low
236 methylation, whereas age-DMCs that declined in methylation (age-loss) were in regions with
237 high methylation (**Fig 2f**).[39,43,44] By overlaying the volcano plots with the age-gain and age-loss
238 information, we see distinct patterns in how these age-DMCs vary with weight (**Fig 2b**), diet (**Fig**
239 **2c**), and genotype lifespan (**Fig 2d**). While the majority of CpGs, including several age-loss CpGs,
240 had negative regression estimates for weight (i.e., decrease in DNAm with unit increase in
241 weight), HFD was associated with higher methylation levels (positive regression estimates)
242 including at several age-DMCs (**Fig 2c**). This pattern of inverse correlation with weight but
243 heightened methylation due to HFD is illustrated by the CpG in the 3'UTR of *Mettl23*
244 (cg10587537) (**Fig 2g**). Taking the 6553 age-DMCs, a comparison of the regression estimates for

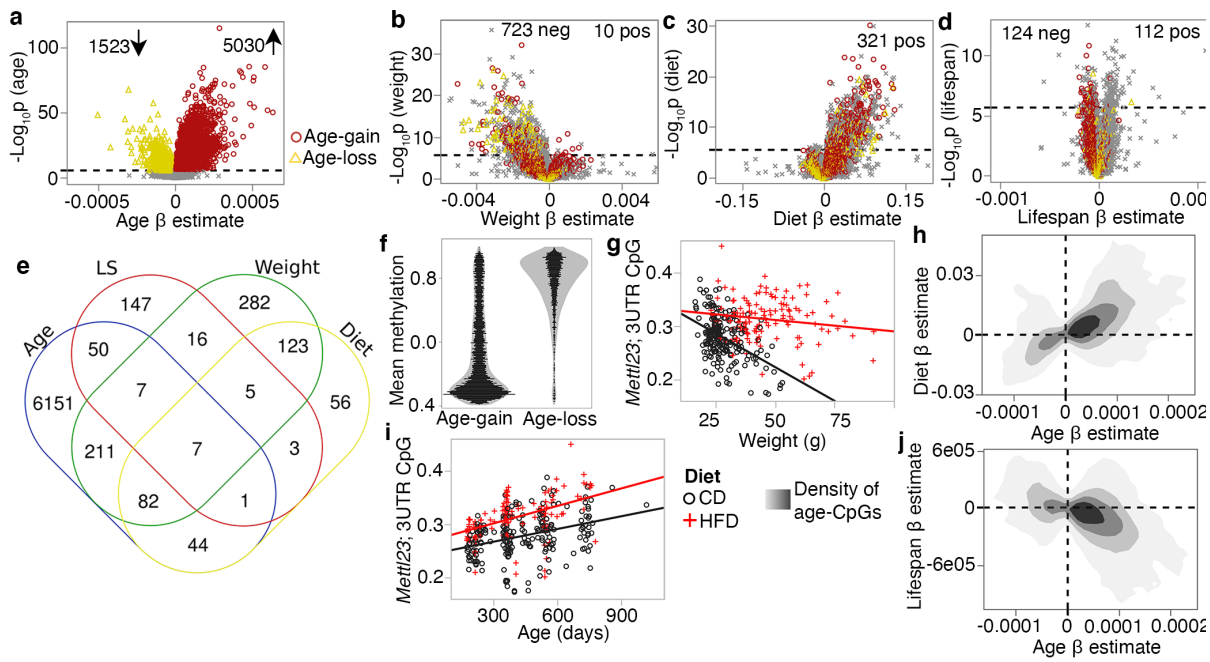


Fig 2. Multivariable analysis of site-specific methylation

(a) Volcano plot comparing regression estimates (change in methylation beta-value per day of age) versus the statistical significance for age effect. Dashed line denotes the Bonferroni p = 0.05 for ~28K tests). Similar volcano plots for predictor variables: **(b)** final body weight (regression estimates are change per gram of weight), **(c)** diet (change in high fat compared to control diet), and **(d)** strain median lifespan (per day increase in median longevity). CpGs that were significantly associated with age are denoted by colored markers (red circles: age-gain; yellow triangles: age-loss). **(e)** Overlap among the lists of differentially methylated CpGs. **(f)** Each dot represents the mean methylation beta-values for the 5030 age-gain, and 1523 age-loss CpGs. **(g)** Correlation between body weight and methylation beta-values for the CpG (cg10587537) located in the 3'UTR of *Mettl23*. Mice on high fat diet (HFD) have higher methylation than mice on control diet (CD), but the inverse correlation with weight is consistent for both groups (r = –0.45, p < .0001 for CD; r = –0.15, p = 0.08 for HFD). **(h)** Contour density plot for the 6553 CpGs that are significantly associated with age (age-DMCs). This relates the pattern of change with age (x-axis) with change on HFD (y-axis). CpGs that gain methylation with age are also increased in methylation by HFD. **(i)** Correlation between age and methylation at the *Mettl23* 3'UTR CpG (r = 0.35 for CD; r = 0.46). **(j)** For the 6553 age-DMCs, the contour density plot relates the pattern of change with age (x-axis) vs. change with median longevity (y-axis). CpGs that gain methylation with age have lower methylation with higher lifespan.

age (i.e., the change in methylation per day of aging) versus diet (difference in HFD relative to CD) shows that the age-gains were augmented in methylation by HFD (**Fig 2h**), and again, this is illustrated by the CpG in the *Mettl23* 3'UTR (**Fig 2i**). For the LS-DMCs, sites that had negative regression estimates for lifespan (i.e., lower DNAm per day increase in strain median longevity) had higher proportion of age-gain CpGs (**Fig 2d**). A comparison between the regression

estimates for age versus the regression estimates for lifespan shows that CpGs that gain methylation with age tended to have lower methylation in strains with longer lifespan (**Fig 2j**).

As in Sziráki et al.,[43] we divided the CpGs by age effect: age-gain, age-loss, and those that do not change strongly with age (age-ns; i.e., the remaining 21413 CpGs that were not classified as

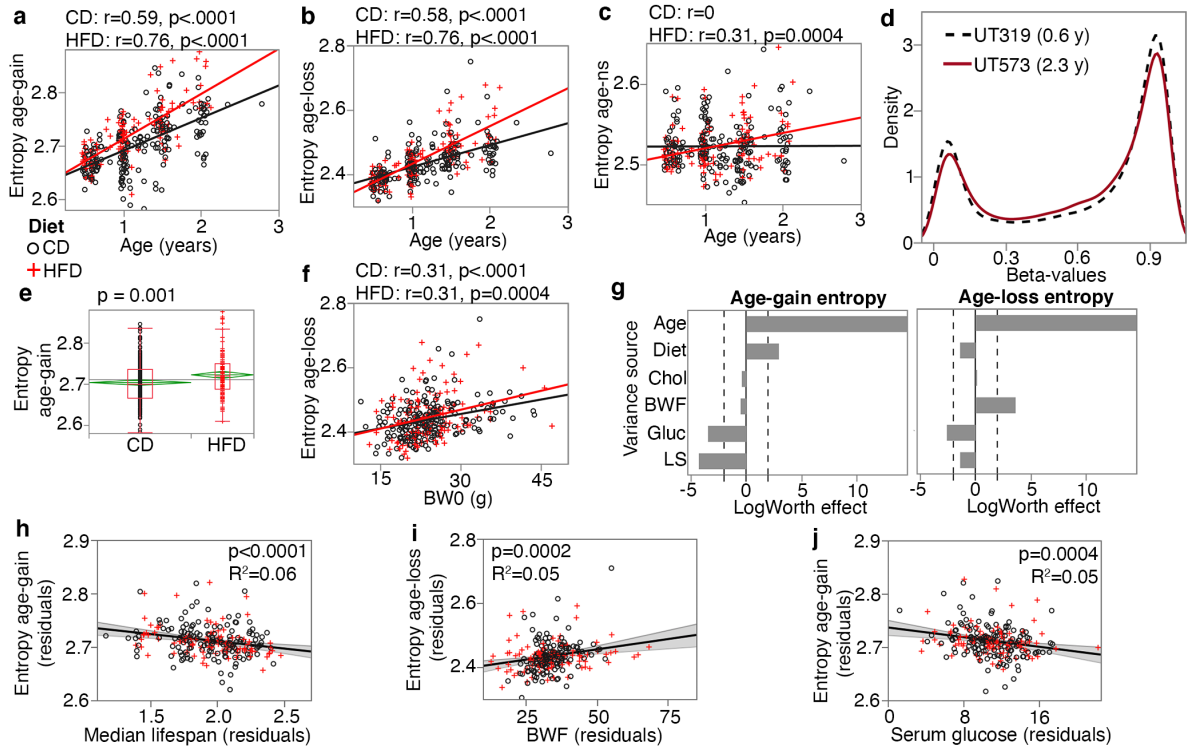


Fig 3. Entropy at age-associated CpGs

Entropy values were calculated for the 5020 age-gain and 1523 age-loss CpGs separately. For both control diet (CD) and high fat diet (HFD), there is significant increase in entropy with age at the **(a)** age-gain and **(b)** age-loss CpGs. **(c)** The HFD mice also showed a slight increase in entropy at CpGs that were not strongly associated with age (age-ns). **(d)** The methylome-wide distribution of beta-values in a young adult mouse (0.6 year old; black dashed line), and an older mouse (2.3 year old; red line); both CD mice. The young mouse has higher peaks at the hypo-methylated (closer to 0.1) and hyper-methylated (around 0.9) beta-values compared to the older mouse. **(e)** The HFD group has higher entropy at the age-gain CpGs compared to the CD group. **(f)** Entropy at age-loss CpGs is higher with higher baseline weight (BW0). **(g)** Relative effects of predictor variables on entropy shown as logworth scores ($-\log_{10}p$). The dashed lines correspond to p = 0.01. Positive values indicate positive regression estimates (for diet, positive value means higher in HFD). BWF is final weight; Chol is serum total cholesterol; Gluc is fasted glucose levels; LS is the strain median lifespan. **(h)** The residual plot (adjusted for age, diet, BWF, glucose, cholesterol, and batch) shows the inverse association between entropy at age-gain sites, and lifespan. Similar residual plots show the association between **(i)** BWF and age-loss entropy, and **(j)** between fasted serum glucose and age-gain entropy.

254 age-DMCs). For these conserved CpGs, both sets of age-DMCs had significant increases in
255 entropy with age regardless of diet (**Fig 3a**, **3b**), and even the age-ns showed a modest entropy
256 gain with age in the HFD group (**Fig 3c**). The reason for this increase in disorder becomes
257 evident when we compare the density plots using the full set of CpGs for one of the younger
258 mice (UT319; 0.56 years old) and one of the older mice (UT573; 2.3 years) (**Fig 3d**). Concordant
259 with previous reports,[43,45] the older sample showed a subtle flattening of the bimodal peaks
260 towards a slightly more hemi-methylated state. The entropy of the age-gain CpGs was modestly
261 but significantly higher in the HFD group ($p = 0.001$; **Fig 3e**). Entropy of the age-loss and age-ns
262 CpGs were not different between the diets. Body weight on the other hand, was associated
263 specifically with the entropy score of the age-loss CpGs, and both higher BW0 (**Fig 3f**) and BWF
264 predicted higher entropy for age-loss CpGs.

265 We applied a multivariable regression to compare the relative effects of age, diet, BWF,
266 glucose, cholesterol, and strain median lifespan (**Fig 3g**; full statistics in **Data S6**). Entropy of
267 age-gain CpGs was increased by HFD but was not associated with BWF. Strain median lifespan
268 showed a significant inverse correlation with the entropy of age-gain CpGs with an explained
269 variance of 6% (**Fig 3h**). Entropy of the age-loss CpGs had a significant positive correlation with
270 BWF (**Fig 3i**), but was not associated with diet, and also had a modestly significant inverse
271 correlation with median lifespan. Cholesterol was unrelated to the entropy values. Glucose on
272 the other hand, showed a significant inverse association with entropy of both the age-gain (**Fig**
273 **3j**) and age-loss CpGs, and this suggests slightly lower entropy with higher fasted glucose.

274 Taken together, our results show that the conserved CpGs are influenced by multiple
275 predictors. HFD augmented the age-dependent changes with a prominent effect on age-gain
276 CpGs. Body weight showed a strong association with the age-loss CpGs. Additionally, strains
277 with longer life expectancy tended to have lower methylation levels at age-gain CpGs with an
278 overall lower entropy state at these CpGs that suggests a more "youthful" methylome for
279 longer lived genotypes.

280 Functional and genomic context of DMCs

281 To uncover the potential biological pathways represented by the DMCs, we performed genomic
282 regions enrichment analyses for the CpGs.[46] The age-gain CpGs were highly enriched in
283 transcription factors, regulators of development and growth, menarche and menstrual phases,
284 energy metabolism, and transcription factor networks such as HNF1 and HNF3B pathways (**Data**
285 **S7**). The age-loss CpGs had somewhat modest enrichment, and represented cell adhesion and
286 cytoskeletal processes, endothelial cell proliferation, and p38 signaling (**Data S7**). The BW-DMCs
287 were enriched in actin and protein metabolism, and WNT, and platelet-derived growth factor
288 (PDGF) and ErbB signaling. Similarly, the diet-DMCs were highly enriched in PDGF, epidermal
289 growth factor (EGFR) and ErbB signaling, as well as the mTOR signaling pathway, and regulation
290 of energy homeostasis (**Data S7**). Seeming to converge on common pathways, the LS-CpGs that
291 were negatively correlated with lifespan had modest enrichment in cell signaling pathways such
292 as EGFR, PDGF, and ErbB signaling. The LS-CpGs with positive correlation with lifespan were
293 highly enriched in lipid metabolic genes, and also included pathways related to chromosome
294 maintenance and telomere expansion (**Data S7**).

295 We next examined the genomic annotations and chromatin states of the DMCs (**Data S8**).
296 Consistent with previous reports,[39,43] age-gain CpGs were enriched in promoter and 5'UTR
297 CpGs, but depleted in 3'UTR, exon, and intergenic CpGs (**Data S8; Fig S3a**). Diet- and weight-
298 DMCs were depleted in promoter regions, and enriched in exons and 3'UTR, and along with the
299 age-loss CpGs, enriched in introns. For chromatin states, we annotated the CpG regions using
300 the 15-states chromatin data for neonatal (P0) mouse liver (**Data S8** has annotations for each
301 site).[47,48] Also included were regions labelled as No Reproducible State or NRS; i.e., regions that
302 were not replicated.[48] Compared to the array content as background, the age-gain CpGs were
303 selectively enrich in polycomb associated heterochromatin (Hc-P) and bivalent promoters (Pr-
304 Bi), chromatin states that were highly depleted among the other DMCs (**Fig S3b**; **Data S8**). In
305 contrast, strong and permissive transcription sites (Tr-S and Tr-P, respectively) were depleted
306 among the age-gain CpGs, and enriched among the BW- and diet-DMCs (**Fig S3b**). Age-loss CpGs
307 were enriched in Tr-P and Tr-I (transcription initiation). Distal enhancers (strong distal or En-Sd,
308 and poised distal or En-Pd) were also highly enriched among the BW- and diet-DMCs, and also
309 showed some enrichment among the age-DMCs (**Fig S3b**).

310 For an overview of the general methylation and variance patterns by chromatin annotations,
311 we used the full set of 27966 CpGs and computed the average methylation beta-values, and
312 average regression coefficients (i.e, change in beta-value per unit change in the respective
313 predictor variable, or contrast between diets). As expected, promoter CpGs and Hc-P were sites
314 with the lowest methylation. Hc-H, Tr-S, and Tr-P had higher methylation, and many of the
315 enhancer sites were in the hemi-methylated zone (**Fig S3c, d, e**). For age effect, mean
316 regression estimates had a significant inverse linear fit with mean methylation ($r = -0.63$, $p =$
317 0.009; **Fig S3c**) and this is consistent with the greater age-loss at hypermethylated CpGs, and
318 greater age-gains at hypomethylated CpGs (**Fig 2f**). The effects of diet and weight were not
319 linearly related to the mean methylation of the chromatin states. Instead, both showed a U-
320 shaped fit with a significant negative quadratic effect for diet ($R^2 = 0.69$, $p = 0.0005$, quadratic
321 estimate = –0.05; **Fig S3d**), and a positive quadratic effect for weight ($R^2 = 0.50$, $p = 0.01$,
322 quadratic estimate = 0.001; **Fig S3e**). Methylation variation as a function on strain longevity did
323 not relate to mean methylation with either a linear or polynomial fit, and indicates that
324 variance due to background genotype is less dependent on the chromatin and mean
325 methylation status. While this is a very low-resolution and broad view of methylation levels and
326 methylation variation, the observations show that while aging results in erosion of the hypo-
327 and hypermethylated peaks, diet and body weight appear to have generally stronger
328 associations with hemi-methylated sites.

329 Genetic analysis of epigenetic age acceleration

330 The EAA traits had moderate heritability at an averaged H^2 of 0.50 (**Table 2**).[34] Another way to
331 gauge level of genetic correlation is to compare between members of strains maintained on
332 different diets. All the EAA traits shared high strain-level correlations between diets, indicating
333 an effect of background genotype that is robust to dietary differences (**Table 2**). The

methylome-wide entropy had a heritability of ~0.30, and had no strain-level correlation between diets.

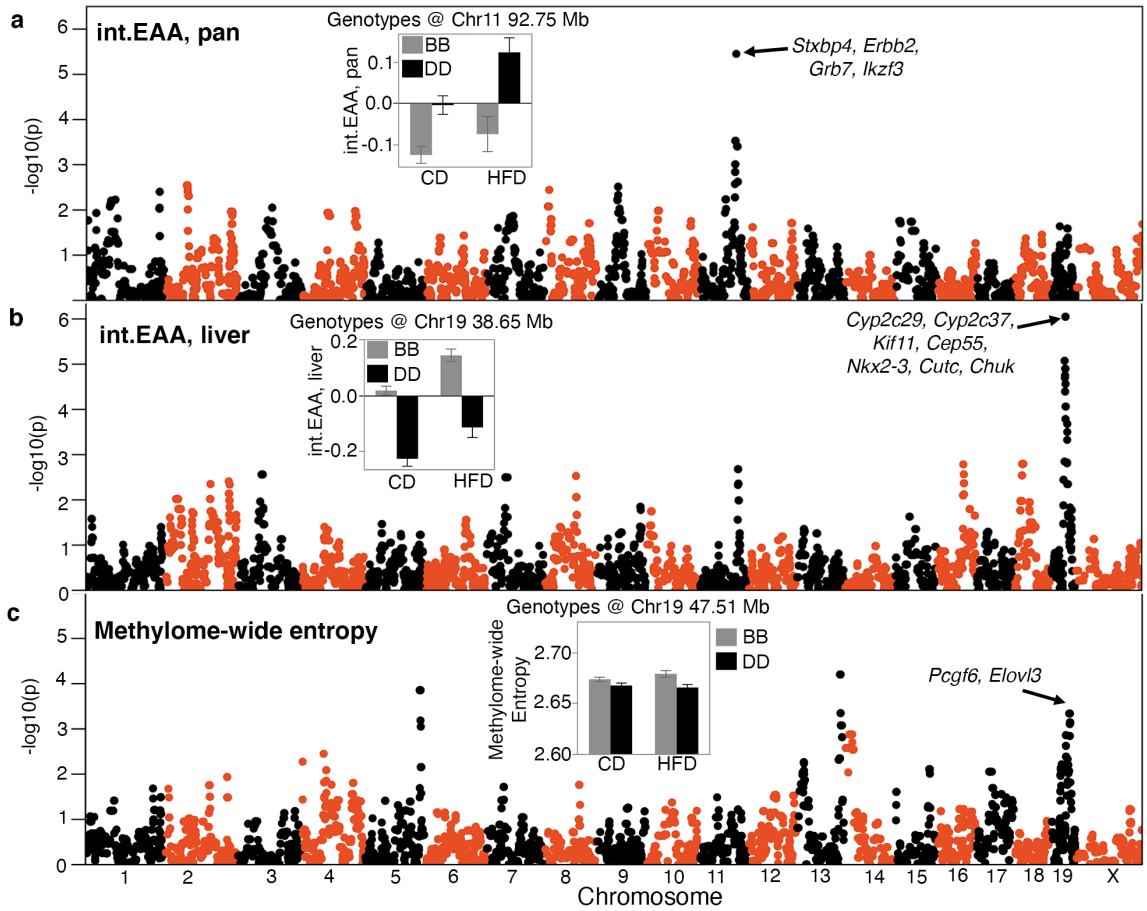


Fig 4. QTL maps for the DNAm readouts

The Manhattan plots represent the location of genotyped markers (x-axis), and linkage – $\log_{10}p$ (y-axis). **(a)** The peak quantitative trait locus (QTL) for age acceleration from the pan-tissue interventional clock (int.EAA) is on chromosome (Chr) 11 at ~93 Mb. The inset shows the mean (± standard error) trait values for BXDs the are homozygous for the C57BL/6J allele (*BB*; grey) versus BXDs homozygous for the DBA/2J allele (*DD*; black) on control diet (CD) and high fat diet (HFD). **(b)** The liver-specific int.EAA has a peak QTL on Chr19 (~38 Mb). Trait means by genotype at this locus are shown in inset; *BB* has higher age acceleration. **(c)** Linkage statistics are weaker for the methylome-wide entropy. However, there is a nominally significant linkage on the Chr19 loci, but the peak markers are at ~47.5 Mb. Here the *BB* genotype has higher entropy.

To uncover genetic loci, we applied QTL mapping using mixed linear modeling that corrects for the BXD kinship structure.[49] First, we performed the QTL mapping for each EAA traits with adjustment for diet and body weight. EAA from the two interventional clocks had the strongest QTLs (**Data S9**). The pan-tissue int.EAA had a significant QTL on Chr11 (90–99 Mb) with the highest linkage at ~93 Mb (p = 3.5E-06; equivalent to a LOD score of 4.7) (**Fig 4a**). Taking a genotype marker at the peak interval (BXD variant ID DA0014408.4 at Chr11, 92.750 Mb)[34], we

342 segregated the BXDs homozygous for either the D2 (*DD*) or the B6 (*BB*) alleles. Strains with *DD*
343 genotype at this locus had significantly higher int.EAA (**Fig 4a** inset). The liver int.EAA had the
344 highest QTL on Chr19 (35–45 Mb) with the most significant linkage at markers between 38–42
345 Mb (p = 9E-07; LOD score of 5.2) (**Fig 4b**). We selected a marker at the peak interval
346 (rs48062674 at Chr19, 38.650 Mb), and the *BB* genotype had significantly higher liver int.EAA
347 compared to *DD* (**Fig 4b** inset).

348 We performed a similar QTL mapping for methylome-wide entropy with adjustment for major
349 covariates (diet, chronological age, and body weight). There were no genome-wide significant
350 QTLs. A region on Chr19 that overlapped the liver int.EAA showed a modest peak (**Fig 4c**; **Data**
351 **S9**). However, the peak markers for entropy were located slightly distal to the peak EAA QTL
352 (~47.5 Mb at rs30567369, minimum p = 0.0005). At this locus, the *BB* genotype had higher
353 average entropy.

354 To identify regulatory loci that are consistent across the different EAA measures, we applied a
355 multi-trait analysis and derived the linkage meta-p-value using a p-value combination for the six
356 EAA traits.[50] The peaks on Chrs 11 and 19 attained the highest consensus p-values (**Fig S4a;**
357 **Data S9**). There was another potential consensus peak at combined $-\log_{10}p > 6$ on Chr3 (~54
358 Mb). We focus on the Chrs 11 and 19 QTLs and refer to these as *EAA QTL on Chr 11* (*Eaa11*),
359 and *EAA QTL on Chr 19* (*Eaa19*). *Eaa11* extends from 90–99 Mb. For *Eaa19*, we delineated a
360 broader interval from 35–48 Mb that also encompasses the peak markers for entropy.

361 We performed marker-specific linkage analyses for each of the clocks using a regression model
362 that adjusted for diet. With the exception of the liver int.EAA, all the EAA traits had nominal to
363 highly significant associations with the representative *Eaa11* marker (DA0014408.4), and the
364 *DD* genotype had higher age acceleration (**Table 3**). Mean plots by genotype and diet shows
365 that this effect was primarily in the CD mice (**Fig S4b**). The effect of this locus appeared to be
366 higher for the pan-tissue clocks compared to the corresponding liver-specific clocks. For
367 proximal *Eaa19*, the representative marker (rs48062674) was associated with all the EAA traits
368 and the *BB* mice had higher age acceleration on both diets (**Fig S4c**). We also tested if these
369 peak markers were associated with the recorded lifespan phenotype and we found no
370 significant association with the observed lifespan of the BXDs.

371 **Table 3: Marker specific linkage analyses for epigenetic age acceleration and body weight**
372 **trajectory**

		Linear regression[1]			
Predictor	**Outcome**	**Estimate**	**Std Error**	**t Ratio**	**p**
Eaa11 DA0014408.4[DD] Chr11, 92.750 Mb (133 *BB* cases, and 173 *DD* cases)	EAA, pan	0.096	0.023	4.184	3.8E-05
	EAA, liver	0.067	0.017	3.880	0.0001
	dev.EAA, pan	0.077	0.025	3.041	0.003
	dev.EAA, liver	0.037	0.020	1.878	0.06
	int.EAA, pan	0.153	0.029	5.278	2.5E-07
	int.EAA, liver	-0.033	0.025	-1.284	0.20
Eaa19 rs48062674[DD] Chr19, 38.650 Mb	EAA, pan	-0.083	0.028	-2.954	0.003
	EAA, liver	-0.137	0.020	-6.972	2.0E-11
	dev.EAA, pan	-0.206	0.029	-7.218	4.3E-12

(238 *BB* cases, and 67 *DD* cases)	dev.EAA, liver	-0.124	0.023	-5.461	9.9E-08
	int.EAA, pan	-0.143	0.035	-4.028	7.1E-05
	int.EAA, liver	-0.250	0.027	-9.238	4.6E-18
	Mixed model for longitudinal change in body weight[2]				
Predictor	**Outcome**	**Estimate**	**Std Error**	**t Ratio**	**p**
Eaa11 DA0014408.4[DD] Number of observations = 6885; number of individuals = 2112	Body weight	0.619	0.345	1.794	0.07
Eaa19 rs48062674[DD] Number of observations = 6132; number of individuals = 1852	Body weight	-1.847	0.374	-4.945	7.6E-07

[1]Regression model: lm(EAA ~ genotype + diet); [2]lmer(weight ~ age + diet + genotype + (1|mouseID)

Association of EAA QTLs with body weight trajectory

Since gain in body weight with age was an accelerator of the clocks, we examined whether the selected markers in *Eaa11* and *Eaa19* were also related to body weight change. We retrieved longitudinal weight data from a larger cohort of the aging BXD mice that were weighed at regular intervals. After excluding heterozygotes, we tested the effect of genotype. Concordant with the higher EAA for the *DD* genotype at *Eaa11* in the CD group, the *DD* genotype in the CD group also had slightly higher mean weight at older adulthood (12 and 18 months; **Fig 5a**). However, this marker had no significant association with body weight when tested using a mixed effects model (p = 0.07; **Table 3**). In *Eaa19*, it was the *BB* genotype that consistently exhibited an accelerated clock on both diets, and also higher entropy, and the *BB* genotype had higher average body weight by 6 months of age (**Fig 5b**), and this locus had a significant influence on the body weight trajectory (p = 7.6E-07; **Table 3**).

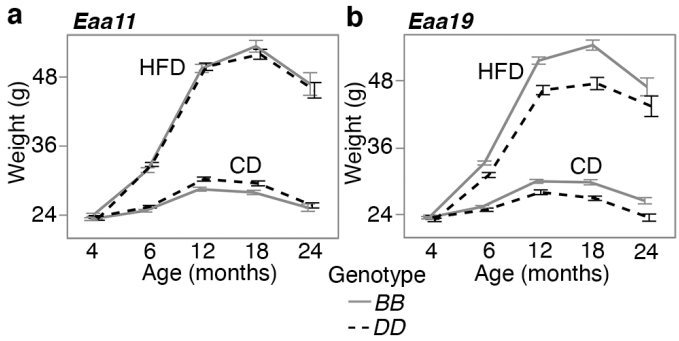


Fig 5. Body weight trajectory by diet and genotype Body weight was measured at regular age intervals (x-axis) from **(a)** 2112 BXD mice that were homozygous at the *Eaa11* marker (DA0014408.4; 842 *BB*, 1279 *DD*), and **(b)** 1852 BXD mice that were homozygous at the proximal *Eaa19* marker (rs48062674; 1252 *BB*, 600 *DD*). Mice were maintained on either control diet (CD) or high fat diet (HFD). The graphs show the segregation of body weight over time by diet and genotype. Mean ± standard error; heterozygotes were excluded.

Candidate genes for epigenetic age acceleration

There are several positional candidate genes in *Eaa11* and *Eaa19*. To narrow the list, we applied two selection

400 criteria: genes that (1) contain missense and/or stop variants, and/or (2) contain non-coding
401 variants and regulated by cis-acting expression QTLs (eQTL). For the eQTL analysis, we utilized
402 an existing liver transcriptome data from the same aging cohort.[32] We identified 24 positional
403 candidates in *Eaa11* that includes *Stxbp4*, *Erbb2* (*Her-2* oncogenic gene), and *Grb7* (growth
404 factor receptor binding) (**Data S10; Fig 4a**). *Eaa19* has 81 such candidates that includes a cluster
405 of cytochrome P450 genes, and *Chuk* (inhibitor of NF-kB) in the proximal region, and *Pcgf6*
406 (epigenetic regulator) and *Elovl3* (lipid metabolic gene) in the distal region (**Data S10; Fig 4b,**
407 **4c**).

408 For further prioritization, we converted the mouse QTL regions to the corresponding syntenic
409 regions in the human genome, and retrieved GWAS annotations for these intervals.[51] We
410 specifically searched for the traits: epigenetic aging, longevity, age of
411 menarche/menopause/puberty, Alzheimer's disease, and age-related cognitive decline and
412 dementia. This highlighted 5 genes in *Eaa11*, and 3 genes in *Eaa19* (**Table S3**). We also
413 identified a GWAS study that found associations between variants near *Myof-Cyp26a1* and
414 human longevity,[41] and a meta-GWAS that found gene-level associations between *Nkx2–3* and
415 *Cutc*, and epigenetic aging (**Table S3**).[27]

416 Gene expression correlates of EAA

417 A subset of the BXD cases had liver RNA-seq data (94 CD, and 59 HFD).[32] Using this set, we
418 performed transcriptome-wide correlation analysis for the general pan-tissue EAA, and the
419 more specific liver int.EAA. To gain insights into biological pathways, we selected the top 2000
420 transcriptome correlates for functional enrichment analysis (**Data S11**). The common themes
421 for both clocks were: (1) there were far fewer negative correlates (223 out of 2000 for pan-
422 tissue EAA, and 337 out of 2000 transcripts for liver int.EAA) than positive correlates, (2) the
423 negative correlates were highly enriched (Bonferroni correct $p < 0.05$) in oxidation-reduction
424 and mitochondrial genes (**Data S12, Data S13**). The pan-tissue general clock was also highly
425 enriched in pathways related to steroid metabolism, epoxygenase p450 pathway, and
426 xenobiotics, which are pathways that are particularly relevant to liver. The p450 genes included
427 candidates that are in *Eaa19* (e.g., *Cyp2c29*, *Cyp2c37*). The positive correlates were enriched in
428 a variety of gene functions including mitosis for both clocks, and immune and inflammatory
429 response for the general pan-tissue clock (functions that are not specific to liver). 563
430 transcripts (315 unique genes) were correlated with both the pan-tissue EAA, and the liver
431 int.EAA. Based on hierarchical clustering (HC) of these common mRNA correlates of EAA, the
432 transcripts could be clustered into 3 groups (**Fig 6a;** heatmap in **Fig S5a**). While none of these
433 were significantly enriched in any particular gene ontology (GO), cluster 3 included several
434 oxidation-reduction genes including the *Eaa11* candidate, *Cyp2c29*, and cluster 2 included
435 several cell cycle genes (**Fig 6a**). To verify that these transcriptomic associations are robust to
436 the effect of diet, we repeated the correlation and enrichment analysis in the CD group only for
437 the pan-tissue general clock ($n = 94$). Again, taking the top 2000 correlates ($|r| \leq 0.22$; $p \leq 0.03$),
438 we found the same enrichment profiles for the positive correlates (immune, cell cycle) and the
439 negative correlates (oxidation-reduction and mitochondrial) (**Data S12**).

440 Liver proteome was also available for 164 of the BXDs, and 53 also had adipose proteome. The
441 liver proteome data quantifies over 32000 protein variants from 3940 unique genes and has

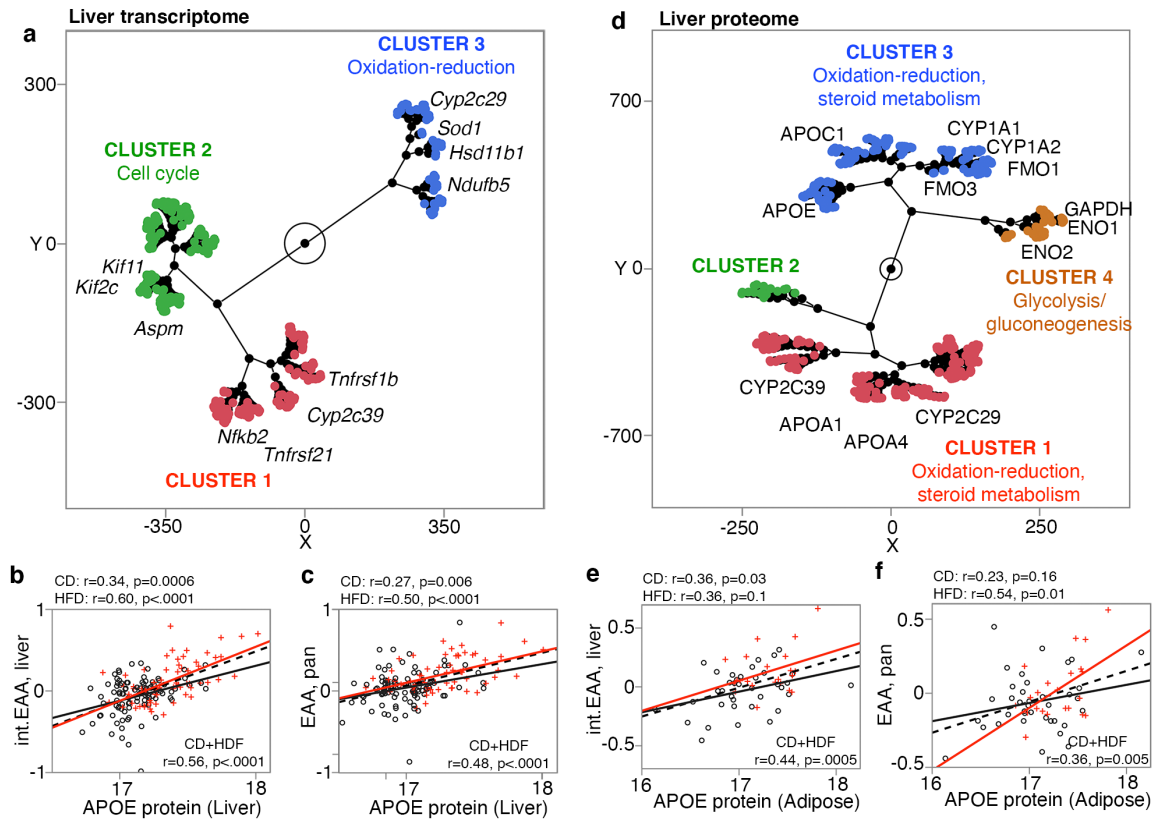


Fig 6. Gene expression correlates of epigenetic age acceleration
(a) mRNAs that were correlated with the acceleration of both the pan-tissue general clock (pan EAA), and the liver interventional clock (liver int.EAA) were grouped based on unsupervised hierarchical clustering (HC). Few representative genes and gene ontologies are highlighted. For liver proteome, the level of APOE was the strongest correlate for both the **(b)** liver int.EAA, and **(c)** the pan-tissue EAA. **(d)** For liver proteins that were correlated with both pan-tissue EAA and liver int.EAA, HC grouped the proteins into clusters that were enriched in oxidation-reduction and lipid metabolism, and a cluster enriched in glycogen metabolism. In adipose tissue, the expression level of the APOE protein was higher with higher age acceleration for both the **(e)** liver int.EAA, and **(f)** the pan-tissue EAA.

been reported in Williams et al.[32] Similar to the transcriptome-wide analysis, we extracted the top 2000 protein correlates of EAA (**Data S14**), and performed functional enrichment analysis (**Data S12, S13**). For both the liver int.EAA and the pan-tissue EAA, the top liver protein correlate was APOE, and higher expression of APOE was associated with higher age acceleration (**Fig 6b, c**). Similar to the transcriptome, the negative correlates of EAA were highly enriched in oxidation-reduction (several cytochrome proteins), steroid metabolism, and epoxygenase 450 pathway. The positive correlates were also highly enriched in oxidation-reduction (several hydroxy-delta-5 steroid dehydrogenases proteins), lipid and carbohydrate metabolism, as well as phospholipid efflux (particularly enriched for the liver int.EAA) (**Data S13**). There was a high degree of overlap at the proteomic level for the two clocks and 1241 proteins variants (332 unique genes) were correlated with both the pan-tissue EAA and the liver int.EAA (**Data S14**). For these common protein correlates, the HC divided the proteins into clusters that

454 represented metabolic pathways mainly related to steroid metabolism, but also glycolysis and
455 gluconeogenesis (**Fig 6d**; heatmap in **Fig S5b**).

456 Finally, we used the adipose proteome data for a proteome-wide correlational analysis for the
457 pan-tissue EAA and liver int.EAA. We took only the top 1000 correlates (due to the small sample
458 size), and a functional enrichment analysis showed consistent enrichment in metabolic
459 pathways related to fatty acids and also carbohydrates, and cell proliferation genes for the pan-
460 tissue EAA (**Data S12, S13**). For the adipose proteome, the cytochrome p450 genes were no
461 longer enriched. However, the overall functional profile highlighted metabolic pathways as
462 important gene expression correlates of EAA. Furthermore, for both the liver and adipose
463 proteomes, APOE levels were highly correlated with EAA that indicates a higher level of this
464 apolipoprotein in both tissues is associated with higher age acceleration (**Fig 6e, 6f**).

465 # Discussion

466 Here we have tested the performance of DNAm clocks derived from highly conserved CpGs, and
467 described the dynamism and variability of site-specific methylation. While age is a major source
468 of variance, we detected joint modulation by diet, body weight, and genotype-by-diet life
469 expectancy. HFD had an age accelerating effect on the clocks, and this is concordant with our
470 previous report where we found more rapid age-associated changes in methylation.[39] This also
471 concurs with studies in humans that have found that obesity accelerates epigenetic aging.[52,53]
472 However, when BWF was included in the regression term, the effect of diet became
473 inconsistent. This suggests that the effect of diet on EAA is mediated by the changes in weight
474 and metabolic traits such as total cholesterol. Body weight in particular, had a strong age-
475 accelerating effect. The effect of weight may manifest early on, and even in the CD group,
476 higher weight gains at younger age (between 4–6 months) was associated with higher EAA later
477 in life.

478 We tested different mouse DNAm clocks, and the main difference between these clocks was
479 the subsets of CpGs that were used for training. It is well-known that DNAm clocks have high
480 level of degeneracy.[3,14] In other words, highly accurate predictors of chronological age can be
481 built from entirely different sets of CpGs and different weight coefficients. This is likely because
482 a large proportion of CpGs undergo some degree of change with age, and combinatorial
483 information from any subset of this is informative of age. For instance, even at a very stringent
484 cutoff of Bonferroni 0.05 that treated the 27966 CpGs as "independent", we still detected 6553
485 CpGs as age-DMC, i.e., close to a quarter of the CpGs we tested. Clocks built from pre-selected
486 CpGs that are at conserved sequences are known to be sensitive to the effects of pro-longevity
487 interventions such as calorie restriction and growth hormone receptor deletion.[3,54] And while
488 all these DNAm clocks achieve reasonably high prediction of chronological age, the age
489 divergence derived from these different clocks (EAA) can capture slightly different facets of
490 biological aging, and the better a clock is at predicting chronological age, the lower its
491 association with mortality risk.[13,14] In the present study, we find that the interventional clocks
492 deviated most from chronological age, and this is expected as these were built from a much
493 smaller set of CpGs (see Methods). The interventional clocks were also associated with BWF
494 and cholesterol, but had weaker associations with BW0. The liver int.EAA had the highest
495 positive correlation with methylome-wide entropy, and was the clock that had the strongest

496 inverse correlation with strain longevity. In contrast, the developmental clocks, which were
497 based on CpGs that change early in life, showed a stronger association with BW0. The contrast
498 between the interventional and developmental clocks suggests that while one is more
499 modifiable, the other is more informative of baseline characteristics that influence aging later in
500 life. The pan-tissue clock, which was not constrained to any preselected set of CpGs or tissue,
501 also performed well in capturing biological aging and was accelerated by both BW0 and BWF,
502 diet (when BW0 was the weight term in the regression model), higher entropy, and had a
503 modest but significant inverse correlation with strain lifespan.

504 Entropy, a measure of noise and information loss, increases as a function of time and age.[10,55-57]
505 In the context of the methylome, the shift to higher entropy represents a tendency for the
506 highly organized hypo- and hypermethylated landscape to erode towards a more hemi-
507 methylated state.[10,43,45] This increase in disorder, particularly across CpGs that are highly
508 conserved, could have important functional consequences. The entropy of age-gain CpGs
509 predicted strain lifespan, and was increase by HFD. Overall, we find that mice belonging to
510 longer-lived BXD strains had a more "youthful" methylome with lower entropy at the age-gain
511 CpGs. The entropy of age-loss CpGs on the other hand, was related to the body weight of mice,
512 and both higher BW0 and BWF were associated with higher entropy. This leads us to suggest
513 that the rate of noise accumulation, an aspect of epigenomic aging, can vary between
514 individuals, and the resilience or susceptibility to this shift towards higher noise may be partly
515 modulate by diet as well as genetic factors.

516 Somewhat surprising was the inverse correlation between the entropy of age-DMCs and fasted
517 glucose. This lower entropy of age-gain CpGs with higher glucose is somewhat counter to the
518 general tendency for strains with shorter lifespan to have higher glucose.[33] In biological
519 systems, entropy is kept at bay by the uptake of chemical energy, and investment in
520 maintenance and repair,[57] and we can only speculate that at least in mice, the higher amount of
521 glucose after overnight fast may be associated with a more ordered methylome. The centrality
522 of bioenergetics for biological systems may explain why we detect this coupling between the
523 DNAm readouts (i.e., the clocks, and entropy), and indices of metabolism including weight, diet,
524 levels of macronutrients, and even expression of metabolic genes. As cogently highlighted by
525 Donohoe and Bultman,[58] many metabolites (e.g., SAM, NAD^+, ATP) are essential co-factors for
526 enzymes that shape the epigenome, and these could serve as nutrient sensors and mechanistic
527 intermediaries that regulate how the epigenome is organized in response to metabolic
528 conditions. Close interactions between macro- and micronutrients, and DNAm is a conserved
529 process and plays a critical role in defining both physiology and body morphology.[59,60] Overall,
530 our results suggests that a higher metabolic state is associated with higher entropy and EAA,
531 and potentially, lower lifespan.

532 For the BXDs, life expectancy is highly dependent on the background genotype, and mean
533 lifespan varies from under 16 months for strains such as BXD8 and BXD13, to over 28 months in
534 strains such as BXD91 and BXD175.[33,36,38] The EAA showed the expected inverse correlation
535 with lifespan, but the effect was modest and only significant for the pan-tissue EAA and the
536 liver int.EAA. The association of lifespan with the entropy of age-gain CpGs was slightly
537 stronger. We must point out that the analysis between the epigenetic readouts and lifespan
538 was an indirect comparison. Unlike the comparison with body weight and metabolic traits,

539 which were traits measured from the same individual, the lifespan data are strain
540 characteristics computed from a parallel cohort of mice that were allowed to survive till natural
541 mortality, and this may partly explain the weaker associations with EAA. Nonetheless, our
542 observations indicate that genotypes with higher life-expectancy have generally lower entropy,
543 and lower methylation levels at the age-gain CpGs, and these properties of the methylome are
544 likely to be partly under genetic modulation.

545 Our goal was to take these different clocks and identify regulatory loci that were the most
546 stable and robust to the slight algorithmic differences in building the clocks. A notable
547 candidate in *Eaa11* is Syntaxin binding protein 4 (*Stxbp4*, aka, *Synip*), located at 90.5 Mb. *Stxbp4*
548 is a high-priority candidate due to the concordant evidence from human genetic studies. The
549 conserved gene in humans is a replicated GWAS hit for the intrinsic rate of epigenetic
550 aging.[24,26,27] In the BXDs, *Stxbp4* contains several non-coding variants, and a missense mutation
551 (rs3668623), and the expression of *Stxbp4* in liver is modulated by a *cis*-eQTL. *Stxbp4* plays a
552 key role in insulin signaling,[61] and has oncogenic activity and implicated in different cancers.[62,63]
553 Furthermore, GWAS have also associated *STXBP4* with age of menarche.[64,65] *Eaa11* corresponds
554 to the 17q12-21 region in humans, and the location of additional oncogenic genes, e.g.,
555 *ERBB2/HER2*, *GRB7*, and *BRCA1*.[66] The mouse *Brca1* gene is a little distal to the peak QTL region
556 and is not considered a candidate here, although it does segregate for two missense variants in
557 the BXDs. *Erbb2* and *Grb7* are in the QTL region, and *Erbb2* contains a missense variant
558 (rs29390172), and *Grb7* is modulated by a *cis*-eQTL. *Nr1d1* is another candidate in *Eaa11*, and
559 the co-activation of *Erbb1*, *Grb7*, and *Nr1d1* has been linked to breast and other cancers.[67,68]

560 *Eaa19* was consistently associated with EAA from all the clocks we evaluated, and also with
561 body weight gains, irrespective of diet. DNAm entropy may also have a weak association with
562 markers at this interval. The EAA traits have peak markers in the proximal part of *Eaa19*
563 (around the cytochrome cluster), and the methylome-wide entropy had a weak peak that was
564 in the distal portion (over candidates like *Elovl3*, *Pcgf3*). Two candidates in *Eaa19* have been
565 implicated in epigenetic aging in humans based on gene-level meta-GWAS: NK homeobox 3
566 (*Nkx2-3*, a developmental gene), and CutC copper transporter (*Cutc*).[27] *Eaa19* is also the
567 location of the *Cyp26a1-Myof* genes, and the human syntenic region is associated with
568 longevity, metabolic traits, and lipid profiles.[41,69,70] Another noteworthy candidate in *Eaa19* is
569 *Chuk*, a regulator of mTORC2, that has been associated with age at menopause.[64,71] *Eaa19*
570 presents a complex and intriguing QTL related to the DNAm readouts that may also influence
571 body weight gains over the course of life. Both *Eaa19* and *Eaa11* exemplify the major challenge
572 that follows when a genetic mapping approach leads to gene- and variant-dense regions.[72,73]
573 Both loci have several biologically relevant genes, and identifying the causal gene (or genes) will
574 require a more fine-scaled functional genomic dissection.

575 The gene expression analyses highlighted metabolic pathways. At the mRNA level, the negative
576 correlates of EAA were highly enriched in metabolic genes related to oxidation-reduction and
577 steroid metabolism, while the positive correlates were enriched in pathways related to mitosis,
578 and immune response for the pan-tissue general EAA. This convergence on metabolic, immune
579 and cell division genes is very consistent with previous reports.[14,28,44] Here we should note that
580 depending on the tissue(s) in which the clocks are trains, and the tissue from which the
581 DNAmAge is estimated, the EAA derivative may put an emphasis on biological pathways or

582 genes that are most relevant to that tissue. For instance, clocks optimized for neural tissue are
583 more closely related to neurodegeneration and neuropathologies.[18,74] With the liver clocks,
584 expression correlates highlighted aspects of metabolism that are relevant to liver function (e.g.,
585 the cytochrome p450 epoxygenase genes), and this is detected both at the transcriptomic, and
586 proteomic levels. For the adipose tissue proteome, the cytochrome genes become less
587 prominent, but the enriched pathways still remained consistent (i.e., oxidation-reduction, lipid
588 and carbohydrate metabolism, and cell proliferation for the positive correlates of the pan-tissue
589 EAA). At the proteome level, we also find several phospholipid efflux genes (APOC1, APOA2,
590 APOC3, APOA1, APOA4, APOE) that are positive correlates of EAA. For both the liver and
591 adipose proteomes, APOE stands out as the top protein correlate of EAA. A recent human study
592 has also identified the APOE locus as the strongest GWAS hit for two measures of biological age
593 acceleration (the phenoAge, and the bioAge).[28] While more specific to liver, the cytochrome
594 P450 genes presents as both positional candidates, and expression correlates of EAA. These
595 genes have high expression in liver, and have major downstream impact on metabolism.[75-77]
596 One caveat is that these CYP genes are part of a gene cluster in *Eaa19* that includes transcripts
597 with *cis*-eQTLs (e.g., *Cyp2c66*, *Cyp2c39*, *Cyp2c68*), and the tight clustering of the genes, and
598 proximity of trait QTL and eQTLs may result in tight co-expression due to linkage
599 disequilibrium.[78] Nonetheless, the cytochrome genes in *Eaa19* are strong candidate modulators
600 of EAA derived from liver tissue that calls for further investigation.

601 Aside from *Eaa11* and *Eaa19*, another locus with evidence for consensus QTL was detected on
602 Chr3. We do not delve into this in the present work, but the Chr3 interval is near genes
603 associated with human epigenetic aging (*Ift80*, *Trim59*, *Kpna4*).[24,27] However, this QTL is
604 dispersed across a large interval, and the peak markers do not exactly overlap these human EAA
605 GWAS hits. While we have focused on *Eaa11* and *Eaa19*, the Chr3 locus presents a potentially
606 important region for EAA.

607 In summary, we have identified two main QTLs—*Eaa11* and *Eaa19*—that contribute to variation
608 in EAA. *Eaa11* contains several genes with oncogenic properties (e.g., *Stxbp4*, *Erbb2*), while
609 *Eaa19* contains a dense cluster of metabolic genes (e.g., *Elovl3*, *Chuk,* the cytochrome genes).
610 We demonstrate that metabolic profile and body weight are closely related to epigenetic aging
611 and methylome entropy. The convergence of evidence from genetic and gene expression
612 analyses indicates that genes involved in metabolism and energy balance contribute to the age-
613 dependent restructuring of the methylome, which in turn forms the basis of the epigenetic
614 clocks.

615 # Materials and Methods

616 ## Biospecimen collection and processing

617 Samples for this study were selected from a larger colony of BXD mice that were housed in a
618 specific pathogen-free (SPF) facility at the University of Tennessee Health Science Center
619 (UTHSC). All animal procedures were in accordance with a protocol approved by the
620 Institutional Animal Care and Use Committee (IACUC) at the UTHSC. Detailed description of
621 housing conditions and diet can be found in.[32,33] Mice were given *ad libitum* access to water,
622 and either standard laboratory chow (Harlan Teklad; 2018, 18.6% protein, 6.2% fat, 75.2%

623 carbohydrates), or high-fat chow (Harlan Teklad 06414; 18.4% protein, 60.3% fat, 21.3%
624 carbohydrate). Animals were first weighed within the first few days of assignment to either
625 diets, and this was mostly but not always prior to introduction to HFD. Following this, animals
626 were weighed periodically, and a final time (BWF) when animals were humanely euthanized
627 (anesthetized with avertin at 0.02 ml per g of weight, followed by perfusion with phosphate-
628 buffered saline) at specific ages for tissue collection. The present work utilizes the biobanked
629 liver specimens that were pulverized and stored in -80 °C, and overlaps samples described in
630 Williams et al.[32] DNA was extracted using the DNeasy Blood & Tissue Kit from Qiagen. Nucleic
631 acid purity was inspected with a NanoDrop spectrophotometer, and quantified using a Qubit
632 fluorometer dsDNA BR Assay.

633 Methylation array and quality checks

634 DNA samples from ~350 BXD mice were profiled on the Illumina HorvathHumanMethylChip40
635 array. Samples were in 96-well plate format (**Data S1**), and the plates were randomized for
636 major covariates such as age and diet. Details of this array are described in Arneson et al.[29] The
637 array contains probes that target ~36K highly conserved CpGs in mammals. Over 33K probes
638 map to homologous regions in the mouse genome. For downstream statistical tests, we further
639 filtered the probes and used only 27966 probes that have been validated for the mouse
640 genome using calibration data generated from synthetic mouse DNA.[29] Data was normalized
641 using the SeSame method.[79] Unsupervised HC was performed to identify outliers and failed
642 arrays, and those were excluded. We also performed strain verification as an additional quality
643 check. While majority of the probes were free of DNA sequence variants, we found 45 probes
644 that overlapped variants in the BXD family. We leveraged these as proxies for genotypes, and
645 performed a principal component analysis (PCA). The top genotype principal components
646 (genoPC1 and genoPC2; **Data S1**) segregated the samples by strain identity, and samples that
647 did not cluster close to the reported strains were removed. After excluding outliers, failed
648 arrays, and samples that failed strain verification, the final liver DNAm data consisted of 339
649 samples. The beta-values for these ~28K probes in the 339 samples show the expected bimodal
650 distribution (**Fig S6a**), but for these highly conserved CpGs, we note a much higher
651 representation of hypermethylated CpGs instead of the slightly hypomethylated state of the
652 methylome when a wider spectrum of CpGs is assayed.[43]

653 BXD-unbiased mouse clock estimation

654 Three different mouse clocks are reported here, and all three are based on penalized regression
655 modeling using glmnet.[80] Training was done in a larger mouse dataset that excluded the
656 BXDs.[30,31,42] The clocks are therefore unbiased to the characteristics of the BXDs. For pan-tissue
657 clocks, all mouse samples were used for training. For the liver specific clocks, the training was
658 limited to data from liver samples.

659 The general DNAmAge clock did not preselect for any CpGs and the full set of CpGs that map to
660 *Mus musculus* was used. First, a log-linear transformation was applied to the chronological age
661 using the function:

662 $$f(Age) = \begin{cases} \frac{Age}{1.2+0.06} + \log(1.2+0.06) - \frac{1.2}{1.2+0.06}, Age > 1.2 \\ \log(Age+0.06), Age \leq 1.2 \end{cases}$$

663 This is similar to the age transformation described in the original Horvath pan-tissue human
664 clock, but with offset at 0.06, and adult mouse age at 1.2.[11] Following this transformation, an
665 elastic net regression was implemented to regress the transformed chronological age on the
666 CpG beta-values in the training data. The alpha was set at 0.5, and the optimal lamda
667 parameter was determined by 10-fold cross-validation (function cv.glmnet). This selected
668 subsets of clock CpGs and coefficients (see **Data S2** for the lists of clock CpG, intercepts, and
669 coefficients). DNAmAge was then calculated as:

670 $$DNAmAge = f^{-1}\left(\frac{b_0 + b_1 CpG_1 + b_2 CpG_2 + \cdots + b_i CpG_i}{b_0 + b_1 + b_2 + \cdots + b_i}\right)$$

671 where b_0 is the intercept, and b_1 to b_i are the coefficients, and CpG_1 to CpG_i denote the beta-
672 values for the respective clock CpGs, and $f^{-1}()$ denotes the inverse function of $f()$.

673 A similar method was used to build the developmental and interventional clocks, but for these,
674 the CpGs were pre-selected. For the liver-specific developmental clock, CpGs that change
675 during mouse development was selected in liver samples based on Pearson correlation with
676 age in mice that were <1.6 months old. The top 1000 negative and top 1000 positive correlates
677 were then classified as "developmental CpGs", and the training was done using only this subset
678 of CpGs. For the pan-tissue dev.DNAmAge, the top 1000 positive and top 1000 negative
679 developmental CpGs were based on a multi-tissues EWAS, also using Pearson correlation with
680 age for mice <1.6 months old, and these are CpGs that are strongly correlated with age during
681 the mouse developmental period when all available tissues are considered.

682 Training for the interventional clock started with 537 CpGs that relate to gold-standard anti-
683 aging interventions (calorie restriction, growth hormone receptor knockout).[42,81] These
684 "interventional CpGs" were identified from an independent mouse liver calorie restriction (n =
685 95), and one growth hormone receptor knockout (n = 71) data that were not included in the
686 clock estimation.[42] Top CpGs associated with these interventions were identified and the 537
687 CpGs are the sites that are consistently associated with these anti-aging interventions. Of the
688 537, 121 CpGs increased in methylation, and 417 decreased in methylation with application of
689 the pro-longevity interventions. Given the small number of CpGs that went into training for the
690 int.DNAmAge, we expected this clock to be less correlated with chronological age, and possibly
691 more responsive variables such as diet.

692 **Entropy calculation**

693 Methylome-wide entropy was calculated from the 27966 probes. The beta-values were
694 discretized into 20 bins, and the Shannon entropy for each sample was estimated using the R
695 package, "entropy" (v1.2.1) with method = "ML": maximum likelihood.[82] The optimal number of
696 bins was determined using the Freedman-Diaconis rule (breaks = "FD" for the hist() function in
697 R). We also estimated the methylome-wide entropy after discretizing into 100 and 2000 bins
698 (values provided in **Data S1**), and the results we report are consistent and robust to the number

699 of bins. For the age-gain, age-loss, and age-ns CpGs, entropy for each set was estimated, also
700 following discretization into 20 bins.

701 Statistics

702 Statistical analyses were done using R or the JMP Pro software (version 15). Association
703 between the epigenetic predictors and continuous variables (body weight, strain lifespan,
704 fasted serum glucose, and total cholesterol) were based on Pearson correlations, and t-test was
705 used to evaluate the effect of categorical predictors (sex, diet). Multivariable regression models
706 were also used to control for covariates (R regression equations provided with **Table S1**, **Data**
707 **S4**, **S6**, and **Table 3**). All these traits are directly accessible from GeneNetwork 2 (GN2; more
708 information on how to retrieve these data from GN2 are provided in **Data S15**).[83,84] Longevity
709 data was obtained from a parallel cohort of BXD mice housed in the same UTHSC colony, and
710 members of this "longevity cohort" were allowed to age until natural death (more detail on the
711 longevity cohort can be found in [33]). Males were excluded and strain-by-diet lifespan summary
712 statistics were derived. Only strain-by-diet groups with 5 or more observations were included in
713 the correlational analyses with the epigenetic predictors.

714 Multivariable EWAS

715 Site-by-site differential methylation analysis (EWAS) was performed on the 27966 CpGs using a
716 multivariable regression model. As such genome-wide explorations are vulnerable to
717 unmeasured confounders, we included the top PC derived from a PCA of the 27966 probes.[85]
718 The top 10 principal components PCs cumulatively accounted for ~62% of the variance (**Fig**
719 **S6b**). A plot of PC1 (19% of variance) and PC2 (14% of variance) showed that PC1 captured
720 some noise due to batch (**Fig S6c**). The remaining top PCs (PC2 onwards) were strongly
721 associated with biological variables, particular age, and also weight and diet (top 10 PCs
722 provided in **Data S1**). For this reason, we included PC1 as a correction factor in the EWAS. The
723 regression model we used was: lm(CpG_i~ age + median lifespan + diet + BWF+ PC1), where CpG_i
724 is the ith CpG from 1 to 27966. As lifespan was from female mice, this EWAS excluded the few
725 male samples.

726 CpG annotation and enrichment

727 Functional annotation and enrichment analyses for the DMCs were done using the genomic
728 region enrichment R package, rGREAT (version 3.0.0)[46] with the array content (i.e., the 27966
729 CpGs) as background. Enrichment p-values are based on hypergeometric tests, and categories
730 with Benjamini-Hochberg adjusted p-values ≤ 0.05 are reported. Annotations were for the
731 GRCm38/mm10 reference genome.

732 For chromatin state annotation, we used bedtools to annotate the 27966 CpGs coordinates
733 using chromatin annotation .bed files for neonatal (P0) mouse liver tissue created by Gorkin et
734 al.[48,86] This provides the 15-states model using ChromHMM,[47] and we downloaded the file for
735 the "replicated set" (here, the regions annotated as NRS are sites that did not produce
736 replicable signal). Enrichment and depletion analyses for genomic annotations, and chromatin
737 annotations were based on the hypergeometric test (phyper R function). The R codes are
738 provided with the results data (**Data S8**).

Genetic analyses

The broad sense heritability within diet was estimated as the fraction of variability that was explained by background genotype.[34,87,88] For this, we applied an anova: aov(EAA ~ strain), and heritability was computed as: $H^2 = SSq_{strain}/(SSq_{strain} + SSq_{residual})$, where SSq_{strain} is the strain sum of squares, and $SSq_{residual}$ is the residual sum of squares.

All QTL mapping was done on the GN2 platform (trait accession IDs provided in **Data S15**).[83] In the GN2 home page, the present set of BXD mice belongs to the **Group: BXD NIA Longevity Study**, and GN2 provides a direct interface to the genotype data. All QTL mapping was done for genotypes with minor allele frequency ≥ 0.05 using the genome-wide efficient mixed model association (GEMMA) algorithm,[49] which corrects for the BXD kinship matrix. For the EAA traits, diet, weight at 6 months, and final weight were fitted as cofactor. Chronological age had not correlation with EAA and this was not included as a cofactor (including age does not change the results). Genome-wide linkage statistics were downloaded for the full set of markers that were available from GN2 (7320 markers in **Data S9**). For the combined p-values, QTL mapping was done separately using GEMMA for each EAA traits, then the Fisher's p-value combination was applied to get the meta-p-value.[50] We used this method to simply highlight loci that had consistent linkage across the different EAA measures. QTL mapping for methylome-wide entropy was done using GEMMA with adjustment for chronological age, diet, weight at 6 months, and final weight.

For marker specific linkage, we selected SNPs located at the peak QTL regions (DA0014408, rs48062674), and grouped the BXDs by their genotypes (F1 hybrids and other heterozygotes were excluded from this), and marker specific linkage was tested using ANOVA and linear regression (R regression equation given in **Table 3**). rs48062674 is a reference variant that is already catalogued in dbSNP,[89] and is used as a marker in the QTL mapping. DA0014408.4 is an updated variant at a recombinant region in the Chr11 interval and within the peak QTL interval.[34] Genotypes at these markers for individual BXD samples are in **Data S1**.

To test the effect of genotype on body weight change, body weight data measured at approximately 4 (baseline), 6, 12, 18, and 24 months were downloaded from GN2 (**Data S15**). Detailed description of these weight data are in Roy et al.[33] We then applied a mixed effects regression model using the lme4 R package[90]: lmer(weight ~ age + diet + genotype + (1|ID)), where ID is the identifier for individual mouse.

Bioinformatic tools for candidate genes selection

Sequence variation between B6 and D2 in the QTL intervals (Chr11:90–99 Mb, and Chr19:35–48 Mb) were retrieved from the Wellcome Sanger Institute Mouse Genomes Project database (release 1505 for GRCm38/mm10).[91-93] Positional candidates were required to contain at least one coding variant (missense and/or nonsense variants), or have non-coding variants with evidence of *cis*-regulation in liver tissue of the BXDs. *Cis*-eQTLs for the candidate genes were obtained from the liver RNA-seq data described in [32]. An interface to search and analyze this transcriptome data is available from GN2, and is catalogued under *Group: BXD NIA Longevity Study*; *Type: Liver mRNA*; and *Dataset: UTHSC BXD Liver RNA-seq (Oct 19) TMP Log2*.

779 For human GWAS annotations, we navigated to the corresponding syntenic regions on the
780 human genome by using the coordinate conversion tool in the UCSC Genome Browser. The
781 Chr11 90–95 Mb interval on the mouse reference genome (GRCm38/mm10) corresponds to
782 human Chr17:50.14–55.75 Mb (GRCh38/hg38) (40.7% of bases; 100% span). The Chr11 95–99
783 Mb interval in the mouse corresponds to human Chr17:47.49–50.14 Mb (29.3% of bases, 57.9%
784 span), and Chr17:38.19–40.39 Mb (20.7% of bases, 44.1% span). Likewise, for the Chr19 QTL,
785 the mm10 35–40 Mb corresponds to hg38 Chr10:89.80–95.06 Mb (32.2% of bases, 89.2% span),
786 40–45 Mb corresponds to hg38 Chr10:95.23–100.98 Mb (46.6% of bases, 95.6% span), and 45–
787 48 Mb corresponds to hg38 Chr10:100.98–104.41 Mb (46.5% of bases, 100% span). We then
788 downloaded the GWAS data for these regions from the NHGRI-EBI GWAS catalogue,[51] and
789 retained the GWAS hits that were related to aging.

790 Transcriptome and proteome analyses

791 The liver RNA-seq data mentioned above was also used for the transcriptome-wide
792 correlational analysis for EAA in the 153 cases that had both DNAm and RNA-seq data. We
793 considered the top 2000 highest mRNA correlates ($|r| = 0.24$, $p = 0.003$ for the pan-tissue EAA;
794 $|r| = 0.3$, $p = 0.0002$ for the liver int.EAA), and the list of transcripts were collapsed to a non-
795 redundant list of gene symbols, and this was uploaded to the DAVID Bioinformatics Database
796 (version 2021 update) for GO enrichment analysis.[94,95] Proteome correlational analysis was
797 carried out using the data: *Group: BXD NIA Longevity Study*; *Type: Liver Proteome*; and *Dataset:*
798 *EPFL/ETHZ BXD Liver Proteome CD-HFD (Nov19)*. Detailed description of this data is in Williams
799 et al.[32] 164 BXD cases had both DNAm and liver proteomics, and similar to the RNA-seq, we
800 selected the top 2000 correlates (($|r| = 0.24$, $p = 0.002$ for both the pan-tissue EAA and liver
801 int.EAA) for enrichment analysis.

802 59 of the BXD cases also have proteome data from adipose tissue (*Group: BXD NIA Longevity*
803 *Study*; *Type: Adipose Proteome*; and *Dataset: Riken-Wu BXD Liver Proteome CD-HFD (Sep20)*).
804 While small in sample number, we used this data to test whether we could recapitulate the
805 same functional enrichment profiles in a different tissue. Details on sample preparation and
806 processing steps for the adipose proteome is provided in the dataset's "Info" page on GN2. In
807 brief, protein was extracted from the adipose samples by first lysis in a buffer with protease
808 inhibitor, followed by homogenization with a glass dounce and sonication. The protein fraction
809 was isolated from the homogenate by centrifugation, and processed for assay on a liquid
810 chromatography tandem mass spectrometry (LC-M/MS) using a modified Phase Transfer
811 Surfactant Method as described in Mostafa et al.[96,97] Samples were measured using a Q
812 Exactive Plus Orbitrap LC–MS/MS System (Thermo Fisher). For each sample, 600 ng was
813 injected and the samples were measured with data-independent acquisition (DIA). A portion of
814 the peptides from the samples were pooled and fractionated using a Pierce High pH Reversed-
815 Phase (HPRP) Peptide Fractionation Kit (Thermo Fisher Scientific) to generate a spectral library.
816 For the HPRP fractions, 450 ng was injected and the samples were measured with data-
817 dependent acquisition (DDA). For protein identification, the raw measurement files were
818 searched against a mouse database using the (uniprot-reviewed_Mus_musculus_10090_.fasta)
819 using Proteome Discoverer v2.4 software (Thermo Fisher Scientific). Filtered output was used to
820 generate a sample-specific spectral library using the Spectronaut software (Biognosys,

Switzerland). Raw files from DIA measurements were used for quantitative data extraction with the generated spectral library, as previously described.[97] The false discovery rate was estimated with the mProphet approach and set to 0.01 at both peptide precursor level and protein level.[98,99] Due to the small sample size, for this dataset, we considered the top 1000 protein correlates of EAA (|r| = 0.25, p = 0.06 for the pan-tissue EAA; |r| = 0.31, p = 0.02 for the liver int.EAA).

Data availability

The full microarray data will be released via NCBI's Gene Expression Omnibus upon official publication. Genome annotations of the CpGs can be found on Github https://github.com/shorvath/MammalianMethylationConsortium. Individual level BXD data, including the processed microarray data are available on www.genenetwork.org on FAIR+ compliant format; data identifiers, and way to retrieve data are described in **Data S15**.

Acknowledgement. We thank the entire UTHSC BXD Aging Colony team, particularly Casey J Chapman, Melinda S McCarty, Jesse Ingles, and everyone else who contributed to the tissue harvest. We thank Evan G Williams for making the gene expression data readily available, and to David Ashbrook for providing the BXD genotypes. We thank the GeneNetwork team, especially Zach Sloan and Arthur Centeno, who have been extremely prompt and effective at assisting with the GeneNetwork interface. We thanks Dr. Garrett Jenkinson for the invaluable guidance he provided for entropy estimation.

Author contributions. KM contributed to the data, conceived portion of the study, and performed statistical analysis and drafted the article. ATL, CZL, AH contributed to the data analysis and in computing the epigenetic clocks. JVS contributed to the lab work. YW contributed to data generation. RWW conceived of the BXD Aging Colony, and provided access to the BXD biospecimen and data. SH developed the array platform, and built the epigenetic clocks and predictor. All authors contributed to, and approved the manuscript.

Research funding. This study was funded by the NIH NIA grants R21AG055841 and R01AG043930

Competing interests. SH is a founder of the non-profit Epigenetic Clock Development Foundation, which plans to license several of his patents from his employer UC Regents. The other authors declare no conflicts of interest.

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.

Reference:

1 Horvath, S. & Raj, K. DNA methylation-based biomarkers and the epigenetic clock theory of ageing. *Nature reviews. Genetics* **19**, 371-384, doi:10.1038/s41576-018-0004-3 (2018).

858 2 Bell, C. G. *et al.* DNA methylation aging clocks: challenges and recommendations.
859 *Genome biology* **20**, 249, doi:10.1186/s13059-019-1824-y (2019).
860 3 Thompson, M. J. *et al.* A multi-tissue full lifespan epigenetic clock for mice. *Aging* **10**,
861 2832-2854, doi:10.18632/aging.101590 (2018).
862 4 Porter, H. L. *et al.* Many chronological aging clocks can be found throughout the
863 epigenome: Implications for quantifying biological aging. *Aging cell* **20**, e13492,
864 doi:10.1111/acel.13492 (2021).
865 5 Zannas, A. S. *et al.* Lifetime stress accelerates epigenetic aging in an urban, African
866 American cohort: relevance of glucocorticoid signaling. *Genome biology* **16**, 266,
867 doi:10.1186/s13059-015-0828-5 (2015).
868 6 Marioni, R. E. *et al.* The epigenetic clock is correlated with physical and cognitive fitness
869 in the Lothian Birth Cohort 1936. *International journal of epidemiology* **44**, 1388-1396,
870 doi:10.1093/ije/dyu277 (2015).
871 7 Dugue, P. A. *et al.* DNA methylation-based biological aging and cancer risk and survival:
872 Pooled analysis of seven prospective studies. *Int J Cancer* **142**, 1611-1619,
873 doi:10.1002/ijc.31189 (2018).
874 8 Lu, A. T. *et al.* DNA methylation GrimAge strongly predicts lifespan and healthspan.
875 *Aging* **11**, 303-327, doi:10.18632/aging.101684 (2019).
876 9 Ryan, J., Wrigglesworth, J., Loong, J., Fransquet, P. D. & Woods, R. L. A systematic review
877 and meta-analysis of environmental, lifestyle and health factors associated with DNA
878 methylation age. *The journals of gerontology. Series A, Biological sciences and medical*
879 *sciences*, doi:10.1093/gerona/glz099 (2019).
880 10 Hannum, G. *et al.* Genome-wide methylation profiles reveal quantitative views of
881 human aging rates. *Molecular cell* **49**, 359-367, doi:10.1016/j.molcel.2012.10.016
882 (2013).
883 11 Horvath, S. DNA methylation age of human tissues and cell types. *Genome biology* **14**,
884 R115, doi:10.1186/gb-2013-14-10-r115 (2013).
885 12 Bocklandt, S. *et al.* Epigenetic predictor of age. *PloS one* **6**, e14821,
886 doi:10.1371/journal.pone.0014821 (2011).
887 13 Zhang, Q. *et al.* Improved precision of epigenetic clock estimates across tissues and its
888 implication for biological ageing. *Genome Med* **11**, 54, doi:10.1186/s13073-019-0667-1
889 (2019).
890 14 Liu, Z. *et al.* Underlying features of epigenetic aging clocks in vivo and in vitro. *Aging cell*
891 **19**, e13229, doi:10.1111/acel.13229 (2020).
892 15 Lee, Y. *et al.* Blood-based epigenetic estimators of chronological age in human adults
893 using DNA methylation data from the Illumina MethylationEPIC array. *BMC genomics* **21**,
894 747, doi:10.1186/s12864-020-07168-8 (2020).
895 16 Horvath, S. *et al.* Epigenetic clock for skin and blood cells applied to Hutchinson Gilford
896 Progeria Syndrome and ex vivo studies. *Aging* **10**, 1758-1775,
897 doi:10.18632/aging.101508 (2018).
898 17 Levine, M. E. *et al.* An epigenetic biomarker of aging for lifespan and healthspan. *Aging*
899 **10**, 573-591, doi:10.18632/aging.101414 (2018).

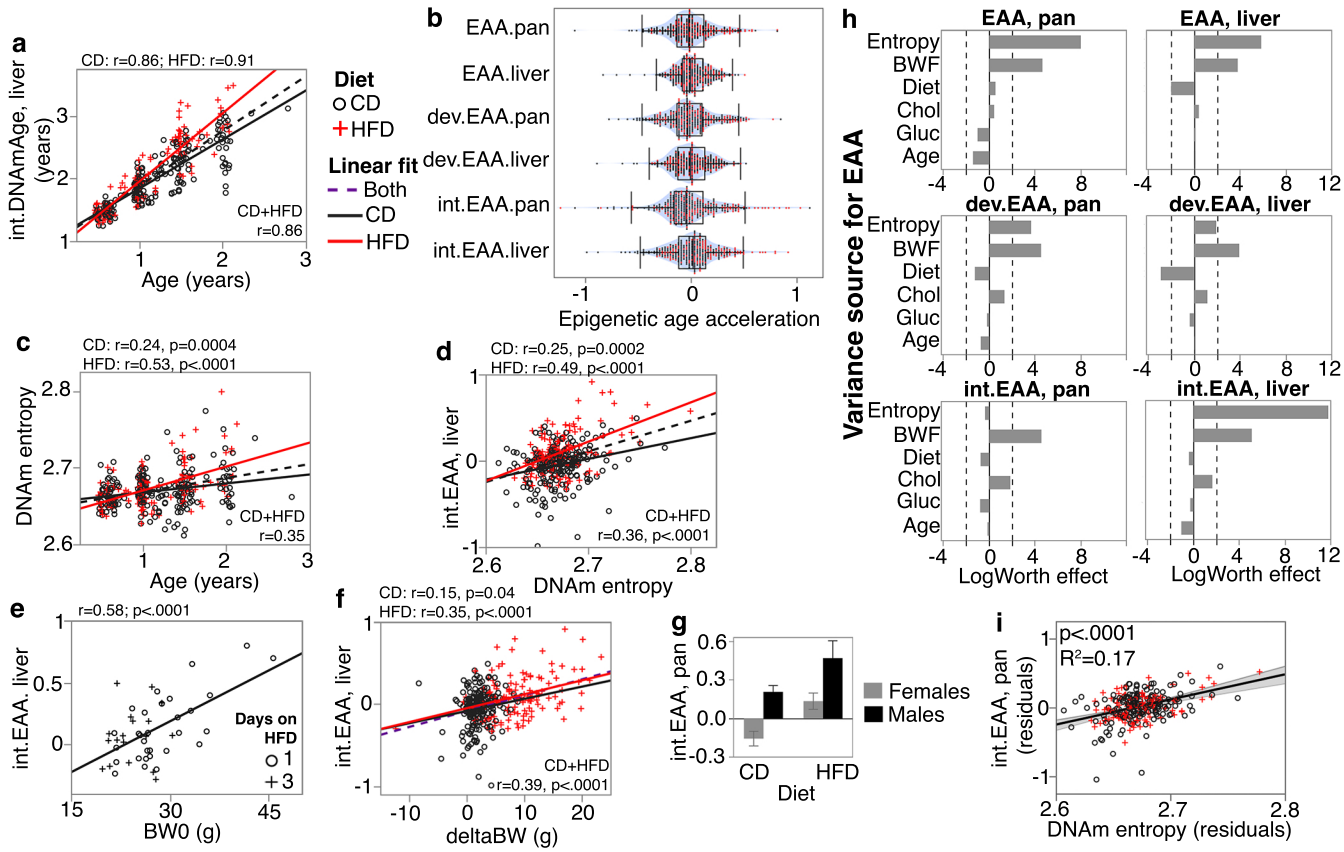
900 18 Shireby, G. L. *et al.* Recalibrating the epigenetic clock: implications for assessing
901 biological age in the human cortex. *Brain* **143**, 3763-3775, doi:10.1093/brain/awaa334
902 (2020).
903 19 Petkovich, D. A. *et al.* Using DNA Methylation Profiling to Evaluate Biological Age and
904 Longevity Interventions. *Cell metabolism* **25**, 954-960 e956,
905 doi:10.1016/j.cmet.2017.03.016 (2017).
906 20 Stubbs, T. M. *et al.* Multi-tissue DNA methylation age predictor in mouse. *Genome*
907 *biology* **18**, 68, doi:10.1186/s13059-017-1203-5 (2017).
908 21 Wang, T. *et al.* Epigenetic aging signatures in mice livers are slowed by dwarfism, calorie
909 restriction and rapamycin treatment. *Genome biology* **18**, 57, doi:10.1186/s13059-017-
910 1186-2 (2017).
911 22 Levine, M. *et al.* A rat epigenetic clock recapitulates phenotypic aging and co-localizes
912 with heterochromatin. *Elife* **9**, doi:10.7554/eLife.59201 (2020).
913 23 Han, Y. *et al.* Epigenetic age-predictor for mice based on three CpG sites. *Elife* **7**,
914 doi:10.7554/eLife.37462 (2018).
915 24 Gibson, J. *et al.* A meta-analysis of genome-wide association studies of epigenetic age
916 acceleration. *PLoS genetics* **15**, e1008104, doi:10.1371/journal.pgen.1008104 (2019).
917 25 Lin, W. Y. Genome-wide association study for four measures of epigenetic age
918 acceleration and two epigenetic surrogate markers using DNA methylation data from
919 Taiwan biobank. *Human molecular genetics*, doi:10.1093/hmg/ddab369 (2021).
920 26 Lu, A. T. *et al.* GWAS of epigenetic aging rates in blood reveals a critical role for TERT.
921 *Nature communications* **9**, 387, doi:10.1038/s41467-017-02697-5 (2018).
922 27 McCartney, D. L. *et al.* Genome-wide association studies identify 137 genetic loci for
923 DNA methylation biomarkers of aging. *Genome biology* **22**, 194, doi:10.1186/s13059-
924 021-02398-9 (2021).
925 28 Kuo, C. L., Pilling, L. C., Liu, Z., Atkins, J. L. & Levine, M. E. Genetic associations for two
926 biological age measures point to distinct aging phenotypes. *Aging cell* **20**, e13376,
927 doi:10.1111/acel.13376 (2021).
928 29 Arneson, A. *et al.* A mammalian methylation array for profiling methylation levels at
929 conserved sequences. *Nature communications* **13**, 783, doi:10.1038/s41467-022-28355-
930 z (2022).
931 30 Lu, A. T. *et al.* Universal DNA methylation age across mammalian tissues. *bioRxiv*,
932 2021.2001.2018.426733, doi:10.1101/2021.01.18.426733 (2021).
933 31 Li, C. Z. *et al.* Epigenetic predictors of maximum lifespan and other life history traits in
934 mammals. *bioRxiv*, 2021.2005.2016.444078, doi:10.1101/2021.05.16.444078 (2021).
935 32 Williams, E. G. *et al.* Multiomic profiling of the liver across diets and age in a diverse
936 mouse population. *Cell Syst*, doi:10.1016/j.cels.2021.09.005 (2021).
937 33 Roy, S. *et al.* Gene-by-environment modulation of lifespan and weight gain in the murine
938 BXD family. *Nat Metab* **3**, 1217-1227, doi:10.1038/s42255-021-00449-w (2021).
939 34 Ashbrook, D. G. *et al.* A Platform for Experimental Precision Medicine: The Extended
940 BXD Mouse Family. *Cell Syst*, doi:10.1016/j.cels.2020.12.002 (2021).
941 35 Peirce, J. L., Lu, L., Gu, J., Silver, L. M. & Williams, R. W. A new set of BXD recombinant
942 inbred lines from advanced intercross populations in mice. *BMC genetics* **5**, 7,
943 doi:10.1186/1471-2156-5-7 (2004).

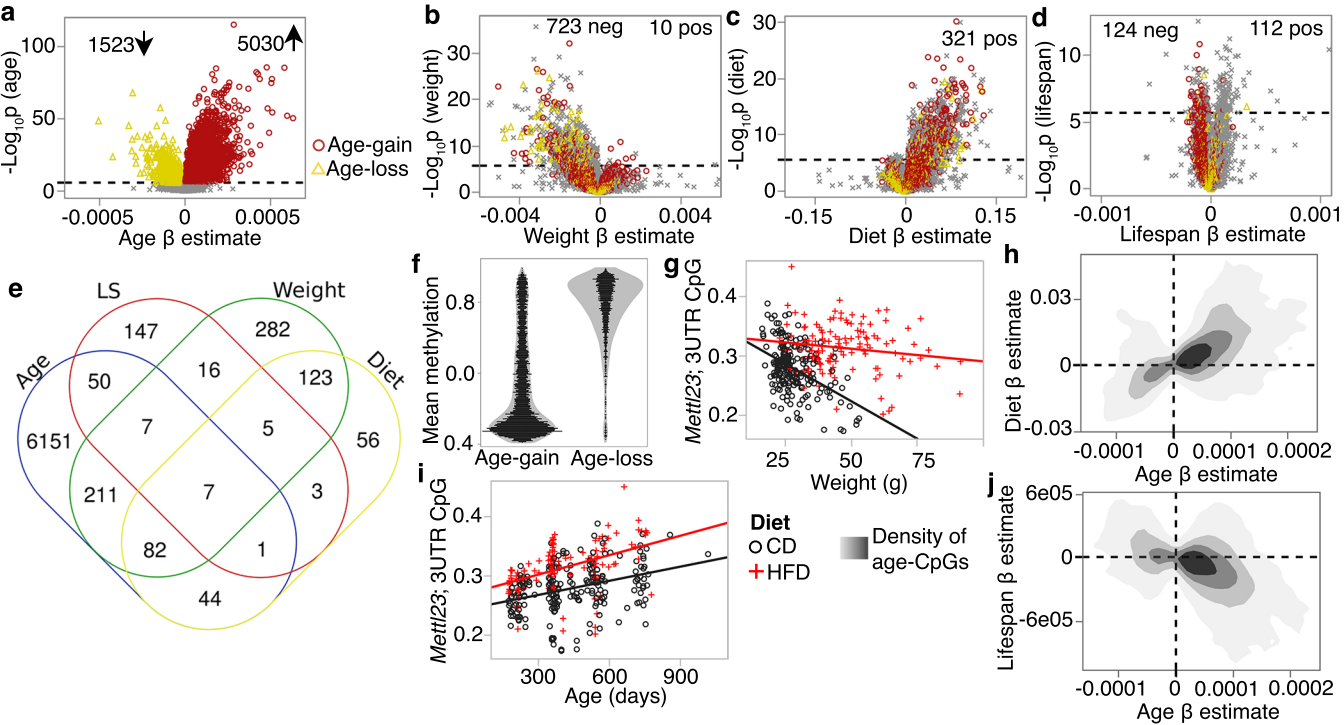
944 36 de Haan, G. & Williams, R. W. A genetic and genomic approach to identify longevity
945 genes in mice. *Mechanisms of ageing and development* **126**, 133-138,
946 doi:10.1016/j.mad.2004.09.012 (2005).
947 37 Hsu, H. C. *et al.* Age-related thymic involution in C57BL/6J x DBA/2J recombinant-inbred
948 mice maps to mouse chromosomes 9 and 10. *Genes and immunity* **4**, 402-410,
949 doi:10.1038/sj.gene.6363982 (2003).
950 38 Lang, D. H. *et al.* Quantitative trait loci (QTL) analysis of longevity in C57BL/6J by DBA/2J
951 (BXD) recombinant inbred mice. *Aging clinical and experimental research* **22**, 8-19
952 (2010).
953 39 Sandoval-Sierra, J. V. *et al.* Body weight and high-fat diet are associated with epigenetic
954 aging in female members of the BXD murine family. *Aging cell*, e13207,
955 doi:10.1111/acel.13207 (2020).
956 40 Benton, M. C. *et al.* Methylome-wide association study of whole blood DNA in the
957 Norfolk Island isolate identifies robust loci associated with age. *Aging* **9**, 753-768,
958 doi:10.18632/aging.101187 (2017).
959 41 Yashin, A. I. *et al.* Genetics of Human Longevity From Incomplete Data: New Findings
960 From the Long Life Family Study. *The journals of gerontology. Series A, Biological*
961 *sciences and medical sciences* **73**, 1472-1481, doi:10.1093/gerona/gly057 (2018).
962 42 Haghani, A. *et al.* Divergent age-related methylation patterns in long and short-lived
963 mammals. *bioRxiv*, 2022.2001.2016.476530, doi:10.1101/2022.01.16.476530 (2022).
964 43 Sziraki, A., Tyshkovskiy, A. & Gladyshev, V. N. Global remodeling of the mouse DNA
965 methylome during aging and in response to calorie restriction. *Aging cell* **17**, e12738,
966 doi:10.1111/acel.12738 (2018).
967 44 Slieker, R. C. *et al.* Age-related accrual of methylomic variability is linked to fundamental
968 ageing mechanisms. *Genome biology* **17**, 191, doi:10.1186/s13059-016-1053-6 (2016).
969 45 Kerepesi, C. *et al.* Epigenetic aging of the demographically non-aging naked mole-rat.
970 *Nature communications* **13**, 355, doi:10.1038/s41467-022-27959-9 (2022).
971 46 McLean, C. Y. *et al.* GREAT improves functional interpretation of cis-regulatory regions.
972 *Nature biotechnology* **28**, 495-501, doi:10.1038/nbt.1630 (2010).
973 47 Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and
974 characterization. *Nat Methods* **9**, 215-216, doi:10.1038/nmeth.1906 (2012).
975 48 Gorkin, D. U. *et al.* An atlas of dynamic chromatin landscapes in mouse fetal
976 development. *Nature* **583**, 744-751, doi:10.1038/s41586-020-2093-3 (2020).
977 49 Zhou, X. & Stephens, M. Efficient multivariate linear mixed model algorithms for
978 genome-wide association studies. *Nature methods* **11**, 407-409,
979 doi:10.1038/nmeth.2848 (2014).
980 50 Peirce, J. L., Broman, K. W., Lu, L. & Williams, R. W. A simple method for combining
981 genetic mapping data from multiple crosses and experimental designs. *PloS one* **2**,
982 e1036, doi:10.1371/journal.pone.0001036 (2007).
983 51 Catalog, G. <https://www.ebi.ac.uk/gwas/> (
984 52 Horvath, S. *et al.* Obesity accelerates epigenetic aging of human liver. *Proceedings of the*
985 *National Academy of Sciences of the United States of America* **111**, 15538-15543,
986 doi:10.1073/pnas.1412759111 (2014).

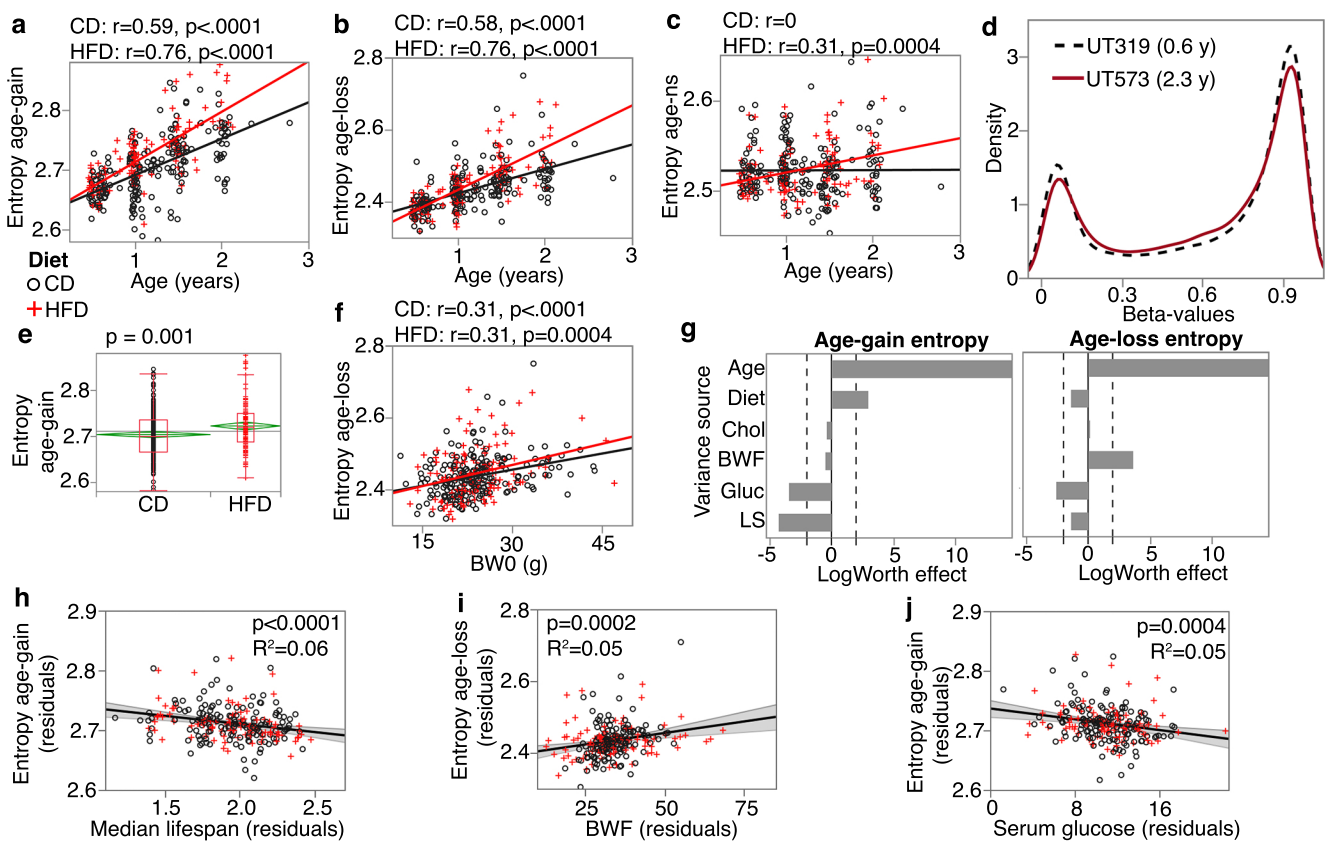
987 53 Nevalainen, T. *et al.* Obesity accelerates epigenetic aging in middle-aged but not in
988 elderly individuals. *Clinical epigenetics* **9**, 20, doi:10.1186/s13148-016-0301-7 (2017).
989 54 Wang, M. & Lemos, B. Ribosomal DNA harbors an evolutionarily conserved clock of
990 biological aging. *Genome research* **29**, 325-333, doi:10.1101/gr.241745.118 (2019).
991 55 Xie, H. *et al.* Genome-wide quantitative assessment of variation in DNA methylation
992 patterns. *Nucleic acids research* **39**, 4099-4108, doi:10.1093/nar/gkr017 (2011).
993 56 Jenkinson, G., Pujadas, E., Goutsias, J. & Feinberg, A. P. Potential energy landscapes
994 identify the information-theoretic nature of the epigenome. *Nature genetics* **49**, 719-
995 729, doi:10.1038/ng.3811 (2017).
996 57 Hayflick, L. Entropy explains aging, genetic determinism explains longevity, and
997 undefined terminology explains misunderstanding both. *PLoS genetics* **3**, e220,
998 doi:10.1371/journal.pgen.0030220 (2007).
999 58 Donohoe, D. R. & Bultman, S. J. Metaboloepigenetics: interrelationships between energy
1000 metabolism and epigenetic control of gene expression. *J Cell Physiol* **227**, 3169-3177,
1001 doi:10.1002/jcp.24054 (2012).
1002 59 Kucharski, R., Maleszka, J., Foret, S. & Maleszka, R. Nutritional control of reproductive
1003 status in honeybees via DNA methylation. *Science* **319**, 1827-1830,
1004 doi:10.1126/science.1153069 (2008).
1005 60 Dolinoy, D. C., Huang, D. & Jirtle, R. L. Maternal nutrient supplementation counteracts
1006 bisphenol A-induced DNA hypomethylation in early development. *Proceedings of the*
1007 *National Academy of Sciences of the United States of America* **104**, 13056-13061,
1008 doi:10.1073/pnas.0703739104 (2007).
1009 61 Holman, G. D. A new deadly Syn? *Curr Biol* **9**, R735-737, doi:10.1016/s0960-
1010 9822(99)80471-0 (1999).
1011 62 Michailidou, K. *et al.* Association analysis identifies 65 new breast cancer risk loci.
1012 *Nature* **551**, 92-94, doi:10.1038/nature24284 (2017).
1013 63 Rokudai, S. *et al.* STXBP4 regulates APC/C-mediated p63 turnover and drives squamous
1014 cell carcinogenesis. *Proceedings of the National Academy of Sciences of the United*
1015 *States of America* **115**, E4806-E4814, doi:10.1073/pnas.1718546115 (2018).
1016 64 Kichaev, G. *et al.* Leveraging Polygenic Functional Enrichment to Improve GWAS Power.
1017 *American journal of human genetics* **104**, 65-75, doi:10.1016/j.ajhg.2018.11.008 (2019).
1018 65 Perry, J. R. *et al.* Parent-of-origin-specific allelic associations among 106 genomic loci for
1019 age at menarche. *Nature* **514**, 92-97, doi:10.1038/nature13545 (2014).
1020 66 Albertsen, H. M. *et al.* A physical map and candidate genes in the BRCA1 region on
1021 chromosome 17q12-21. *Nature genetics* **7**, 472-479, doi:10.1038/ng0894-472 (1994).
1022 67 Kauraniemi, P. & Kallioniemi, A. Activation of multiple cancer-associated genes at the
1023 ERBB2 amplicon in breast cancer. *Endocr Relat Cancer* **13**, 39-49,
1024 doi:10.1677/erc.1.01147 (2006).
1025 68 Tanaka, S. *et al.* Coexpression of Grb7 with epidermal growth factor receptor or
1026 Her2/erbB2 in human advanced esophageal carcinoma. *Cancer research* **57**, 28-31
1027 (1997).
1028 69 de Vries, P. S. *et al.* Multiancestry Genome-Wide Association Study of Lipid Levels
1029 Incorporating Gene-Alcohol Interactions. *Am J Epidemiol* **188**, 1033-1054,
1030 doi:10.1093/aje/kwz005 (2019).

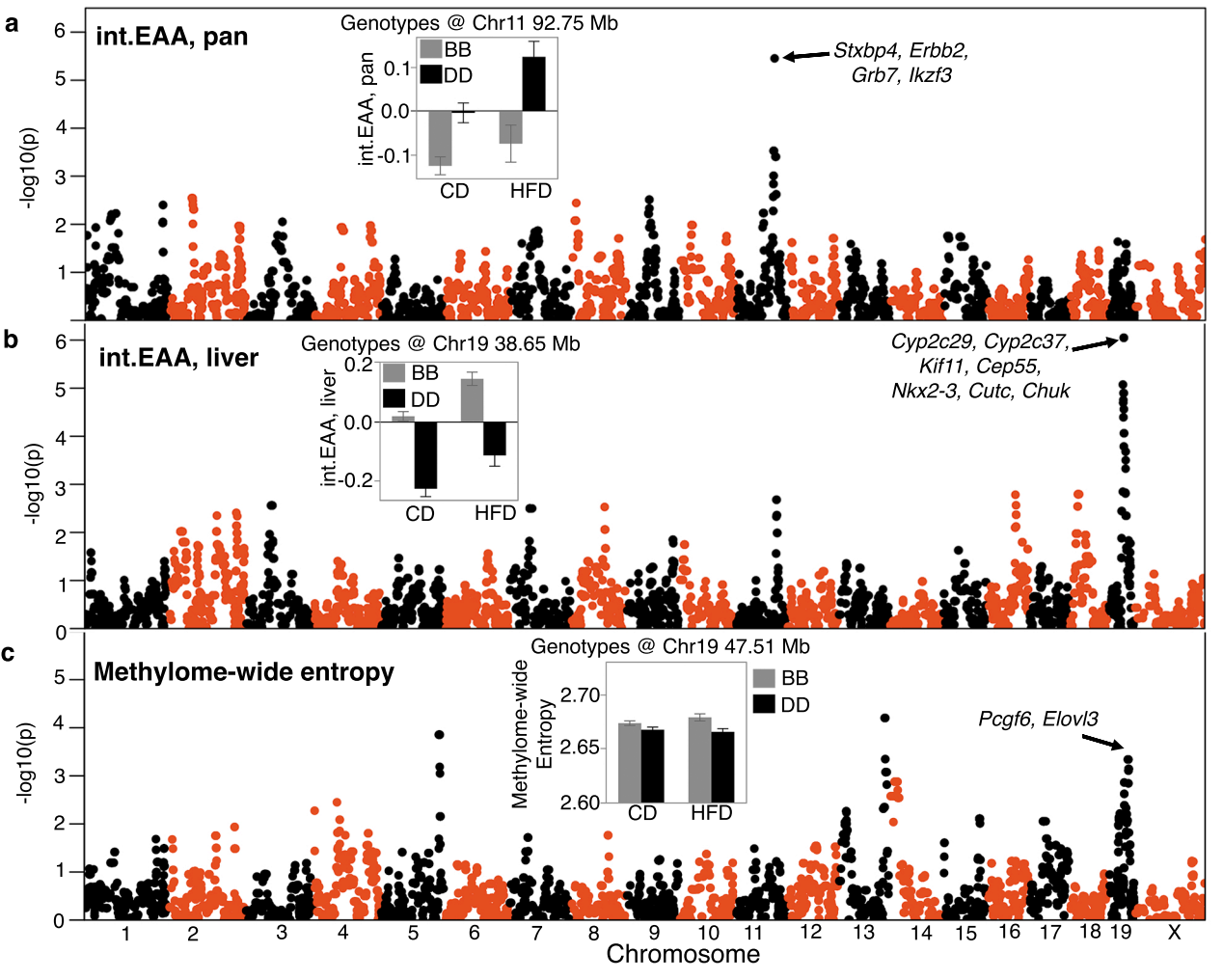
1031 70 Richardson, T. G. *et al.* Evaluating the relationship between circulating lipoprotein lipids
1032 and apolipoproteins with risk of coronary heart disease: A multivariable Mendelian
1033 randomisation analysis. *PLoS Med* **17**, e1003062, doi:10.1371/journal.pmed.1003062
1034 (2020).
1035 71 Xu, Y. *et al.* IKK interacts with rictor and regulates mTORC2. *Cell Signal* **25**, 2239-2245,
1036 doi:10.1016/j.cellsig.2013.07.008 (2013).
1037 72 Gallagher, M. D. & Chen-Plotkin, A. S. The Post-GWAS Era: From Association to Function.
1038 *American journal of human genetics* **102**, 717-730, doi:10.1016/j.ajhg.2018.04.002
1039 (2018).
1040 73 Lappalainen, T. Functional genomics bridges the gap between quantitative genetics and
1041 molecular biology. *Genome research* **25**, 1427-1431, doi:10.1101/gr.190983.115 (2015).
1042 74 Grodstein, F. *et al.* The association of epigenetic clocks in brain tissue with brain
1043 pathologies and common aging phenotypes. *Neurobiol Dis* **157**, 105428,
1044 doi:10.1016/j.nbd.2021.105428 (2021).
1045 75 Olona, A. *et al.* Epoxygenase inactivation exacerbates diet and aging-associated
1046 metabolic dysfunction resulting from impaired adipogenesis. *Mol Metab* **11**, 18-32,
1047 doi:10.1016/j.molmet.2018.03.003 (2018).
1048 76 Schuck, R. N. *et al.* The cytochrome P450 epoxygenase pathway regulates the hepatic
1049 inflammatory response in fatty liver disease. *PloS one* **9**, e110162,
1050 doi:10.1371/journal.pone.0110162 (2014).
1051 77 Wang, Q. *et al.* Time serial transcriptome reveals Cyp2c29 as a key gene in
1052 hepatocellular carcinoma development. *Cancer Biol Med* **17**, 401-417,
1053 doi:10.20892/j.issn.2095-3941.2019.0335 (2020).
1054 78 Mozhui, K. *et al.* Dissection of a QTL hotspot on mouse distal chromosome 1 that
1055 modulates neurobehavioral phenotypes and gene expression. *PLoS genetics* **4**,
1056 e1000260, doi:10.1371/journal.pgen.1000260 (2008).
1057 79 Zhou, W., Triche, T. J., Jr., Laird, P. W. & Shen, H. SeSAMe: reducing artifactual detection
1058 of DNA methylation by Infinium BeadChips in genomic deletions. *Nucleic acids research*
1059 **46**, e123, doi:10.1093/nar/gky691 (2018).
1060 80 Friedman, J., Hastie, T. & Tibshirani, R. Regularization Paths for Generalized Linear
1061 Models via Coordinate Descent. *J Stat Softw* **33**, 1-22 (2010).
1062 81 Coschigano, K. T. *et al.* Deletion, but not antagonism, of the mouse growth hormone
1063 receptor results in severely decreased body weights, insulin, and insulin-like growth
1064 factor I levels and increased life span. *Endocrinology* **144**, 3799-3810,
1065 doi:10.1210/en.2003-0374 (2003).
1066 82 Hausser, J. & Strimmer, K. Entropy Inference and the James-Stein Estimator, with
1067 Application to Nonlinear Gene Association Networks. *J. Mach. Learn. Res.* **10**, 1469–
1068 1484 (2009).
1069 83 *GeneNetwork 2*, <http://genenetwork.org> (
1070 84 Mulligan, M. K., Mozhui, K., Prins, P. & Williams, R. W. GeneNetwork: A Toolbox for
1071 Systems Genetics. *Methods in molecular biology* **1488**, 75-120, doi:10.1007/978-1-4939-
1072 6427-7_4 (2017).

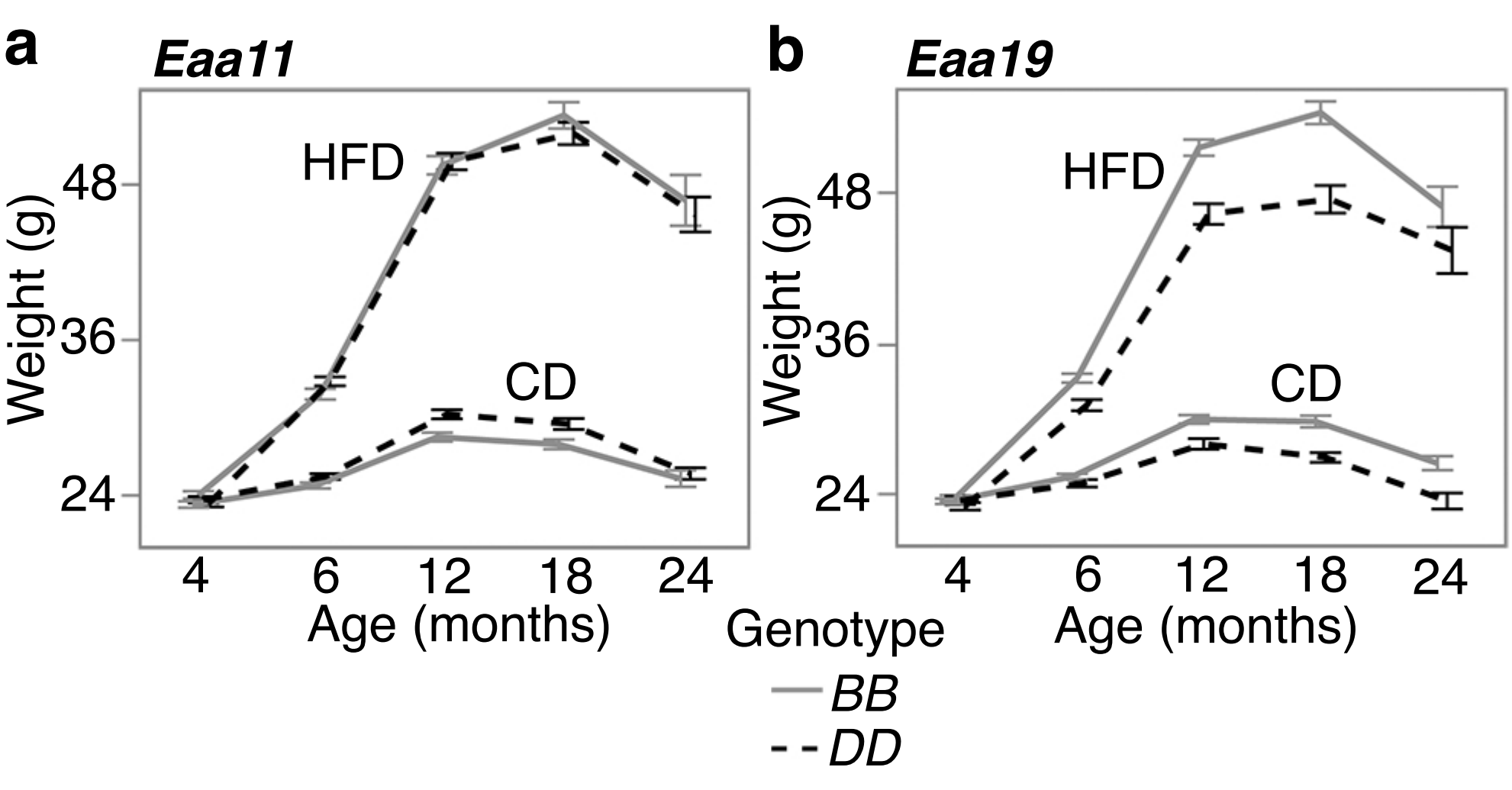
85 McClay, J. L. *et al.* A methylome-wide study of aging using massively parallel sequencing of the methyl-CpG-enriched genomic fraction from blood in over 700 subjects. *Human molecular genetics* **23**, 1175-1185, doi:10.1093/hmg/ddt511 (2014).
86 Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842, doi:10.1093/bioinformatics/btq033 (2010).
87 Ashbrook, D. G. *et al.* Born to Cry: A Genetic Dissection of Infant Vocalization. *Front Behav Neurosci* **12**, 250, doi:10.3389/fnbeh.2018.00250 (2018).
88 Belknap, J. K. Effect of within-strain sample size on QTL detection and mapping using recombinant inbred mouse strains. *Behavior genetics* **28**, 29-38 (1998).
89 Sherry, S. T. *et al.* dbSNP: the NCBI database of genetic variation. *Nucleic acids research* **29**, 308-311, doi:10.1093/nar/29.1.308 (2001).
90 Bates, D., Maechler, M., Bolker, B. & Walker, S. lme4: Linear mixed-effects models using Eigen and S4. . *R package version 1.1-7, http://CRAN.R-project.org/package=lme4*.
91 *Wellcome Sanger Institute Mouse Genome Project*, <https://www.sanger.ac.uk/sanger/Mouse_SnpViewer/rel-1505> (
92 Keane, T. M. *et al.* Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature* **477**, 289-294, doi:10.1038/nature10413 (2011).
93 Yalcin, B. *et al.* Sequence-based characterization of structural variation in the mouse genome. *Nature* **477**, 326-329, doi:10.1038/nature10432 (2011).
94 Huang da, W., Sherman, B. T. & Lempicki, R. A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic acids research* **37**, 1-13, doi:10.1093/nar/gkn923 (2009).
95 Huang da, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols* **4**, 44-57, doi:10.1038/nprot.2008.211 (2009).
96 Masuda, T., Tomita, M. & Ishihama, Y. Phase transfer surfactant-aided trypsin digestion for membrane proteome analysis. *J Proteome Res* **7**, 731-740, doi:10.1021/pr700658q (2008).
97 Mostafa, D. *et al.* Loss of beta-cell identity and diabetic phenotype in mice caused by disruption of CNOT3-dependent mRNA deadenylation. *Commun Biol* **3**, 476, doi:10.1038/s42003-020-01201-y (2020).
98 Reiter, L. *et al.* mProphet: automated data processing and statistical validation for large-scale SRM experiments. *Nature methods* **8**, 430-435, doi:10.1038/nmeth.1584 (2011).
99 Rosenberger, G. *et al.* Statistical control of peptide and protein error rates in large-scale targeted data-independent acquisition analyses. *Nature methods* **14**, 921-927, doi:10.1038/nmeth.4398 (2017).

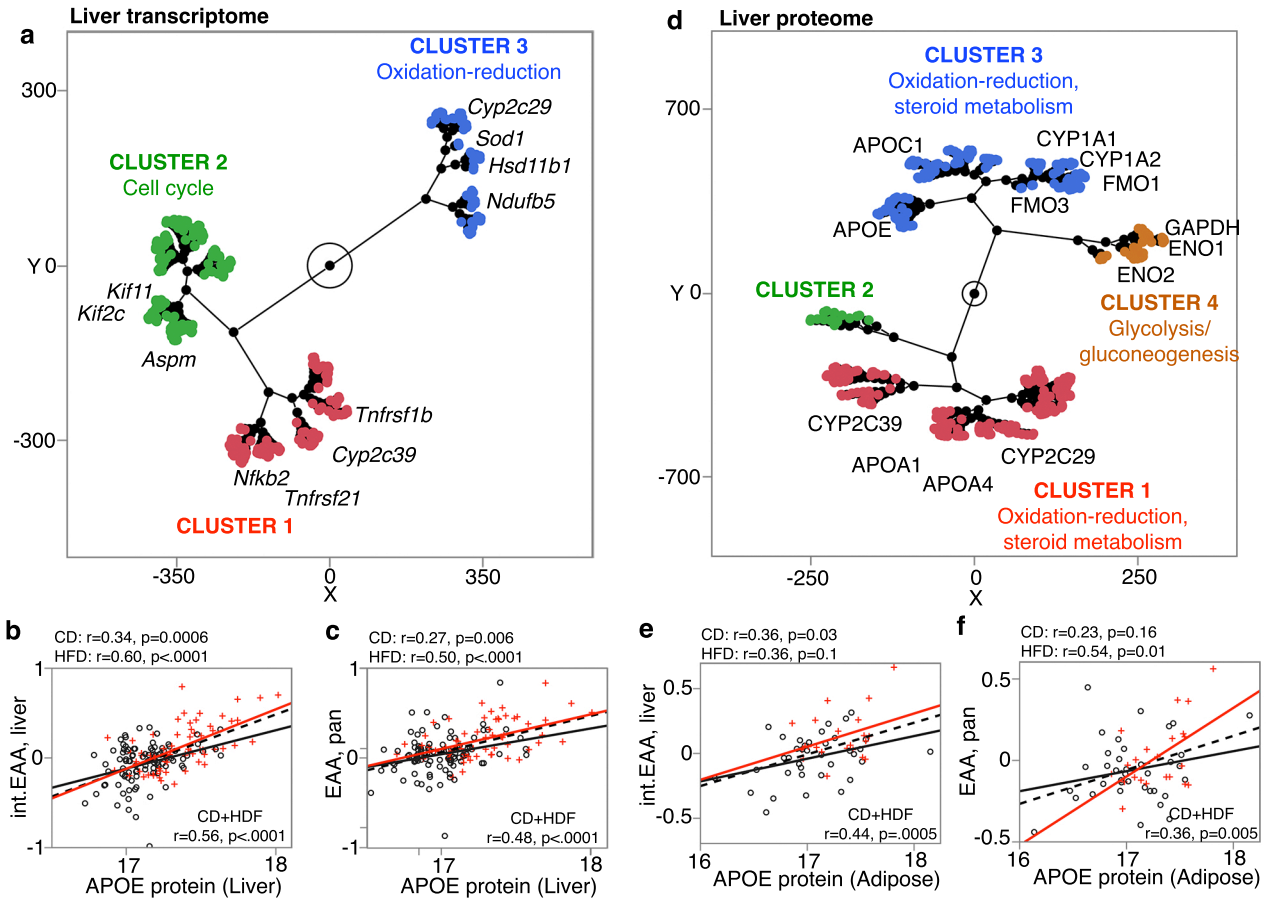

a
CD: r=0.86; HFD: r=0.91
int.DNAmAge, liver (years)
Age (years)
CD+HFD
r=0.86
Diet
CD
HFD
Linear fit
Both
CD
HFD
b
EAA.pan
EAA.liver
dev.EAA.pan
dev.EAA.liver
int.EAA.pan
int.EAA.liver
Epigenetic age acceleration
c
CD: r=0.24, p=0.0004
HFD: r=0.53, p<.0001
DNAm entropy
Age (years)
CD+HFD
r=0.35
d
CD: r=0.25, p=0.0002
HFD: r=0.49, p<.0001
int.EAA, liver
DNAm entropy
CD+HFD
r=0.36, p<.0001
e
r=0.58; p<.0001
int.EAA. liver
BW0 (g)
Days on HFD
1
3
f
CD: r=0.15, p=0.04
HFD: r=0.35, p<.0001
int.EAA, liver
deltaBW (g)
CD+HFD
r=0.39, p<.0001
g
int.EAA, pan
Females
Males
CD
HFD
Diet
h
EAA, pan
EAA, liver
dev.EAA, pan
dev.EAA, liver
int.EAA, pan
int.EAA, liver
Variance source for EAA
Entropy
BWF
Diet
Chol
Gluc
Age
LogWorth effect
i
p<.0001
R²=0.17
int.EAA, pan (residuals)
DNAm entropy (residuals)


a
1523
5030
-Log10p (age)
Age β estimate
Age-gain
Age-loss
b
723 neg
10 pos
-Log10p (weight)
Weight β estimate
c
321 pos
-Log10p (diet)
Diet β estimate
d
124 neg
112 pos
-Log10p (lifespan)
Lifespan β estimate
e
Age
LS
Weight
Diet
147
282
50
16
123
6151
7
5
56
211
7
3
82
1
44
f
Mean methylation
Age-gain
Age-loss
g
Mettl23; 3UTR CpG
Weight (g)
h
Diet β estimate
Age β estimate
i
Mettl23; 3UTR CpG
Age (days)
Diet
CD
HFD
Density of age-CpGs
j
Lifespan β estimate
Age β estimate


a
CD: r=0.59, p<.0001
HFD: r=0.76, p<.0001
Entropy age-gain
Age (years)
Diet
CD
HFD
b
CD: r=0.58, p<.0001
HFD: r=0.76, p<.0001
Entropy age-loss
Age (years)
c
CD: r=0
HFD: r=0.31, p=0.0004
Entropy age-ns
Age (years)
d
UT319 (0.6 y)
UT573 (2.3 y)
Density
Beta-values
e
p = 0.001
Entropy age-gain
CD
HFD
f
CD: r=0.31, p<.0001
HFD: r=0.31, p=0.0004
Entropy age-loss
BW0 (g)
g
Age-gain entropy
Age-loss entropy
Variance source
Age
Diet
Chol
BWF
Gluc
LS
LogWorth effect
h
p<0.0001
R²=0.06
Entropy age-gain (residuals)
Median lifespan (residuals)
i
p=0.0002
R²=0.05
Entropy age-loss (residuals)
BWF (residuals)
j
p=0.0004
R²=0.05
Entropy age-gain (residuals)
Serum glucose (residuals)


a
int.EAA, pan
Genotypes @ Chr11 92.75 Mb
BB
DD
int.EAA, pan
0.1
0.0
-0.1
CD
HFD
Stxbp4, Erbb2,
Grb7, Ikzf3
-log10(p)
b
int.EAA, liver
Genotypes @ Chr19 38.65 Mb
BB
DD
int.EAA, liver
0.2
0.0
-0.2
CD
HFD
Cyp2c29, Cyp2c37,
Kif11, Cep55,
Nkx2-3, Cutc, Chuk
-log10(p)
c
Methylome-wide entropy
Genotypes @ Chr19 47.51 Mb
BB
DD
Methylome-wide Entropy
2.70
2.65
2.60
CD
HFD
Pcgf6, Elovl3
-log10(p)
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 X
Chromosome


a
Eaa11
HFD
CD
Weight (g)
48
36
24
4
6
12
18
24
Age (months)
b
Eaa19
HFD
CD
Weight (g)
48
36
24
4
6
12
18
24
Age (months)
Genotype
BB
DD


a
Liver transcriptome
CLUSTER 3
Oxidation-reduction
Cyp2c29
Sod1
Hsd11b1
Ndufb5
CLUSTER 2
Cell cycle
Kif11
Kif2c
Aspm
Tnfrsf1b
Cyp2c39
Nfkb2
Tnfrsf21
CLUSTER 1
300
Y 0
-300
-350
0
350
X
d
Liver proteome
CLUSTER 3
Oxidation-reduction,
steroid metabolism
APOC1
CYP1A1
CYP1A2
FMO1
FMO3
APOE
GAPDH
ENO1
ENO2
CLUSTER 4
Glycolysis/
gluconeogenesis
CLUSTER 2
CYP2C39
APOA1
APOA4
CYP2C29
CLUSTER 1
Oxidation-reduction,
steroid metabolism
700
Y 0
-700
-250
0
250
X
b
CD: r=0.34, p=0.0006
HFD: r=0.60, p<.0001
CD+HDF
r=0.56, p<.0001
int.EAA, liver
APOE protein (Liver)
c
CD: r=0.27, p=0.006
HFD: r=0.50, p<.0001
CD+HDF
r=0.48, p<.0001
EAA, pan
APOE protein (Liver)
e
CD: r=0.36, p=0.03
HFD: r=0.36, p=0.1
CD+HDF
r=0.44, p=.0005
int.EAA, liver
APOE protein (Adipose)
f
CD: r=0.23, p=0.16
HFD: r=0.54, p=0.01
CD+HDF
r=0.36, p=0.005
EAA, pan
APOE protein (Adipose)