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]

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]

The DNAm 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 **Supplementary file 1**.

DNAm clocks, entropy, and chronological age prediction

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

We next estimated the methylome-wide entropy as a measure of randomness and information loss. This was computed from 27966 probes that provide high-quality data and have been validated to perform well in mice.[29] Consistent with previous reports,[10,43] this property increased with chronological age, and age accounted for about 6% (in CD) to 28% (in HFD) of the variance in entropy (**Figure 1c**). As direct correlates of chronological age, all the DNAmAge estimates also had significant positive correlations with entropy (**Table 1**). We hypothesized that higher entropy levels will be associated with higher EAA, and based on this bivariate comparison, most of the EAA showed a significant positive correlation with entropy (**Figure 1d; Supplementary file 3**).

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

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

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

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

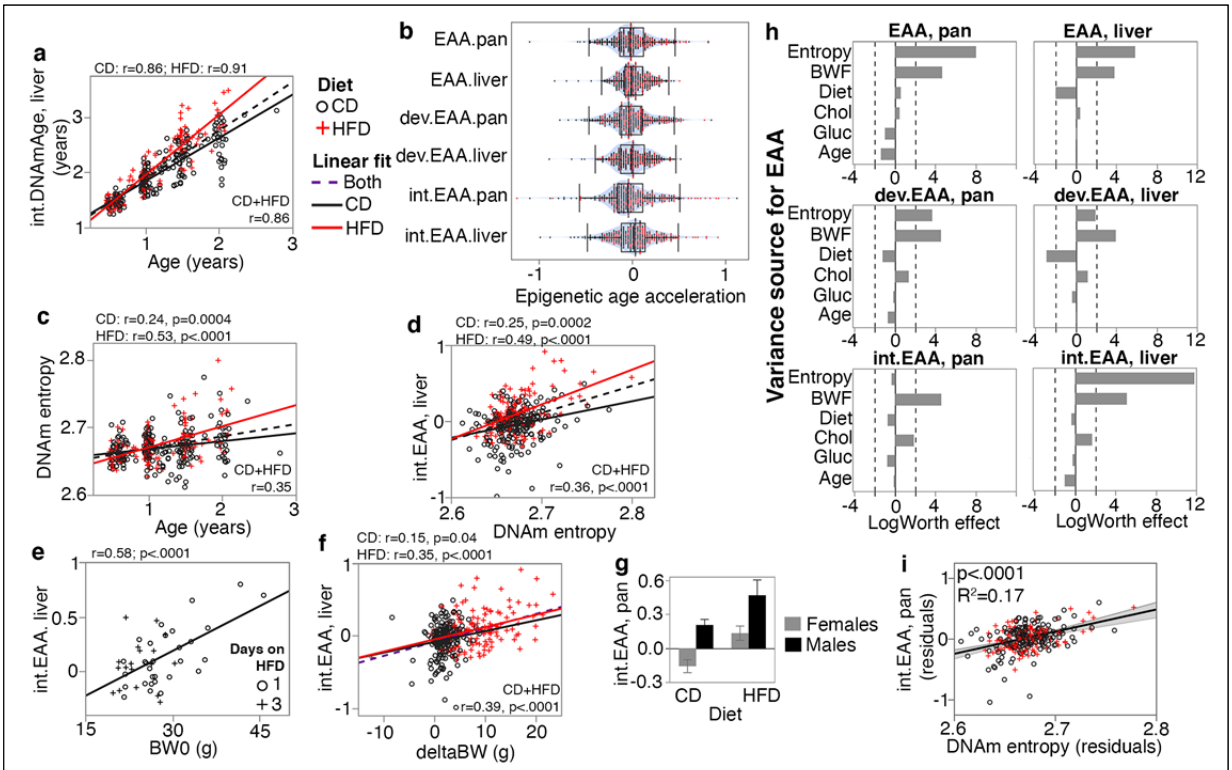


Figure 1. Correlates and modifiers of epigenetic clocks and methylome-wide entropy **(a)** Correlation between chronological age and predicted age (shown for the liver intervention clock or int.DNAmAge). Black circles are control diet (CD, n = 210); red crosses are high fat diet mice (HFD, n = 128) **(b)** 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.

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

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	

[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 CD and 128 HFD)

[b] Pearson correlation between strain means for n = 29 BXD genotypes kept on CD and HFD

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

Sex. Four BXD genotypes (B6D2F1, D2B6F1, BXD102, B6) had cases from both males and females. We used these to test for sex effects. All the clocks showed significant age acceleration in male mice, and this effect was particularly strong for the both dev.EAA, and the pan-tissue int.EAA (**Figure 1g; Supplementary file 3**). This effect was independent of the higher BWF of males, and the higher age-acceleration in males was detected after adjusting for BWF (**Supplementary file 4a**). There was no significant difference in entropy between the sexes.

Metabolic measures. 278 cases with DNAm data also had fasted serum glucose and total cholesterol,[32,33] and we examined whether these metabolic traits were associated with either the EAA measures or methylome entropy. Since these are highly dependent on diet, weight, and age, we applied a multivariable model to jointly examine how the different metabolic variables (cholesterol and glucose, as well as diet and weight) and entropy relate to EAA after adjusting for age. To test the robustness of associations, we also include the methylation array batch as another covariate (**Supplementary file 1** has batch information; **Supplementary file 5** has the full statistics). **Figure 1h** shows the relative strengths and directions of associations between these variables and the EAA traits. Except for the pan-tissue interventional clock, entropy had a strong positive association with EAA. For example, a plot of residuals between entropy and the liver int.EAA indicates that after adjusting for all the other covariates, the methylome-wide entropy explains 17% of the variance in int.EAA (**Figure 1i**). Since diet strongly influences BWF, the inclusion of BWF in the regression diminished the effect of diet. For the two clocks that were not influenced by diet (the liver EAA and liver dev.EAA), adjusting for the effect of BWF resulted in an inverse association with diet (i.e., the residual EAA values after accounting for BWF were slightly lower in the HFD group). Fasted glucose did not have a significant effect on EAA. Cholesterol had a positive association with the interventional clocks but the effects were modest (residual $R^2 = 0.02$ and $p = 0.02$ for cholesterol and the pan-tissue int.EAA).

We also performed a similar analysis with BW0 instead of BWF (**Figure 1-figure supplement 1**), and here, HFD remained as an accelerator of the pan-tissue EAA and liver int.EAA. Cholesterol

also became a significant positive correlate of EAA for the interventional clocks. This would suggest that the effect of diet on EAA is mostly mediated by its impact on physiological and metabolic traits, and BWF becomes a prominent predictor of EAA.

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

How the epigenetic readouts relate to strain longevity

We next obtained longevity data from a parallel cohort of female BXD mice that were allowed to age on CD or HFD.[33] Since the strain lifespan was determined from female BXDs, we restricted this to only the female cases. For strains with natural death data from $n \geq 5$, we computed the minimum (minLS), 25th quartile (25Q-LS), mean, median lifespan, 75th quartile (75Q-LS), and maximum lifespan (maxLS) (**Supplementary file 1**). Specifically, we postulated an accelerated clock for strains with shorter lifespan (i.e., inverse correlation). Overall, the EAA measures showed the expected inverse correlation trend with the lifespan statistics (**Supplementary file 4b**). However, these correlations were weak. The correlations were significant only for the pan-tissue general clock (**Figure 1-figure supplement 2a**) and the liver intervention clock, with explained variance in lifespan of only ~3% (**Figure 1-figure supplement 2b, 2c**). When separated by diet, these correlations became weaker indicating that while we see the expected inverse relationship, the EAA is only weakly predictive of strain longevity. Entropy estimated from the full set of CpGs was unrelated to strain longevity.

Multifactor variance of the conserved CpGs

Both the entropy and clock readouts capture the overall variation across multiple CpGs, and to gain insights into the underlying variance patterns, we performed a multivariable epigenome-wide association study (EWAS). For this, we applied a site-by-site regression on the 27966 validated CpGs,[29] and tested for association with age, BWF, diet, and genotype-dependent strain median lifespan (full set of probes, annotations, and EWAS results in **Supplementary file 6**).

Age was clearly the most influential variable, and this is apparent from the volcano plots (**Figure 2a–d**). We used a cutoff of Bonferroni $p \leq 0.05$ to define differentially methylated CpGs (DMCs), and 6553 CpGs were associated with age (referred to as age-DMCs), 733 with weight (weight-DMCs), 321 with diet (diet-DMCs), and 236 with genotype-dependent lifespan (LS-DMCs). We note extensive overlap among the lists of DMCs that shows that variation at these CpGs are

multifactorial in nature (**Figure 1e**). Majority of the age-DMCs (77%) gained methylation (or age-gain), and consistent with previous observations, age-gain CpGs tended to be in regions with low methylation, whereas age-DMCs that declined in methylation (age-loss) were in regions with high methylation (**Figure 2f**).[39,43,44] By overlaying the volcano plots with the age-gain and age-loss information, we see distinct patterns in how these age-DMCs vary with

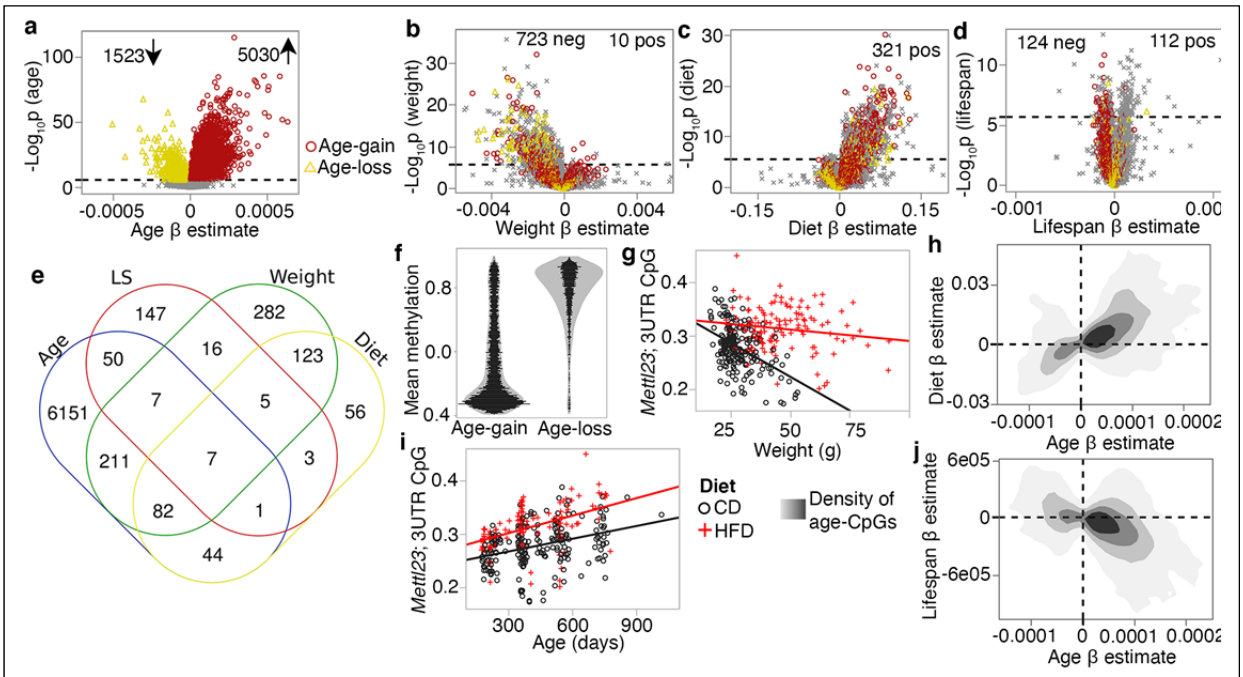


Figure 2. Multivariable analysis of site-specific methylation

(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.

weight (**Figure 2b**), diet (**Figure 2c**), and genotype lifespan (**Figure 2d**). While the majority of CpGs, including several age-loss CpGs, had negative regression estimates for weight (i.e., decrease in DNAm with unit increase in weight), HFD was associated with higher methylation levels (positive regression estimates) including at several age-DMCs (**Figure 2c**). This pattern of inverse correlation with weight but heightened methylation due to HFD is illustrated by the CpG

in the 3'UTR of *Mettl23* (cg10587537) (**Figure 2g**). Taking the 6553 age-DMCs, a comparison of the regression estimates for 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 (**Figure 2h**), and again, this is illustrated by the CpG in the *Mettl23* 3'UTR (**Figure 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 (**Figure 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 (**Figure 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 age-DMCs). For these conserved CpGs, both sets of age-DMCs had significant increases in entropy with age regardless of diet (**Figure 3a**, **3b**), and even the age-ns showed a modest entropy gain with age in the HFD group (**Figure 3c**). The reason for this increase in disorder becomes evident when we compare the density plots using the full set of CpGs for one of the younger mice (UT319; 0.56 years old) and one of the older mice (UT573; 2.3 years) (**Figure 3d**). Concordant with previous reports,[43,45] the older sample showed a subtle flattening of the bimodal peaks towards a slightly more hemi-methylated state. The entropy of the age-gain CpGs was modestly but significantly higher in the HFD group ($p = 0.001$; **Figure 3e**). Entropy of the age-loss and age-ns CpGs were not different between the diets. Body weight on the other hand, was associated specifically with the entropy score of the age-loss CpGs, and both higher BW0 (**Figure 3f**) and BWF predicted higher entropy for age-loss CpGs.

We applied a multivariable regression to compare the relative effects of age, diet, BWF, glucose, cholesterol, and strain median lifespan (**Figure 3g**; full statistics in **Supplementary file 7**). Entropy of age-gain CpGs was increased by HFD but was not associated with BWF. Strain median lifespan showed a significant inverse correlation with the entropy of age-gain CpGs with an explained variance of 6% (**Figure 3h**). Entropy of the age-loss CpGs had a significant positive correlation with BWF (**Figure 3i**), but was not associated with diet, and also had a modestly significant inverse correlation with median lifespan. Cholesterol was unrelated to the entropy values. Glucose on the other hand, showed a significant inverse association with entropy of both the age-gain (**Figure 3j**) and age-loss CpGs, and this suggests slightly lower entropy with higher fasted glucose.

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

Functional and genomic context of DMCs

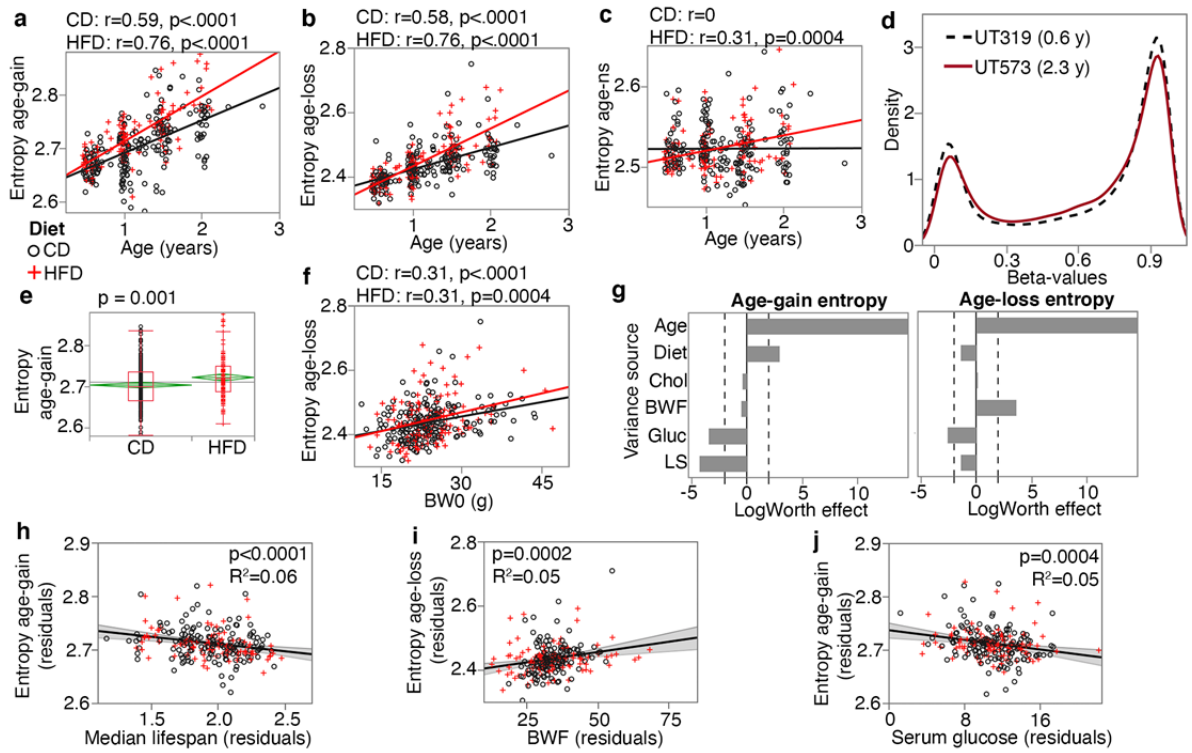


Figure 3. Entropy at age-associated CpGs

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.

To uncover the potential biological pathways represented by the DMCs, we performed genomic regions enrichment analyses for the CpGs.[46] The age-gain CpGs were highly enriched in transcription factors, regulators of development and growth, menarche and menstrual phases, energy metabolism, and transcription factor networks such as HNF1 and HNF3B pathways (**Supplementary file 8**). The age-loss CpGs had somewhat modest enrichment, and represented

cell adhesion and cytoskeletal processes, endothelial cell proliferation, and p38 signaling. The BW-DMCs were enriched in actin and protein metabolism, and WNT, and platelet-derived growth factor (PDGF) and ErbB signaling. Similarly, the diet-DMCs were highly enriched in PDGF, epidermal growth factor (EGFR) and ErbB signaling, as well as the mTOR signaling pathway, and regulation of energy homeostasis (**Supplementary file 8**). Seeming to converge on common pathways, the LS-CpGs that were negatively correlated with lifespan had modest enrichment in cell signaling pathways such as EGFR, PDGF, and ErbB signaling. The LS-CpGs with positive correlation with lifespan were highly enriched in lipid metabolic genes, and also included pathways related to chromosome maintenance and telomere expansion (**Supplementary file 8**).

We next examined the genomic annotations and chromatin states of the DMCs. Consistent with previous reports,[39,43] age-gain CpGs were enriched in promoter and 5'UTR CpGs, but depleted in 3'UTR, exon, and intergenic CpGs (**Figure 2-figure supplement 1a**; **Supplementary file 9**). Diet- and weight-DMCs were depleted in promoter regions, and enriched in exons and 3'UTR, and along with the age-loss CpGs, enriched in introns. For chromatin states, we annotated the CpG regions using the 15-states chromatin data for neonatal (P0) mouse liver.[47,48] Also included were regions labelled as No Reproducible State or NRS; i.e., regions that were not replicated (**Supplementary file 6** has annotations for each site).[48] Compared to the array content as background, the age-gain CpGs were selectively enrich in polycomb associated heterochromatin (Hc-P) and bivalent promoters (Pr-Bi), chromatin states that were highly depleted among the other DMCs (**Figure 2-figure supplement 1b**). In contrast, strong and permissive transcription sites (Tr-S and Tr-P, respectively) were depleted among the age-gain CpGs, and enriched among the BW- and diet-DMCs. Age-loss CpGs were enriched in Tr-P and Tr-I (transcription initiation). Distal enhancers (strong distal or En-Sd, and poised distal or En-Pd) were also highly enriched among the BW- and diet-DMCs, and also showed some enrichment among the age-DMCs.

For an overview of the general methylation and variance patterns by chromatin annotations, we used the full set of 27966 CpGs and computed the average methylation beta-values, and average regression coefficients (i.e, change in beta-value per unit change in the respective predictor variable, or contrast between diets). As expected, promoter CpGs and Hc-P were sites with the lowest methylation. Hc-H, Tr-S, and Tr-P had higher methylation, and many of the enhancer sites were in the hemi-methylated zone. For age effect, mean regression estimates had a significant inverse linear fit with mean methylation ($r = –0.63$, $p = 0.009$; **Figure 2-figure supplement 1c**) and this is consistent with the greater age-loss at hypermethylated CpGs, and greater age-gains at hypomethylated CpGs (**Figure 2f**). The effects of diet and weight were not linearly related to the mean methylation of the chromatin states. Instead, both showed a U-shaped fit with a significant negative quadratic effect for diet ($R^2 = 0.69$, $p = 0.0005$, quadratic estimate = –0.05; **Figure 2-figure supplement 1d**), and a positive quadratic effect for weight ($R^2 = 0.50$, $p = 0.01$, quadratic estimate = 0.001; **Figure 2-figure supplement 1e**). Methylation variation as a function on strain longevity did not relate to mean methylation with either a linear or polynomial fit, and indicates that variance due to background genotype is less dependent on the chromatin and mean methylation status. While this is a very low-resolution and broad view of methylation levels and methylation variation, the observations show that

while aging results in erosion of the hypo- and hypermethylated peaks, diet and body weight appear to have generally stronger associations with hemi-methylated sites.

Genetic analysis of epigenetic age acceleration

The EAA traits had moderate heritability at an averaged h^2 of 0.50 (**Table 2**).[34] Another way to gauge level of genetic correlation is to compare between members of strains maintained on different diets. All the EAA traits shared high strain-level correlations between diets, indicating 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.

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

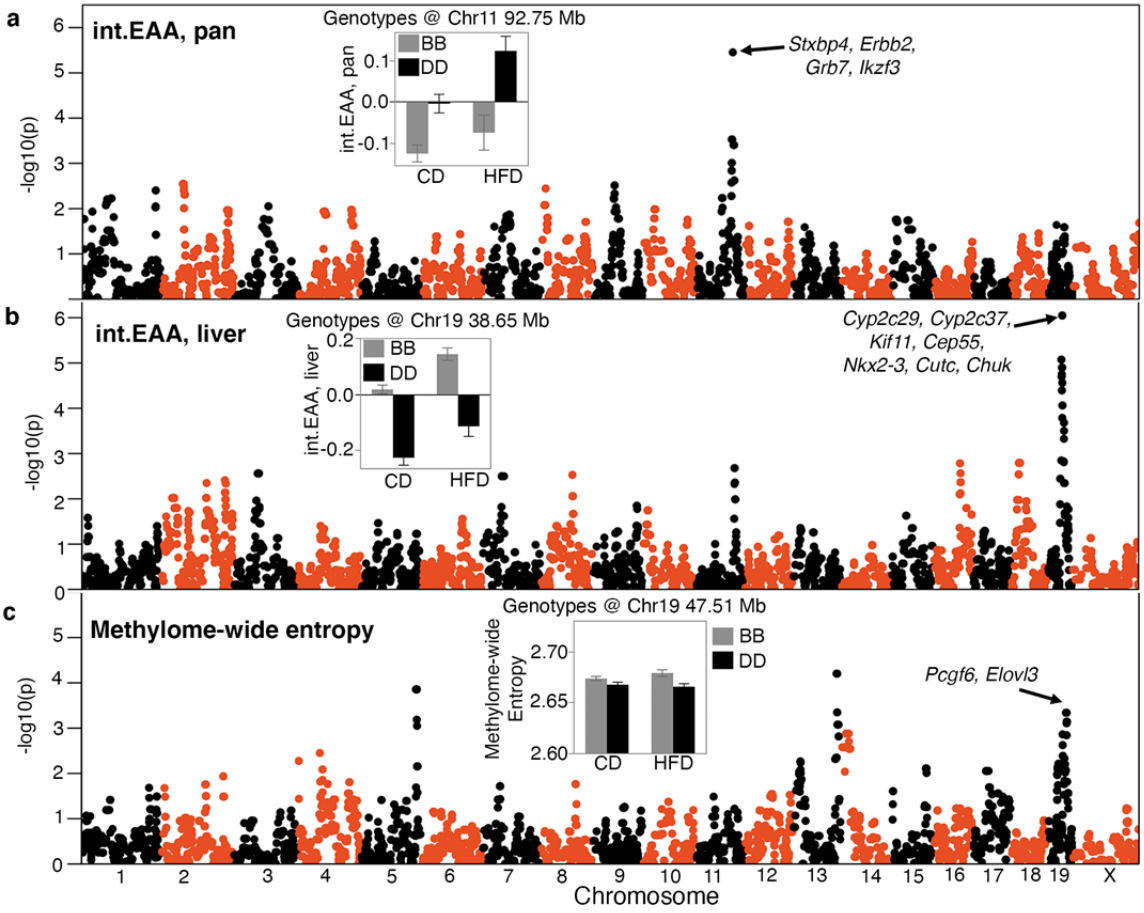


Figure 4. QTL maps for the DNAm readouts

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.

adjustment for diet and body weight. EAA from the two interventional clocks had the strongest QTLs (**Supplementary file 10**). 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) (**Figure 4a**). Taking a genotype marker at the peak interval (BXD variant ID DA0014408.4 at Chr11, 92.750 Mb)[34], we segregated the BXDs homozygous for either the D2 (*DD*) or the B6 (*BB*) alleles. Strains with *DD* genotype at this locus had significantly higher int.EAA (**Figure 4a** inset). The liver int.EAA had the highest QTL on Chr19 (35–45 Mb) with the most significant linkage at markers between 38–42 Mb (p = 9E-07; LOD score of 5.2) (**Figure 4b**). We selected a marker at the peak interval (rs48062674 at Chr19, 38.650 Mb), and the *BB* genotype had significantly higher liver int.EAA compared to *DD* (**Figure 4b** inset).

We performed a similar QTL mapping for methylome-wide entropy with adjustment for major covariates (diet, chronological age, and body weight). There were no genome-wide significant QTLs. A region on Chr19 that overlapped the liver int.EAA showed a modest peak (**Figure 4c**; **Supplementary file 10**). However, the peak markers for entropy were located slightly distal to the peak EAA QTL (~47.5 Mb at rs30567369, minimum p = 0.0005). At this locus, the *BB* genotype had higher average entropy.

To identify regulatory loci that are consistent across the different EAA measures, we applied a multi-trait analysis and derived the linkage meta-p-value using a p-value combination for the six EAA traits.[50] The peaks on Chrs 11 and 19 attained the highest consensus p-values (**Figure 4-figure supplement 1a**). There was another potential consensus peak at combined $-\log_{10}p > 6$ on Chr3 (~54 Mb). We focus on the Chrs 11 and 19 QTLs and refer to these as *EAA QTL on Chr 11* (Eaa11), and EAA QTL on Chr 19 (Eaa19). Eaa11 extends from 90–99 Mb. For Eaa19, we delineated a broader interval from 35–48 Mb that also encompasses the peak markers for entropy.

We performed marker-specific linkage analyses for each of the clocks using a regression model that adjusted for diet. With the exception of the liver int.EAA, all the EAA traits had nominal to highly significant associations with the representative Eaa11 marker (DA0014408.4), and the *DD* genotype had higher age acceleration (**Table 3**). Mean plots by genotype and diet shows that this effect was primarily in the CD mice (**Figure 4-figure supplement 1b**). The effect of this locus appeared to be higher for the pan-tissue clocks compared to the corresponding liver-specific clocks. For proximal Eaa19, the representative marker (rs48062674) was associated with all the EAA traits and the *BB* mice had higher age acceleration on both diets (**Figure 4-figure supplement 1c**). We also tested if these peak markers were associated with the recorded lifespan phenotype and we found no significant association with the observed lifespan of the BXDs.

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

		Linear regression[1]			
Predictor	**Outcome**	**Estimate**	**Std Error**	**t Ratio**	**p**
Eaa11 DA0014408.4[DD] Chr11, 92.750 Mb	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

(133 *BB* cases, and 173 *DD* cases)	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 (238 *BB* cases, and 67 *DD* cases)	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
	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; **Figure 5a**). However, this marker had no significant association with body weight when tested using a mixed effects model (p

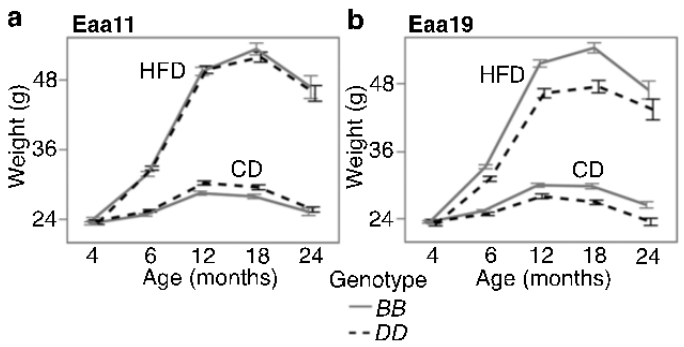


Figure 5. Body weight trajectory by diet and genotype

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

= 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 (**Figure 5b**), and this locus had a significant influence on the body weight trajectory (p = 7.6E-07; **Table 3**).

Candidate genes for epigenetic age acceleration

There are several positional candidate genes in Eaa11 and Eaa19. To narrow the list, we applied two selection criteria: genes that (1) contain missense and/or stop variants, and/or (2) contain non-coding variants and regulated by cis-acting expression QTLs (eQTL). For the eQTL analysis, we utilized an existing liver transcriptome data from the same aging cohort.[32] We identified 24 positional candidates in Eaa11 that includes *Stxbp4*, *Erbb2* (*Her-2* oncogenic gene), and *Grb7* (growth factor receptor binding) (**Supplementary file 11; Figure 4a**). Eaa19 has 81 such candidates that includes a cluster of cytochrome P450 genes, and *Chuk* (inhibitor of NF-kB) in the proximal region, and *Pcgf6* (epigenetic regulator) and *Elovl3* (lipid metabolic gene) in the distal region (**Supplementary file 11; Figure 4b, 4c**).

For further prioritization, we converted the mouse QTL regions to the corresponding syntenic regions in the human genome, and retrieved GWAS annotations for these intervals.[51] We specifically searched for the traits: epigenetic aging, longevity, age of menarche/menopause/puberty, Alzheimer's disease, and age-related cognitive decline and dementia. This highlighted 5 genes in Eaa11, and 3 genes in Eaa19 (**Supplementary file 4c**). We also identified a GWAS study that found associations between variants near *Myof-Cyp26a1* and human longevity,[41] and a meta-GWAS that found gene-level associations between *Nkx2–3* and *Cutc*, and epigenetic aging (**Supplementary file 4c**).[27]

Gene expression correlates of EAA

A subset of the BXD cases had liver RNA-seq data (94 CD, and 59 HFD).[32] Using this set, we performed transcriptome-wide correlation analysis for the general pan-tissue EAA, and the more specific liver int.EAA. To gain insights into biological pathways, we selected the top 2000 transcriptome correlates for functional enrichment analysis (**Supplementary file 12a**). The common themes for both clocks were: (1) there were far fewer negative correlates (223 out of 2000 for pan-tissue EAA, and 337 out of 2000 transcripts for liver int.EAA) than positive correlates, (2) the negative correlates were highly enriched (Bonferroni correct $p < 0.05$) in oxidation-reduction and mitochondrial genes (**Supplementary file 12b, 12c**). The pan-tissue general clock was also highly enriched in pathways related to steroid metabolism, epoxygenase p450 pathway, and xenobiotics, which are pathways that are particularly relevant to liver. The p450 genes included candidates that are in Eaa19 (e.g., *Cyp2c29*, *Cyp2c37*). The positive correlates were enriched in a variety of gene functions including mitosis for both clocks, and immune and inflammatory response for the general pan-tissue clock (functions that are not specific to liver). 563 transcripts (315 unique genes) were correlated with both the pan-tissue EAA, and the liver int.EAA. Based on hierarchical clustering (HC) of these common mRNA correlates of EAA, the transcripts could be clustered into 3 groups (**Figure 6a;** heatmap in **Figure 6-figure supplement 1a**). While none of these were significantly enriched in any particular gene ontology (GO), cluster 3 included several oxidation-reduction genes including the Eaa11

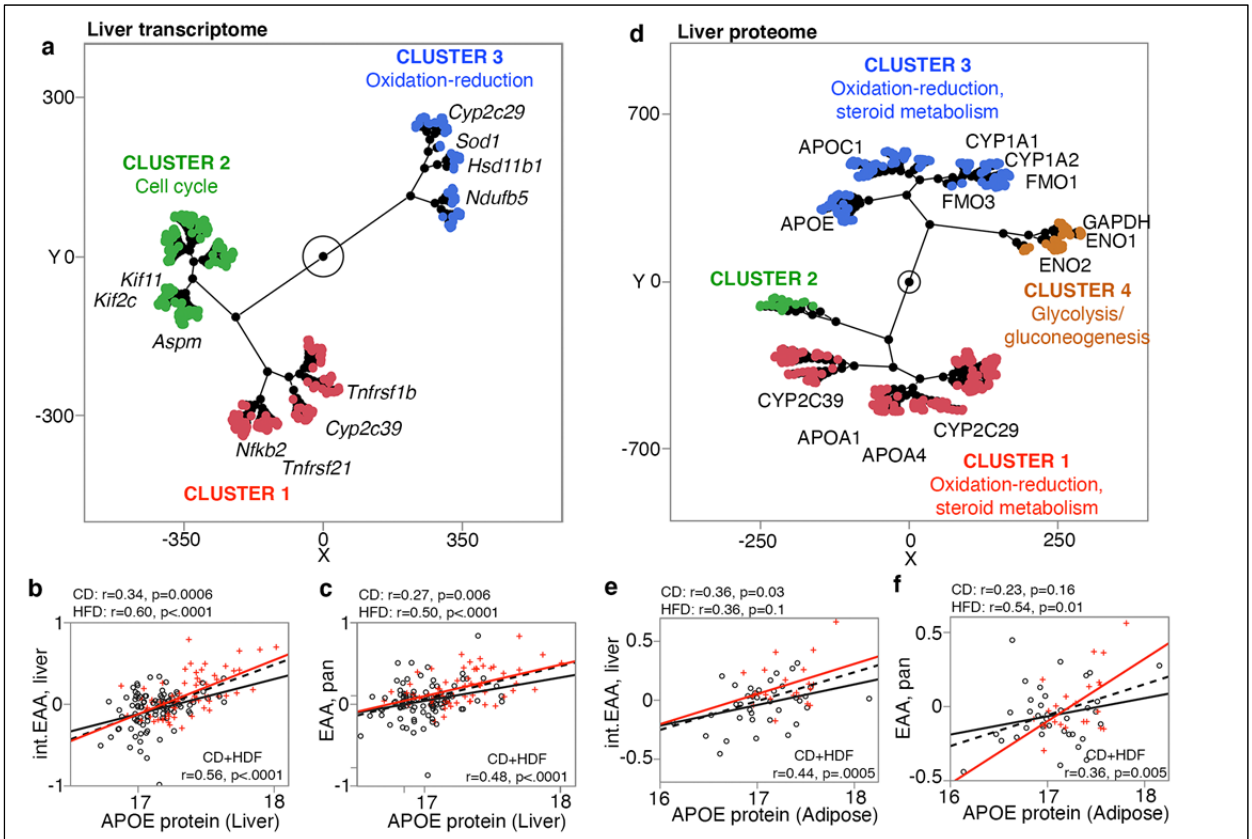


Figure 6. Gene expression correlates of epigenetic age acceleration

(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.

candidate, *Cyp2c29*, and cluster 2 included several cell cycle genes (**Figure 6a**). To verify that these transcriptomic associations are robust to the effect of diet, we repeated the correlation and enrichment analysis in the CD group only for the pan-tissue general clock (n = 94). Again, taking the top 2000 correlates ($|r| \leq 0.22$; $p \leq 0.03$), we found the same enrichment profiles for the positive correlates (immune, cell cycle) and the negative correlates (oxidation-reduction and mitochondrial).

Liver proteome was also available for 164 of the BXDs, and 53 also had adipose proteome. The liver proteome data quantifies over 32000 protein variants from 3940 unique genes and has been reported in Williams et al.[32] Similar to the transcriptome-wide analysis, we extracted the top 2000 protein correlates of EAA (**Supplementary file 12d**), and performed functional enrichment analysis (**Supplementary file 12b, 12c**). 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 (**Figure 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). 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 (**Supplementary file 12d**). For these common protein correlates, the HC divided the proteins into clusters that represented metabolic pathways mainly related to steroid metabolism, but also glycolysis and gluconeogenesis (**Figure 6d**; heatmap in **Figure 6-figure supplement 1b**).

Finally, we used the adipose proteome data for a proteome-wide correlational analysis for the pan-tissue EAA and liver int.EAA. We took only the top 1000 correlates (due to the small sample size), and a functional enrichment analysis showed consistent enrichment in metabolic pathways related to fatty acids and also carbohydrates, and cell proliferation genes for the pan-tissue EAA (**Supplementary file 12b, 12c**). For the adipose proteome, the cytochrome p450 genes were no longer enriched. However, the overall functional profile highlighted metabolic pathways as important gene expression correlates of EAA. Furthermore, for both the liver and adipose proteomes, APOE levels were highly correlated with EAA that indicates a higher level of this apolipoprotein in both tissues is associated with higher age acceleration (**Figure 6e, 6f**).

Discussion

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

We tested different mouse DNAm clocks, and the main difference between these clocks was the subsets of CpGs that were used for training. It is well-known that DNAm clocks have high level of degeneracy.[3,14] In other words, highly accurate predictors of chronological age can be built from entirely different sets of CpGs and different weight coefficients. This is likely because a large proportion of CpGs undergo some degree of change with age, and combinatorial information from any subset of this is informative of age. For instance, even at a very stringent cutoff of Bonferroni 0.05 that treated the 27966 CpGs as “independent”, we still detected 6553 CpGs as age-DMC, i.e., close to a quarter of the CpGs we tested. Clocks built from pre-selected

CpGs that are at conserved sequences are known to be sensitive to the effects of pro-longevity interventions such as calorie restriction and growth hormone receptor deletion.[3,54] And while all these DNAm clocks achieve reasonably high prediction of chronological age, the age divergence derived from these different clocks (EAA) can capture slightly different facets of biological aging, and the better a clock is at predicting chronological age, the lower its association with mortality risk.[13,14] In the present study, we find that the interventional clocks deviated most from chronological age, and this is expected as these were built from a much smaller set of CpGs (see Methods). The interventional clocks were also associated with BWF and cholesterol, but had weaker associations with BW0. The liver int.EAA had the highest positive correlation with methylome-wide entropy, and was the clock that had the strongest inverse correlation with strain longevity. In contrast, the developmental clocks, which were based on CpGs that change early in life, showed a stronger association with BW0. The contrast between the interventional and developmental clocks suggests that while one is more modifiable, the other is more informative of baseline characteristics that influence aging later in life. The pan-tissue clock, which was not constrained to any preselected set of CpGs or tissue, also performed well in capturing biological aging and was accelerated by both BW0 and BWF, diet (when BW0 was the weight term in the regression model), higher entropy, and had a modest but significant inverse correlation with strain lifespan.

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

Somewhat surprising was the inverse correlation between the entropy of age-DMCs and fasted glucose. This lower entropy of age-gain CpGs with higher glucose is somewhat counter to the general tendency for strains with shorter lifespan to have higher glucose.[33] In biological systems, entropy is kept at bay by the uptake of chemical energy, and investment in maintenance and repair,[57] and we can only speculate that at least in mice, the higher amount of glucose after overnight fast may be associated with a more ordered methylome. The centrality of bioenergetics for biological systems may explain why we detect this coupling between the DNAm readouts (i.e., the clocks, and entropy), and indices of metabolism including weight, diet, levels of macronutrients, and even expression of metabolic genes. As cogently highlighted by Donohoe and Bultman,[58] many metabolites (e.g., SAM, NAD^+, ATP) are essential co-factors for enzymes that shape the epigenome, and these could serve as nutrient sensors and mechanistic intermediaries that regulate how the epigenome is organized in response to metabolic conditions. Close interactions between macro- and micronutrients, and DNAm is a conserved

process and plays a critical role in defining both physiology and body morphology.[59,60] Overall, our results suggests that a higher metabolic state is associated with higher entropy and EAA, and potentially, lower lifespan.

For the BXDs, life expectancy is highly dependent on the background genotype, and mean lifespan varies from under 16 months for strains such as BXD8 and BXD13, to over 28 months in strains such as BXD91 and BXD175.[33,36,38] The EAA showed the expected inverse correlation with lifespan, but the effect was modest and only significant for the pan-tissue EAA and the liver int.EAA. The association of lifespan with the entropy of age-gain CpGs was slightly stronger. We must point out that the analysis between the epigenetic readouts and lifespan was an indirect comparison. Unlike the comparison with body weight and metabolic traits, which were traits measured from the same individual, the lifespan data are strain characteristics computed from a parallel cohort of mice that were allowed to survive till natural mortality, and this may partly explain the weaker associations with EAA. Nonetheless, our observations indicate that genotypes with higher life-expectancy have generally lower entropy, and lower methylation levels at the age-gain CpGs, and these properties of the methylome are likely to be partly under genetic modulation.

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

Eaa19 was consistently associated with EAA from all the clocks we evaluated, and also with body weight gains, irrespective of diet. DNAm entropy may also have a weak association with markers at this interval. The EAA traits have peak markers in the proximal part of Eaa19 (around the cytochrome cluster), and the methylome-wide entropy had a weak peak that was in the distal portion (over candidates like *Elovl3*, *Pcgf3*). Two candidates in Eaa19 have been implicated in epigenetic aging in humans based on gene-level meta-GWAS: NK homeobox 3 (*Nkx2-3*, a developmental gene), and CutC copper transporter (*Cutc*).[27] Eaa19 is also the location of the *Cyp26a1-Myof* genes, and the human syntenic region is associated with longevity, metabolic traits, and lipid profiles.[41,69,70] Another noteworthy candidate in Eaa19 is *Chuk*, a regulator of mTORC2, that has been associated with age at menopause.[64,71] Eaa19 presents a complex and intriguing QTL related to the DNAm readouts that may also influence body weight gains over the course of life. Both Eaa19 and Eaa11 exemplify the major challenge

that follows when a genetic mapping approach leads to gene- and variant-dense regions.[72,73] Both loci have several biologically relevant genes, and identifying the causal gene (or genes) will require a more fine-scaled functional genomic dissection.

The gene expression analyses highlighted metabolic pathways. At the mRNA level, the negative correlates of EAA were highly enriched in metabolic genes related to oxidation-reduction and steroid metabolism, while the positive correlates were enriched in pathways related to mitosis, and immune response for the pan-tissue general EAA. This convergence on metabolic, immune and cell division genes is very consistent with previous reports.[14,28,44] Here we should note that depending on the tissue(s) in which the clocks are trains, and the tissue from which the DNAmAge is estimated, the EAA derivative may put an emphasis on biological pathways or genes that are most relevant to that tissue. For instance, clocks optimized for neural tissue are more closely related to neurodegeneration and neuropathologies.[18,74] With the liver clocks, expression correlates highlighted aspects of metabolism that are relevant to liver function (e.g., the cytochrome p450 epoxygenase genes), and this is detected both at the transcriptomic, and proteomic levels. For the adipose tissue proteome, the cytochrome genes become less prominent, but the enriched pathways still remained consistent (i.e., oxidation-reduction, lipid and carbohydrate metabolism, and cell proliferation for the positive correlates of the pan-tissue EAA). At the proteome level, we also find several phospholipid efflux genes (APOC1, APOA2, APOC3, APOA1, APOA4, APOE) that are positive correlates of EAA. For both the liver and adipose proteomes, APOE stands out as the top protein correlate of EAA. A recent human study has also identified the APOE locus as the strongest GWAS hit for two measures of biological age acceleration (the phenoAge, and the bioAge).[28] While more specific to liver, the cytochrome P450 genes presents as both positional candidates, and expression correlates of EAA. These genes have high expression in liver, and have major downstream impact on metabolism.[75-77] One caveat is that these CYP genes are part of a gene cluster in Eaa19 that includes transcripts with *cis*-eQTLs (e.g., *Cyp2c66*, *Cyp2c39*, *Cyp2c68*), and the tight clustering of the genes, and proximity of trait QTL and eQTLs may result in tight co-expression due to linkage disequilibrium.[78] Nonetheless, the cytochrome genes in Eaa19 are strong candidate modulators of EAA derived from liver tissue that calls for further investigation.

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

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

Materials and Methods

Biospecimen collection and processing

Samples for this study were selected from a larger colony of BXD mice that were housed in a specific pathogen-free (SPF) facility at the University of Tennessee Health Science Center (UTHSC). All animal procedures were in accordance with a protocol approved by the Institutional Animal Care and Use Committee (IACUC) at the UTHSC. Detailed description of housing conditions and diet can be found in.[32,33] Mice were given *ad libitum* access to water, and either standard laboratory chow (Harlan Teklad; 2018, 18.6% protein, 6.2% fat, 75.2% carbohydrates), or high-fat chow (Harlan Teklad 06414; 18.4% protein, 60.3% fat, 21.3% carbohydrate). Animals were first weighed within the first few days of assignment to either diets, and this was mostly but not always prior to introduction to HFD. Following this, animals were weighed periodically, and a final time (BWF) when animals were humanely euthanized (anesthetized with avertin at 0.02 ml per g of weight, followed by perfusion with phosphate-buffered saline) at specific ages for tissue collection. The present work utilizes the biobanked liver specimens that were pulverized and stored in -80 °C, and overlaps samples described in Williams et al.[32] DNA was extracted using the DNeasy Blood & Tissue Kit from Qiagen. Nucleic acid purity was inspected with a NanoDrop spectrophotometer, and quantified using a Qubit fluorometer dsDNA BR Assay.

Methylation array and quality checks

DNA samples from ~350 BXD mice were profiled on the Illumina HorvathHumanMethylChip40 array. Samples were in 96-well plate format (**Supplementary file 1**), and the plates were randomized for major covariates such as age and diet. Details of this array are described in Arneson et al.[29,79] The array contains probes that target ~36K highly conserved CpGs in mammals. Over 33K probes map to homologous regions in the mouse genome. For downstream statistical tests, we further filtered the probes and used only 27966 probes that have been validated for the mouse genome using calibration data generated from synthetic mouse DNA.[29] Data was normalized using the SeSame method.[80] Unsupervised HC was performed to identify outliers and failed arrays, and those were excluded. We also performed strain verification as an additional quality check. While majority of the probes were free of DNA sequence variants, we found 45 probes that overlapped variants in the BXD family. We leveraged these as proxies for genotypes, and performed a principal component analysis (PCA). The top genotype principal components (genoPC1 and genoPC2; **Supplementary file 1**) segregated the samples by strain identity, and samples that did not cluster close to the reported strains were removed. After excluding outliers, failed arrays, and samples that failed strain verification, the final liver DNAm data consisted of 339 samples. The beta-values for these ~28K probes in the 339 samples show the expected bimodal distribution (**Figure 2-figure supplement 2a**), but for these highly conserved CpGs, we note a much higher representation of hypermethylated CpGs instead of the slightly hypomethylated state of the methylome when a wider spectrum of CpGs is assayed.[43]

BXD-unbiased mouse clock estimation

Three different mouse clocks are reported here, and all three are based on penalized regression modeling using glmnet.[81] Training was done in a larger mouse dataset that excluded the BXDs.[30,31,42] The clocks are therefore unbiased to the characteristics of the BXDs. For pan-tissue clocks, all mouse samples were used for training. For the liver specific clocks, the training was limited to data from liver samples.

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

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

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

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

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

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

Training for the interventional clock started with 537 CpGs that relate to gold-standard anti-aging interventions (calorie restriction, growth hormone receptor knockout).[42,82] These "interventional CpGs" were identified from an independent mouse liver calorie restriction (n = 95), and one growth hormone receptor knockout (n = 71) data that were not included in the clock estimation.[42] Top CpGs associated with these interventions were identified and the 537 CpGs are the sites that are consistently associated with these anti-aging interventions. Of the 537, 121 CpGs increased in methylation, and 417 decreased in methylation with application of the pro-longevity interventions. Given the small number of CpGs that went into training for the int.DNAmAge, we expected this clock to be less correlated with chronological age, and possibly more responsive variables such as diet.

Entropy calculation

Methylome-wide entropy was calculated from the 27966 probes. The beta-values were discretized into 20 bins, and the Shannon entropy for each sample was estimated using the R package, "entropy" (v1.2.1) with method = "ML": maximum likelihood.[83] The optimal number of bins was determined using the Freedman-Diaconis rule (breaks = "FD" for the hist() function in R). We also estimated the methylome-wide entropy after discretizing into 100 and 2000 bins (values provided in **Supplementary file 1**), and the results we report are consistent and robust to the number of bins. For the age-gain, age-loss, and age-ns CpGs, entropy for each set was estimated, also following discretization into 20 bins.

Statistics

Statistical analyses were done using R or the JMP Pro software (version 15). Association between the epigenetic predictors and continuous variables (body weight, strain lifespan, fasted serum glucose, and total cholesterol) were based on Pearson correlations, and t-test was used to evaluate the effect of categorical predictors (sex, diet). Multivariable regression models were also used to control for covariates (R regression equations provided with relevant tables and supplementary files). All these traits are directly accessible from GeneNetwork 2 (GN2; more information on how to retrieve these data from GN2 are provided in **Supplementary file 13**).[84,85] Longevity data was obtained from a parallel cohort of BXD mice housed in the same UTHSC colony, and members of this "longevity cohort" were allowed to age until natural death (more detail on the longevity cohort can be found in [33]). Males were excluded and strain-by-diet lifespan summary statistics were derived. Only strain-by-diet groups with 5 or more observations were included in the correlational analyses with the epigenetic predictors.

Multivariable EWAS

Site-by-site differential methylation analysis (EWAS) was performed on the 27966 CpGs using a multivariable regression model. As such genome-wide explorations are vulnerable to unmeasured confounders, we included the top PC derived from a PCA of the 27966 probes.[86] The top 10 principal components PCs cumulatively accounted for ~62% of the variance (**Figure 2-figure supplement 2b**). A plot of PC1 (19% of variance) and PC2 (14% of variance) showed that PC1 captured some noise due to batch (**Figure 2-figure supplement 2b**). The remaining top PCs (PC2 onwards) were strongly associated with biological variables, particular age, and also weight and diet (top 10 PCs provided in **Supplementary file 1**). For this reason, we included PC1 as a correction factor in the EWAS. The regression model we used was: lm(CpG_i~ age + median lifespan + diet + BWF+ PC1), where CpG_i is the i^{th} CpG from 1 to 27966. As lifespan was from female mice, this EWAS excluded the few male samples.

CpG annotation and enrichment

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

For chromatin state annotation, we used bedtools to annotate the 27966 CpGs coordinates using chromatin annotation .bed files for neonatal (P0) mouse liver tissue created by Gorkin et al.[48,87] This provides the 15-states model using ChromHMM,[47] and we downloaded the file for the "replicated set" (here, the regions annotated as NRS are sites that did not produce replicable signal). Enrichment and depletion analyses for genomic annotations, and chromatin annotations were based on the hypergeometric test (phyper R function). The R codes are provided with the results data (**Supplementary file 9**).

Genetic analyses

Heritability within diet was estimated as the fraction of variability that was explained by background genotype.[34,88,89] 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 **Supplementary file 13**).[84] 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 **Supplementary file 10**). 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,[90] 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 **Supplementary file 1**.

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 (**Supplementary file 13**). 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[91]: 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).[92-94] 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*.

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

Transcriptome and proteome analyses

The liver RNA-seq data mentioned above was also used for the transcriptome-wide correlational analysis for EAA in the 153 cases that had both DNAm and RNA-seq data. We considered the top 2000 highest mRNA correlates ($|r| = 0.24$, $p = 0.003$ for the pan-tissue EAA; $|r| = 0.3$, $p = 0.0002$ for the liver int.EAA), and the list of transcripts were collapsed to a non-redundant list of gene symbols, and this was uploaded to the DAVID Bioinformatics Database (version 2021 update) for GO enrichment analysis.[95,96] Proteome correlational analysis was carried out using the data: *Group: BXD NIA Longevity Study*; *Type: Liver Proteome*; and *Dataset: EPFL/ETHZ BXD Liver Proteome CD-HFD (Nov19)*. Detailed description of this data is in Williams et al.[32] 164 BXD cases had both DNAm and liver proteomics, and similar to the RNA-seq, we selected the top 2000 correlates (($|r| = 0.24$, $p = 0.002$ for both the pan-tissue EAA and liver int.EAA) for enrichment analysis.

59 of the BXD cases also have proteome data from adipose tissue (*Group: BXD NIA Longevity Study*; *Type: Adipose Proteome*; and *Dataset: Riken-Wu BXD Liver Proteome CD-HFD (Sep20)*). While small in sample number, we used this data to test whether we could recapitulate the same functional enrichment profiles in a different tissue. Details on sample preparation and processing steps for the adipose proteome is provided in the dataset's "Info" page on GN2. In brief, protein was extracted from the adipose samples by first lysis in a buffer with protease inhibitor, followed by homogenization with a glass dounce and sonication. The protein fraction was isolated from the homogenate by centrifugation, and processed for assay on a liquid chromatography tandem mass spectrometry (LC-M/MS) using a modified Phase Transfer Surfactant Method as described in Mostafa et al.[97,98] Samples were measured using a Q Exactive Plus Orbitrap LC–MS/MS System (Thermo Fisher). For each sample, 600 ng was injected and the samples were measured with data-independent acquisition (DIA). A portion of the peptides from the samples were pooled and fractionated using a Pierce High pH Reversed-

Phase (HPRP) Peptide Fractionation Kit (Thermo Fisher Scientific) to generate a spectral library. For the HPRP fractions, 450 ng was injected and the samples were measured with data-dependent acquisition (DDA). For protein identification, the raw measurement files were searched against a mouse database using the (uniprot-reviewed_Mus_musculus_10090_.fasta) using Proteome Discoverer v2.4 software (Thermo Fisher Scientific). Filtered output was used to 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.[98] The false discovery rate was estimated with the mProphet approach and set to 0.01 at both peptide precursor level and protein level.[99,100] 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 normalized microarray data and raw files are available from the NCBI Gene Expression Omnibus (accession ID GSE199979). The HorvathMammalMethylChip40 array manifest files and genome annotations of CpGs can be found on Github at https://github.com/shorvath/MammalianMethylationConsortium.[79] Individual level BXD data, including the processed microarray data are available on www.genenetwork.org[84] on FAIR+ compliant format; data identifiers, and way to retrieve data are described in **Supplementary file 13**.

Acknowledgement. We thank the entire UTHSC BXD Aging Colony team, particularly Dr. Suheeta Roy, Casey J Chapman, Melinda S McCarty, Jesse Ingles, and everyone else who contributed to the tissue harvest. We thank Dr. Lu Lu, who leads the main BXD Colony effort. We thank Dr. Evan G Williams for making the gene expression data readily available, and to Dr. 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, and Dr. Pjotr Prins for his role in implementing GEMMA. We thank Dr. Garrett Jenkinson for the invaluable guidance he provided for entropy estimation.

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, University of California Regents. The Regents of the University of California filed a patent application (publication number WO2020150705) related to the HorvathMammalMethylChip40 and clock computation for which SH is named an inventor. 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. Protocol numbers 12-148.0 (2012–2015), 15-124.0 (2015–2018), and 18-094.0 (2018–present).

Additional files list.

Supplementary file 1. Individual-level sample information (file name: SupplementaryFile1_sampleInfo.xlsx)

Supplementary file 2. CpGs and coefficients for the mouse clocks (file name: SupplementaryFile2_ClockCpGs.csv)

Supplementary file 3. Covariates of the DNA methylation based readouts (file name: SupplementaryFile3_Covariates.xlsx)

Supplementary file 4. Supplementary Tables 4a, 4b, and 4c (file name: SupplementaryFile4_Tables.docx)

- **Supplementary file 4a.** Sex differences in epigenetic aging after correction for body weight
- **Supplementary file 4b.** Pearson correlations between epigenetic age acceleration and strain-level longevity summaries
- **Supplementary file 4c.** High priority candidate genes in QTLs for epigenetic age acceleration

Supplementary file 5. Multivariable regression analysis of epigenetic age acceleration (file name: SupplementaryFile5_RegressionOutputEAA.xlsx)

Supplementary file 6. Epigenome-wide association study results and annotations for 27996 CpG probes (file name: SupplementaryFile6_EWAS.csv)

Supplementary file 7. Multivariable regression analysis of entropy by age-effect (file name: SupplementaryFile7_RegressionOutputEntropy.xlsx)

Supplementary file 8. Genomic regions enrichments analyses for the differentially methylated CpGs (file: SupplementaryFile8_rGREATresults.csv)

Supplementary file 9. Enrichment/depletion in genomic regions and chromatin states for the differentially methylated CpGs (file: SupplementaryFile9_GenomicStates.xlsx)

Supplementary file 10. QTL analysis of epigenetic age acceleration and methylome-wide entropy (file: SupplementaryFile10_QTLresults.xlsx)

Supplementary file 11. Positional candidate genes in Eaa11 and Eaa19 (file: SupplementaryFile11_QTLcandidates.csv)

Supplementary file 12. Transcriptome and proteome analysis (file: SupplementaryFile12_GeneExpression.xlsx)

- **Supplementary file 12a.** Top 2000 liver transcriptome-wide correlates of liver int.EAA and pan-tissue EAA
- **Supplementary file 12b.** Functional enrichment among gene expression correlates of pan-tissue general clock (pan-tissue EAA)
- **Supplementary file 12c.** Functional enrichment among gene expression correlates of liver interventional clock (liver int.EAA)
- **Supplementary file 12d.** Top 2000 liver proteome correlates of liver int.EAA and pan-tissue EAA

Supplementary file 13. Data access (file: SupplementaryFile13_DataAccess.xlsx)

Figure Supplements Legends.

Figure 1-figure supplement 1. Relative effects of different predictor variables on epigenetic age acceleration (EAA)

Logworth scores of the predictors ($-\log_{10}p$) with dashed lines corresponding to p = 0.01. Positive logworth values indicate positive regression estimates, and negative values indicate negative regression estimates (for diet, positive means higher in high fat diet compared to control diet). BW0 is baseline weight; Chol is serum total cholesterol, Gluc is fasted glucose levels.

Figure 1-figure supplement 2. BXD strains with shorter life expectancy have slightly more accelerated clocks

This inverse correlation is depicted for the **(a)** 75th quartile age at natural death, **(b)** the minimum lifespan, and **(c)** the median lifespan (analysis in 302 female samples with lifespan data). CD is control diet; HFD is high fat diet. The negative correlations are modest with explained variance values, r2, of about ~3%.

Figure 2-figure supplement 1. Genomic and chromatin states of differentially methylated CpGs

Enrichment is **(a)** genomic location, and **(b)** chromatin states among the differentially methylated CpGs (DMC) (expansions for the chromatin states are provided in **Supplementary file 9**). Asterisks denote hypergeometric enrichment p < 0.001. For the 15 chromatin states (and regions with no replicable signal, NRS), we compare the methylation levels, and mean regression estimates for the effects of **(c)** age, **(d)** diet, and **(e)** body weight.

Figure 2-figure supplement 2. Array quality check

(a) Density plot for the 339 cases using the full set of CpG probes. **(b)** Variance explained by the top 10 principal components (PCs) derived from the full set of probes. **(c)** Plot between component 1 and 2 shows the PC1 captures some batch effect. Here batch is the 96-well plates.

Figure 4-figure supplement 1. Consensus QTL mapping for epigenetic age acceleration

(s) The Manhattan plot displays the combined meta p-values for epigenetic age acceleration (EAA). These meta p-values are based on a simple p-value combination for the six EAA QTL traits, and is mainly to highlight regions with the highest consensus QTLs. The highest peaks are on chromosomes 11 (Eaa11), and 19 (Eaa19). **(b)** BXDs were segregated by genotype at a representative marker in Eaa11 (variant at 92.750 Mb). In the control diet group (CD), mean EAA (± standard error) is higher for mice with the *DD* genotype. Only the EAA derived from the liver interventional clock (int.EAA) shows no difference between the genotypes. **(c)** BXDs were segregated by the genotype at a marker in Eaa19 (38.650 Mb). Mean EAA is higher in the *BB* genotype, and this genotype effect is seen for all the clocks on both diets. Bars are standard error.

Figure 6-figure supplement 1. Hierarchical clustering heatmaps for the top expression correlates of epigenetic age acceleration. The dendrograms represent the liver expression of **(a)** mRNA, and **(b)** proteins that are correlated with age acceleration derived from both the pan-tissue general clock, and the liver interventional clock.

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).

2 Bell, C. G. *et al.* DNA methylation aging clocks: challenges and recommendations. *Genome biology* **20**, 249, doi:10.1186/s13059-019-1824-y (2019).

3 Thompson, M. J. *et al.* A multi-tissue full lifespan epigenetic clock for mice. *Aging* **10**, 2832-2854, doi:10.18632/aging.101590 (2018).

4 Porter, H. L. *et al.* Many chronological aging clocks can be found throughout the epigenome: Implications for quantifying biological aging. *Aging cell* **20**, e13492, doi:10.1111/acel.13492 (2021).

5 Zannas, A. S. *et al.* Lifetime stress accelerates epigenetic aging in an urban, African American cohort: relevance of glucocorticoid signaling. *Genome biology* **16**, 266, doi:10.1186/s13059-015-0828-5 (2015).

6 Marioni, R. E. *et al.* The epigenetic clock is correlated with physical and cognitive fitness in the Lothian Birth Cohort 1936. *International journal of epidemiology* **44**, 1388-1396, doi:10.1093/ije/dyu277 (2015).

7 Dugue, P. A. *et al.* DNA methylation-based biological aging and cancer risk and survival: Pooled analysis of seven prospective studies. *Int J Cancer* **142**, 1611-1619, doi:10.1002/ijc.31189 (2018).

8 Lu, A. T. *et al.* DNA methylation GrimAge strongly predicts lifespan and healthspan. *Aging* **11**, 303-327, doi:10.18632/aging.101684 (2019).

9 Ryan, J., Wrigglesworth, J., Loong, J., Fransquet, P. D. & Woods, R. L. A systematic review and meta-analysis of environmental, lifestyle and health factors associated with DNA methylation age. *The journals of gerontology. Series A, Biological sciences and medical sciences*, doi:10.1093/gerona/glz099 (2019).

10 Hannum, G. *et al.* Genome-wide methylation profiles reveal quantitative views of human aging rates. *Molecular cell* **49**, 359-367, doi:10.1016/j.molcel.2012.10.016 (2013).

11 Horvath, S. DNA methylation age of human tissues and cell types. *Genome biology* **14**, R115, doi:10.1186/gb-2013-14-10-r115 (2013).

12 Bocklandt, S. *et al.* Epigenetic predictor of age. *PloS one* **6**, e14821, doi:10.1371/journal.pone.0014821 (2011).

13 Zhang, Q. *et al.* Improved precision of epigenetic clock estimates across tissues and its implication for biological ageing. *Genome Med* **11**, 54, doi:10.1186/s13073-019-0667-1 (2019).

14 Liu, Z. *et al.* Underlying features of epigenetic aging clocks in vivo and in vitro. *Aging cell* **19**, e13229, doi:10.1111/acel.13229 (2020).

15 Lee, Y. *et al.* Blood-based epigenetic estimators of chronological age in human adults using DNA methylation data from the Illumina MethylationEPIC array. *BMC genomics* **21**, 747, doi:10.1186/s12864-020-07168-8 (2020).

16 Horvath, S. *et al.* Epigenetic clock for skin and blood cells applied to Hutchinson Gilford Progeria Syndrome and ex vivo studies. *Aging* **10**, 1758-1775, doi:10.18632/aging.101508 (2018).

17 Levine, M. E. *et al.* An epigenetic biomarker of aging for lifespan and healthspan. *Aging* **10**, 573-591, doi:10.18632/aging.101414 (2018).
18 Shireby, G. L. *et al.* Recalibrating the epigenetic clock: implications for assessing biological age in the human cortex. *Brain* **143**, 3763-3775, doi:10.1093/brain/awaa334 (2020).
19 Petkovich, D. A. *et al.* Using DNA Methylation Profiling to Evaluate Biological Age and Longevity Interventions. *Cell metabolism* **25**, 954-960 e956, doi:10.1016/j.cmet.2017.03.016 (2017).
20 Stubbs, T. M. *et al.* Multi-tissue DNA methylation age predictor in mouse. *Genome biology* **18**, 68, doi:10.1186/s13059-017-1203-5 (2017).
21 Wang, T. *et al.* Epigenetic aging signatures in mice livers are slowed by dwarfism, calorie restriction and rapamycin treatment. *Genome biology* **18**, 57, doi:10.1186/s13059-017-1186-2 (2017).
22 Levine, M. *et al.* A rat epigenetic clock recapitulates phenotypic aging and co-localizes with heterochromatin. *Elife* **9**, doi:10.7554/eLife.59201 (2020).
23 Han, Y. *et al.* Epigenetic age-predictor for mice based on three CpG sites. *Elife* **7**, doi:10.7554/eLife.37462 (2018).
24 Gibson, J. *et al.* A meta-analysis of genome-wide association studies of epigenetic age acceleration. *PLoS genetics* **15**, e1008104, doi:10.1371/journal.pgen.1008104 (2019).
25 Lin, W. Y. Genome-wide association study for four measures of epigenetic age acceleration and two epigenetic surrogate markers using DNA methylation data from Taiwan biobank. *Human molecular genetics*, doi:10.1093/hmg/ddab369 (2021).
26 Lu, A. T. *et al.* GWAS of epigenetic aging rates in blood reveals a critical role for TERT. *Nature communications* **9**, 387, doi:10.1038/s41467-017-02697-5 (2018).
27 McCartney, D. L. *et al.* Genome-wide association studies identify 137 genetic loci for DNA methylation biomarkers of aging. *Genome biology* **22**, 194, doi:10.1186/s13059-021-02398-9 (2021).
28 Kuo, C. L., Pilling, L. C., Liu, Z., Atkins, J. L. & Levine, M. E. Genetic associations for two biological age measures point to distinct aging phenotypes. *Aging cell* **20**, e13376, doi:10.1111/acel.13376 (2021).
29 Arneson, A. *et al.* A mammalian methylation array for profiling methylation levels at conserved sequences. *Nature communications* **13**, 783, doi:10.1038/s41467-022-28355-z (2022).
30 Lu, A. T. *et al.* Universal DNA methylation age across mammalian tissues. *bioRxiv*, 2021.2001.2018.426733, doi:10.1101/2021.01.18.426733 (2021).
31 Li, C. Z. *et al.* Epigenetic predictors of maximum lifespan and other life history traits in mammals. *bioRxiv*, 2021.2005.2016.444078, doi:10.1101/2021.05.16.444078 (2021).
32 Williams, E. G. *et al.* Multiomic profiling of the liver across diets and age in a diverse mouse population. *Cell Syst*, doi:10.1016/j.cels.2021.09.005 (2021).
33 Roy, S. *et al.* Gene-by-environment modulation of lifespan and weight gain in the murine BXD family. *Nat Metab* **3**, 1217-1227, doi:10.1038/s42255-021-00449-w (2021).
34 Ashbrook, D. G. *et al.* A Platform for Experimental Precision Medicine: The Extended BXD Mouse Family. *Cell Syst*, doi:10.1016/j.cels.2020.12.002 (2021).

35 Peirce, J. L., Lu, L., Gu, J., Silver, L. M. & Williams, R. W. A new set of BXD recombinant inbred lines from advanced intercross populations in mice. *BMC genetics* **5**, 7, doi:10.1186/1471-2156-5-7 (2004).
36 de Haan, G. & Williams, R. W. A genetic and genomic approach to identify longevity genes in mice. *Mechanisms of ageing and development* **126**, 133-138, doi:10.1016/j.mad.2004.09.012 (2005).
37 Hsu, H. C. *et al.* Age-related thymic involution in C57BL/6J x DBA/2J recombinant-inbred mice maps to mouse chromosomes 9 and 10. *Genes and immunity* **4**, 402-410, doi:10.1038/sj.gene.6363982 (2003).
38 Lang, D. H. *et al.* Quantitative trait loci (QTL) analysis of longevity in C57BL/6J by DBA/2J (BXD) recombinant inbred mice. *Aging clinical and experimental research* **22**, 8-19 (2010).
39 Sandoval-Sierra, J. V. *et al.* Body weight and high-fat diet are associated with epigenetic aging in female members of the BXD murine family. *Aging cell*, e13207, doi:10.1111/acel.13207 (2020).
40 Benton, M. C. *et al.* Methylome-wide association study of whole blood DNA in the Norfolk Island isolate identifies robust loci associated with age. *Aging* **9**, 753-768, doi:10.18632/aging.101187 (2017).
41 Yashin, A. I. *et al.* Genetics of Human Longevity From Incomplete Data: New Findings From the Long Life Family Study. *The journals of gerontology. Series A, Biological sciences and medical sciences* **73**, 1472-1481, doi:10.1093/gerona/gly057 (2018).
42 Haghani, A. *et al.* Divergent age-related methylation patterns in long and short-lived mammals. *bioRxiv*, 2022.2001.2016.476530, doi:10.1101/2022.01.16.476530 (2022).
43 Sziraki, A., Tyshkovskiy, A. & Gladyshev, V. N. Global remodeling of the mouse DNA methylome during aging and in response to calorie restriction. *Aging cell* **17**, e12738, doi:10.1111/acel.12738 (2018).
44 Slieker, R. C. *et al.* Age-related accrual of methylomic variability is linked to fundamental ageing mechanisms. *Genome biology* **17**, 191, doi:10.1186/s13059-016-1053-6 (2016).
45 Kerepesi, C. *et al.* Epigenetic aging of the demographically non-aging naked mole-rat. *Nature communications* **13**, 355, doi:10.1038/s41467-022-27959-9 (2022).
46 McLean, C. Y. *et al.* GREAT improves functional interpretation of cis-regulatory regions. *Nature biotechnology* **28**, 495-501, doi:10.1038/nbt.1630 (2010).
47 Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. *Nat Methods* **9**, 215-216, doi:10.1038/nmeth.1906 (2012).
48 Gorkin, D. U. *et al.* An atlas of dynamic chromatin landscapes in mouse fetal development. *Nature* **583**, 744-751, doi:10.1038/s41586-020-2093-3 (2020).
49 Zhou, X. & Stephens, M. Efficient multivariate linear mixed model algorithms for genome-wide association studies. *Nature methods* **11**, 407-409, doi:10.1038/nmeth.2848 (2014).
50 Peirce, J. L., Broman, K. W., Lu, L. & Williams, R. W. A simple method for combining genetic mapping data from multiple crosses and experimental designs. *PloS one* **2**, e1036, doi:10.1371/journal.pone.0001036 (2007).

51 Buniello, A. *et al.* The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic acids research* **47**, D1005-D1012, doi:10.1093/nar/gky1120 (2019).

52 Horvath, S. *et al.* Obesity accelerates epigenetic aging of human liver. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 15538-15543, doi:10.1073/pnas.1412759111 (2014).

53 Nevalainen, T. *et al.* Obesity accelerates epigenetic aging in middle-aged but not in elderly individuals. *Clinical epigenetics* **9**, 20, doi:10.1186/s13148-016-0301-7 (2017).

54 Wang, M. & Lemos, B. Ribosomal DNA harbors an evolutionarily conserved clock of biological aging. *Genome research* **29**, 325-333, doi:10.1101/gr.241745.118 (2019).

55 Xie, H. *et al.* Genome-wide quantitative assessment of variation in DNA methylation patterns. *Nucleic acids research* **39**, 4099-4108, doi:10.1093/nar/gkr017 (2011).

56 Jenkinson, G., Pujadas, E., Goutsias, J. & Feinberg, A. P. Potential energy landscapes identify the information-theoretic nature of the epigenome. *Nature genetics* **49**, 719-729, doi:10.1038/ng.3811 (2017).

57 Hayflick, L. Entropy explains aging, genetic determinism explains longevity, and undefined terminology explains misunderstanding both. *PLoS genetics* **3**, e220, doi:10.1371/journal.pgen.0030220 (2007).

58 Donohoe, D. R. & Bultman, S. J. Metaboloepigenetics: interrelationships between energy metabolism and epigenetic control of gene expression. *J Cell Physiol* **227**, 3169-3177, doi:10.1002/jcp.24054 (2012).

59 Kucharski, R., Maleszka, J., Foret, S. & Maleszka, R. Nutritional control of reproductive status in honeybees via DNA methylation. *Science* **319**, 1827-1830, doi:10.1126/science.1153069 (2008).

60 Dolinoy, D. C., Huang, D. & Jirtle, R. L. Maternal nutrient supplementation counteracts bisphenol A-induced DNA hypomethylation in early development. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 13056-13061, doi:10.1073/pnas.0703739104 (2007).

61 Holman, G. D. A new deadly Syn? *Curr Biol* **9**, R735-737, doi:10.1016/s0960-9822(99)80471-0 (1999).

62 Michailidou, K. *et al.* Association analysis identifies 65 new breast cancer risk loci. *Nature* **551**, 92-94, doi:10.1038/nature24284 (2017).

63 Rokudai, S. *et al.* STXBP4 regulates APC/C-mediated p63 turnover and drives squamous cell carcinogenesis. *Proceedings of the National Academy of Sciences of the United States of America* **115**, E4806-E4814, doi:10.1073/pnas.1718546115 (2018).

64 Kichaev, G. *et al.* Leveraging Polygenic Functional Enrichment to Improve GWAS Power. *American journal of human genetics* **104**, 65-75, doi:10.1016/j.ajhg.2018.11.008 (2019).

65 Perry, J. R. *et al.* Parent-of-origin-specific allelic associations among 106 genomic loci for age at menarche. *Nature* **514**, 92-97, doi:10.1038/nature13545 (2014).

66 Albertsen, H. M. *et al.* A physical map and candidate genes in the BRCA1 region on chromosome 17q12-21. *Nature genetics* **7**, 472-479, doi:10.1038/ng0894-472 (1994).

67 Kauraniemi, P. & Kallioniemi, A. Activation of multiple cancer-associated genes at the ERBB2 amplicon in breast cancer. *Endocr Relat Cancer* **13**, 39-49, doi:10.1677/erc.1.01147 (2006).

68 Tanaka, S. *et al.* Coexpression of Grb7 with epidermal growth factor receptor or Her2/erbB2 in human advanced esophageal carcinoma. *Cancer research* **57**, 28-31 (1997).
69 de Vries, P. S. *et al.* Multiancestry Genome-Wide Association Study of Lipid Levels Incorporating Gene-Alcohol Interactions. *Am J Epidemiol* **188**, 1033-1054, doi:10.1093/aje/kwz005 (2019).
70 Richardson, T. G. *et al.* Evaluating the relationship between circulating lipoprotein lipids and apolipoproteins with risk of coronary heart disease: A multivariable Mendelian randomisation analysis. *PLoS Med* **17**, e1003062, doi:10.1371/journal.pmed.1003062 (2020).
71 Xu, Y. *et al.* IKK interacts with rictor and regulates mTORC2. *Cell Signal* **25**, 2239-2245, doi:10.1016/j.cellsig.2013.07.008 (2013).
72 Gallagher, M. D. & Chen-Plotkin, A. S. The Post-GWAS Era: From Association to Function. *American journal of human genetics* **102**, 717-730, doi:10.1016/j.ajhg.2018.04.002 (2018).
73 Lappalainen, T. Functional genomics bridges the gap between quantitative genetics and molecular biology. *Genome research* **25**, 1427-1431, doi:10.1101/gr.190983.115 (2015).
74 Grodstein, F. *et al.* The association of epigenetic clocks in brain tissue with brain pathologies and common aging phenotypes. *Neurobiol Dis* **157**, 105428, doi:10.1016/j.nbd.2021.105428 (2021).
75 Olona, A. *et al.* Epoxygenase inactivation exacerbates diet and aging-associated metabolic dysfunction resulting from impaired adipogenesis. *Mol Metab* **11**, 18-32, doi:10.1016/j.molmet.2018.03.003 (2018).
76 Schuck, R. N. *et al.* The cytochrome P450 epoxygenase pathway regulates the hepatic inflammatory response in fatty liver disease. *PloS one* **9**, e110162, doi:10.1371/journal.pone.0110162 (2014).
77 Wang, Q. *et al.* Time serial transcriptome reveals Cyp2c29 as a key gene in hepatocellular carcinoma development. *Cancer Biol Med* **17**, 401-417, doi:10.20892/j.issn.2095-3941.2019.0335 (2020).
78 Mozhui, K. *et al.* Dissection of a QTL hotspot on mouse distal chromosome 1 that modulates neurobehavioral phenotypes and gene expression. *PLoS genetics* **4**, e1000260, doi:10.1371/journal.pgen.1000260 (2008).
79 Horvath, S., Haghani, A., Arneson, A. & Ernst, J. Mammalian Methylation Consortium, https://github.com/shorvath/MammalianMethylationConsortium, v1.0.0. (2021).
80 Zhou, W., Triche, T. J., Jr., Laird, P. W. & Shen, H. SeSAMe: reducing artifactual detection of DNA methylation by Infinium BeadChips in genomic deletions. *Nucleic acids research* **46**, e123, doi:10.1093/nar/gky691 (2018).
81 Friedman, J., Hastie, T. & Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw* **33**, 1-22 (2010).
82 Coschigano, K. T. *et al.* Deletion, but not antagonism, of the mouse growth hormone receptor results in severely decreased body weights, insulin, and insulin-like growth factor I levels and increased life span. *Endocrinology* **144**, 3799-3810, doi:10.1210/en.2003-0374 (2003).

83 Hausser, J. & Strimmer, K. Entropy Inference and the James-Stein Estimator, with Application to Nonlinear Gene Association Networks. *J. Mach. Learn. Res.* **10**, 1469–1484 (2009).

84 Williams, R. W. GeneNetwork 2, http://genenetwork.org. (2022).

85 Mulligan, M. K., Mozhui, K., Prins, P. & Williams, R. W. GeneNetwork: A Toolbox for Systems Genetics. *Methods in molecular biology* **1488**, 75-120, doi:10.1007/978-1-4939-6427-7_4 (2017).

86 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).

87 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).

88 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).

89 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).

90 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).

91 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, v1.1-28.

92 *Wellcome Sanger Institute Mouse Genome Project, https://www.sanger.ac.uk/sanger/Mouse_SnpViewer/rel-1505*.

93 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).

94 Yalcin, B. *et al.* Sequence-based characterization of structural variation in the mouse genome. *Nature* **477**, 326-329, doi:10.1038/nature10432 (2011).

95 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).

96 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).

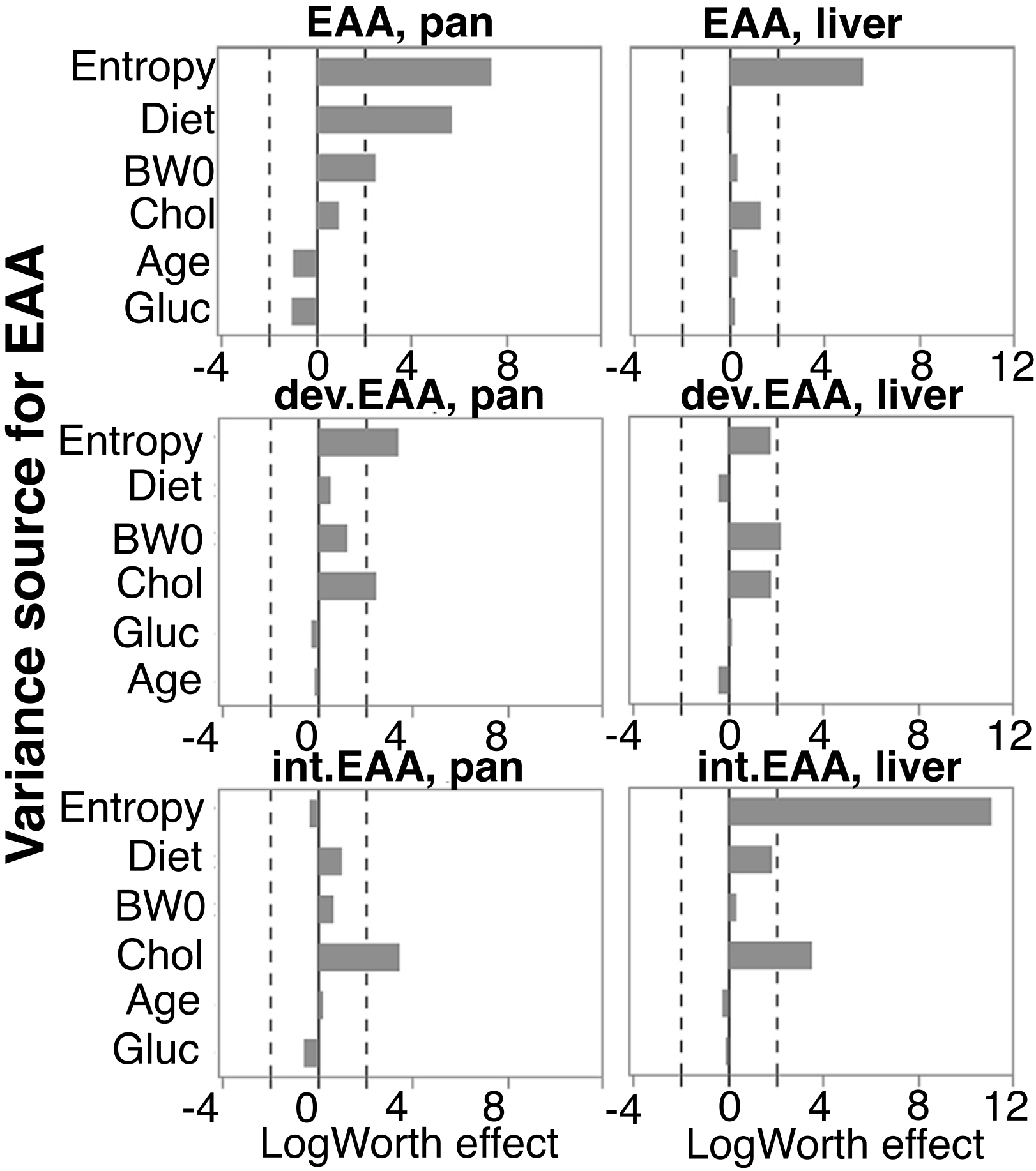
97 Masuda, T., Tomita, M. & Ishihama, Y. Phase transfer surfactant-aided trypsin digestion for membrane proteome analysis. *J Proteome Res* **7**, 731-740, doi:10.1021/pr700658q (2008).

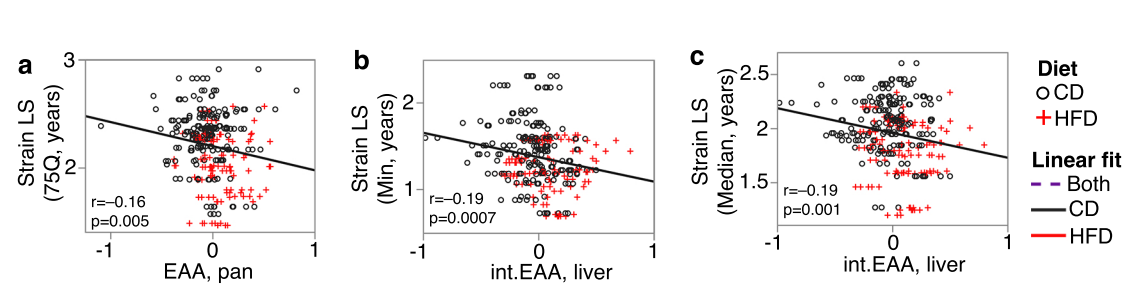
98 Mostafa, D. *et al.* Loss of beta-cell identity and diabetic phenotype in mice caused by disruption of CNOT3-dependent mRNA deadenylation. *Commun Biol* **3**, 476, doi:10.1038/s42003-020-01201-y (2020).

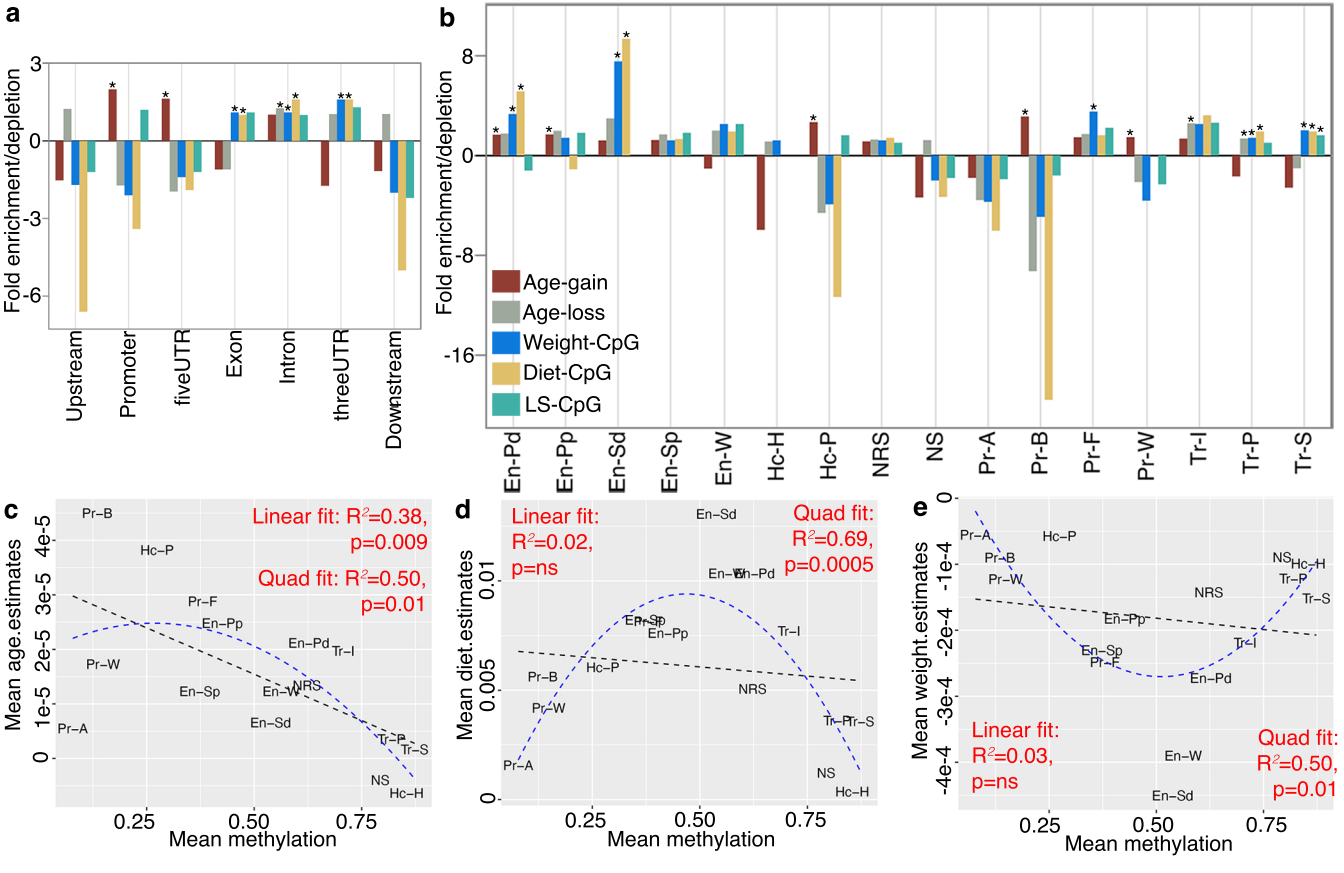
99 Reiter, L. *et al.* mProphet: automated data processing and statistical validation for large-scale SRM experiments. *Nature methods* **8**, 430-435, doi:10.1038/nmeth.1584 (2011).

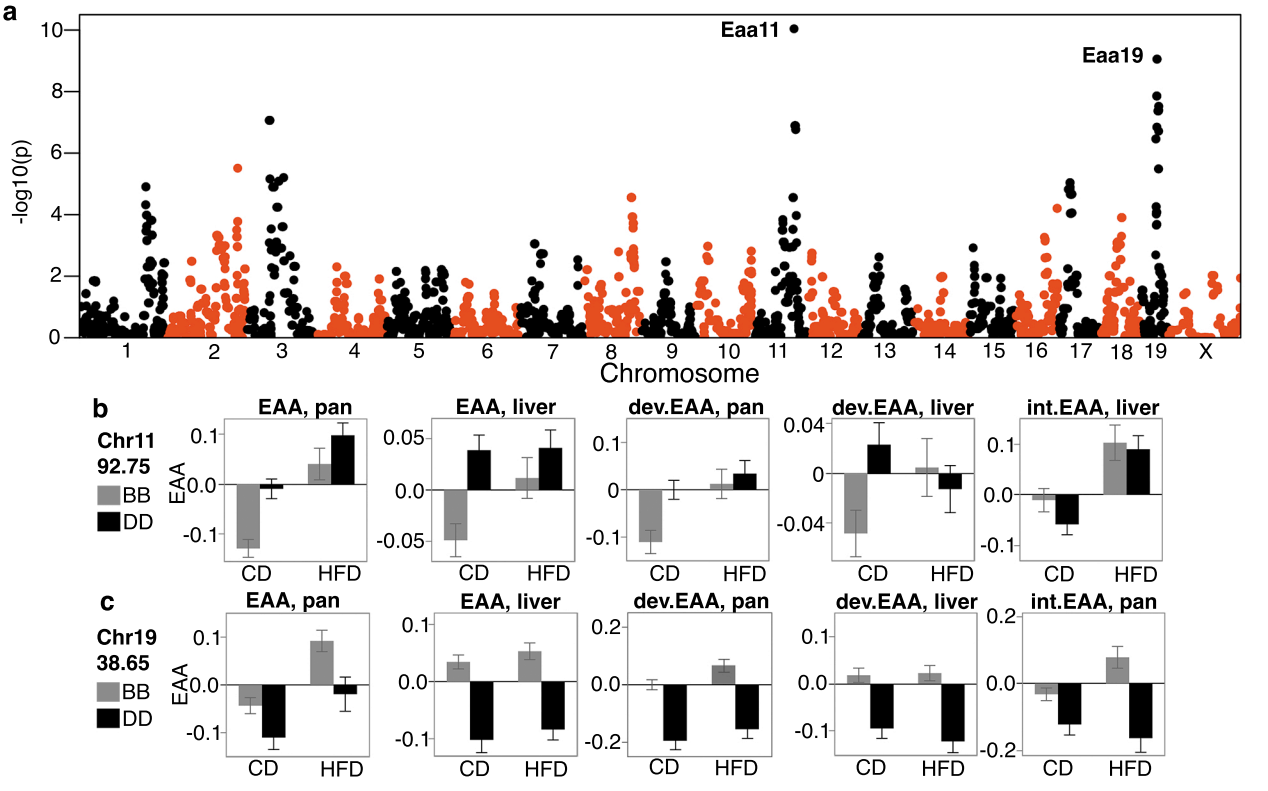
100 Rosenberger, G. *et al.* Statistical control of peptide and protein error rates in large-scale targeted data-independent acquisition analyses. *Nature methods* **14**, 921-927, doi:10.1038/nmeth.4398 (2017).

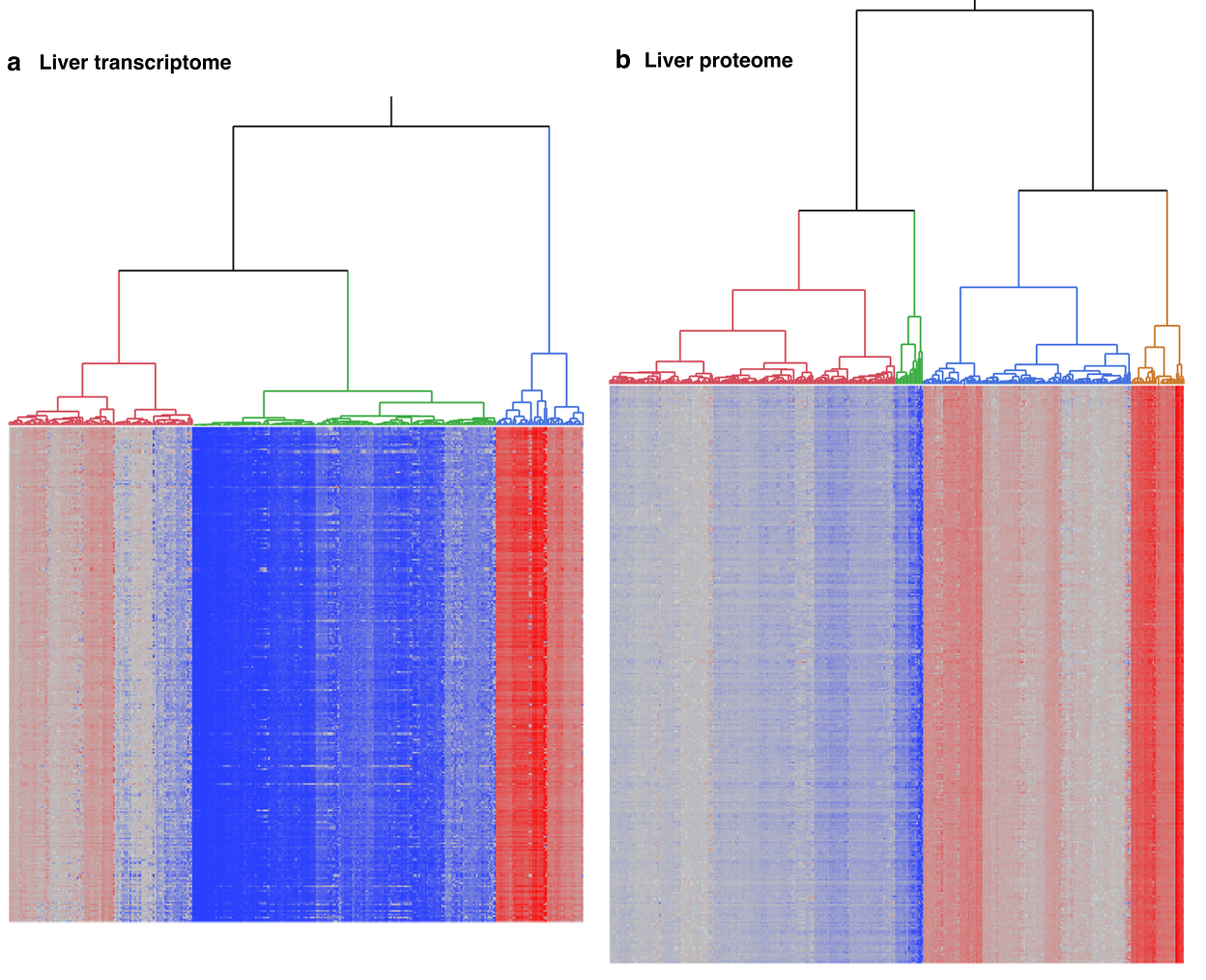
1206

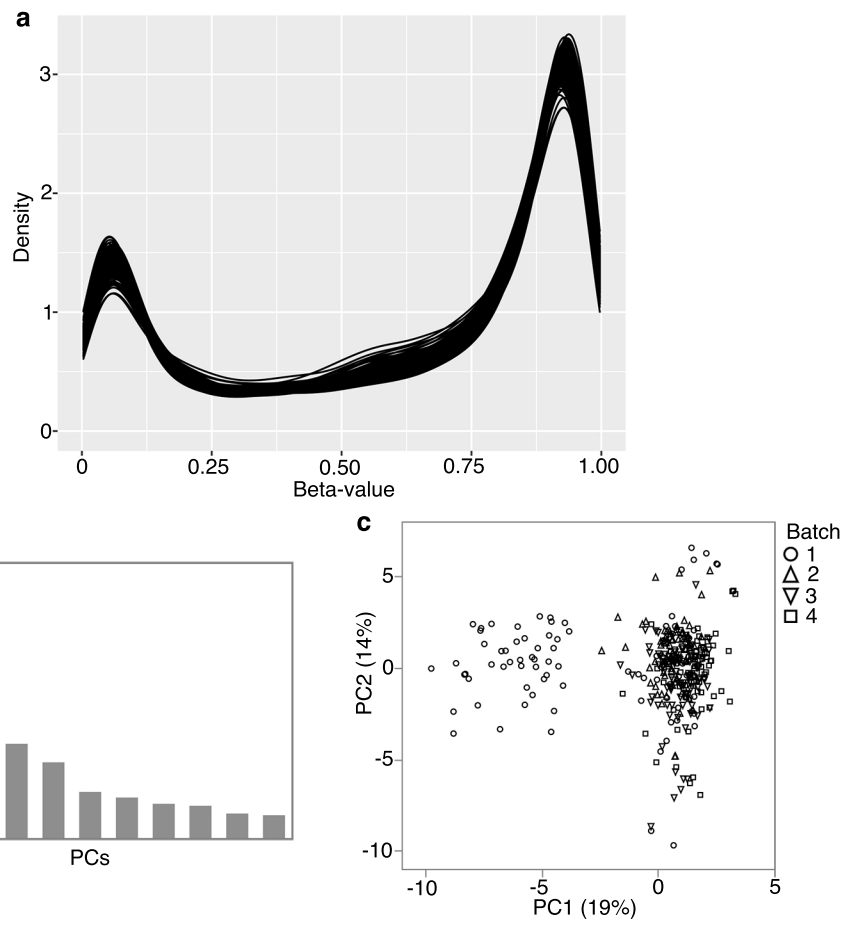

EAA, pan
EAA, liver
dev.EAA, pan
dev.EAA, liver
int.EAA, pan
int.EAA, liver
Variance source for EAA
Entropy
Diet
BW0
Chol
Age
Gluc
-4
0
4
8
12
LogWorth effect


a
Strain LS
(75Q, years)
3
2
r=−0.16
p=0.005
-1
0
1
EAA, pan
b
Strain LS
(Min, years)
2
1
r=−0.19
p=0.0007
-1
0
1
int.EAA, liver
c
Strain LS
(Median, years)
2.5
2
1.5
r=−0.19
p=0.001
-1
0
1
int.EAA, liver
Diet
CD
HFD
Linear fit
Both
CD
HFD


a
Fold enrichment/depletion
3
0
-3
-6
Upstream
Promoter
fiveUTR
Exon
Intron
threeUTR
Downstream
b
Fold enrichment/depletion
8
0
-8
-16
Age-gain
Age-loss
Weight-CpG
Diet-CpG
LS-CpG
En-Pd
En-Pp
En-Sd
En-Sp
En-W
Hc-H
Hc-P
NRS
NS
Pr-A
Pr-B
Pr-F
Pr-W
Tr-I
Tr-P
Tr-S
c
Linear fit: R2=0.38, p=0.009
Quad fit: R2=0.50, p=0.01
Mean age.estimates
Mean methylation
d
Linear fit: R2=0.02, p=ns
Quad fit: R2=0.69, p=0.0005
Mean diet.estimates
Mean methylation
e
Linear fit: R2=0.03, p=ns
Quad fit: R2=0.50, p=0.01
Mean weight.estimates
Mean methylation


a
Eaa11
Eaa19
-log10(p)
Chromosome
b
Chr11
92.75
BB
DD
EAA
EAA, pan
EAA, liver
dev.EAA, pan
dev.EAA, liver
int.EAA, liver
CD
HFD
c
Chr19
38.65
BB
DD
EAA
EAA, pan
EAA, liver
dev.EAA, pan
dev.EAA, liver
int.EAA, pan
CD
HFD


a Liver transcriptome
b Liver proteome


a
Density
Beta-value
0
1
2
3
0
0.25
0.50
0.75
1.00
b
Variance explained
PCs
0
0.1
0.2
c
PC2 (14%)
PC1 (19%)
Batch
1
2
3
4
-10
-5
0
5