Genetic Analyses of Epigenetic Predictors that Estimate Aging, Metabolic Traits, and Lifespan

Khyobeni Mozhui [1, 2*], Ake T. Lu[3], Caesar Z Li [3], Amin Haghani [4], Jose Vladimir Sandoval-Sierra [1], 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

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

Abstract

DNA methylation (DNAm) clocks are accurate molecular biomarkers of aging. However, the clock mechanisms remain unclear. Here, we used a pan-mammalian microarray to assay DNAm in liver from 339 predominantly female mice belonging to the BXD family. We computed epigenetic clocks and maximum lifespan predictor (predicted-maxLS), and examined associations with DNAm entropy, diet, weight, metabolic traits, and genetic variation. The epigenetic age acceleration (EAA) derived from the clocks, and predicted-maxLS were correlated with lifespan of the BXD strains. Quantitative trait locus (QTL) analyses uncovered significant QTLs on chromosome (Chr) 11 that encompasses the *Erbb2/Her2* oncogenic region, and on Chr19 that contains a cytochrome P450 cluster. Both loci harbor candidate genes associated with EAA in humans (*STXBP4*, *NKX2*-3, *CUTC)*. Transcriptome and proteome analyses revealed enrichment in oxidation-reduction, metabolic, and mitotic genes. Our results highlight loci that are concordant in human and mouse, and demonstrate intimate links between metabolism, body weight, and epigenetic aging.

Keywords: epigenetic clock, lifespan, aging, DNA methylation, QTL, weight, diet

Introduction

Epigenetic clocks are widely used molecular biomarkers of aging[1]. These biological clocks are based on the methylation status across an ensemble of "clock CpGs" that are collectively used to derive a DNA methylation (DNAm) based estimate of 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 is a measure of the rate of biological aging. Denoted as epigenetic age acceleration (EAA), a more accelerated measure (positive EAA) suggests an older biological age. While DNAmAge predicts age, its age-adjusted counterpart, EAA, is associated with health, fitness, exposure to stressors, body mass index (BMI), and even life expectancy[2-6].

DNAm clocks were initially reported for humans[7,8]. Since then, the age estimator has been extended to model organisms[9-11], and different variants of human clocks have also been developed. Some clocks are tissue specific, others are pan-tissue, and others perform well at predicting health and life expectancy[5,8,12-14].

A new microarray platform was recently developed to profile CpGs that have high conservation across mammalian clades. This pan-mammalian DNAm array (HorvathMammalMethylChip40) provides a common platform to measure DNAm, and has been used to build universal epigenetic clocks that can estimate age across a variety of tissues and mammalian species[15,16]. Another remarkable development with this array is the novel lifespan predictor that can estimate the maximum lifespan of over 190 mammals at high accuracy[17].

Here, we examine these novel clocks, lifespan predictor, and methylome entropy in a cohort of mice belonging to the BXD family that were maintained on either normal chow or high-fat diet (HFD)[18,19]. The BXDs are a well-established mouse genetic reference panel that were first created as a family of recombinant inbred (RI) strains by crossing two inbred progenitors: C57BL/6J (B6) and DBA/2J (D2). The family has been expanded to ~150 fully sequenced progeny strains[20,21]. Members of the BXD family vary greatly in their metabolic profiles, aging rates, and natural life expectancy[18,19,22-24]. The genetic variation, and the availability of accompanying deep -omic data make the BXDs a unique experimental population for dissecting the genetic modulators of epigenetic aging. Previously, we explored the aging methylome in a small number of BXD cases and found that HFD and higher body weight were associated with higher age-dependent changes in methylation[25]. In the present work, our goals were to (1) test the accuracy of the DNAm measures in predicting age, lifespan, and association with diet and metabolic characteristics, and (2) apply quantitative trait locus (QTL) mapping and gene expression analyses to uncover loci and genes that contribute to these DNAm biomarkers.

Our results are consistent with a faster clock for cases on HFD, and with higher body weight. Both the DNAmAge and lifespan predictors were correlated with the genotype-dependent life expectancy of female BXDs. We report QTLs on chromosomes (Chrs) 11 and 19. A strong candidate gene in the chromosome (Chr) 11 interval (referred to as *Eaaq11*) is *Stxbp4*, a gene that has been consistently associated with EAA by human genome-wide association studies (GWAS)[26-28]. The Chr19 QTL (*Eaaq19*) also harbors strong contenders including *Cyp26a1*, *Myof*, *Cutc*, and *Nkx2–3*, and the conserved genes in humans have been associated with longevity and

EAA[28-30]. *Eaaq19* may also have an effect on body weight change with age. We performed gene expression analyses to clarify the physiology associated with the DNAm traits, and this, perhaps unsurprisingly, highlighted metabolic networks as strong expression correlates of epigenetic aging.

Results

Description of samples

The present study uses liver DNAm data from 339 predominantly female mice (18 males only) belonging to 45 isogenic members of the BXD family, including F1 hybrids, and both parental strains. 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 the HFD (see Roy et al for details [18]). Tissues were collected at approximately six months intervals (see Williams et al. [19]). Individual-level data of cases used in this study are in **Data S1**.

Correlation with chronological age

For biological age prediction, three different types of mouse DNAm clocks were computed, each as a pair: liver-specific, and pan-tissue (**Table 1**). These are: (1) a general DNAm clock (referred to simply as DNAmAge): clock trained without pre-selecting for any specific CpG subsets; (2) developmental clock (dev.DNAmAge): built from CpGs that change during development; and (3) interventional clock (int.DNAmAge): built from CpGs that change in response to aging related interventions such as caloric restriction, HFD, or dwarfing alleles[9,11,25]. These clocks were trained either in an independent mouse dataset that did not include the BXDs and were therefore unbiased to BXD characteristics (unbiased mouse clocks), or trained in a subset of the BXD CD mice and used to estimate age in the full BXD cohort (BXD-biased

Table 1. Epigenetic predictors and correlation with chronological age and entropy

Clock type	Training set[1]	DNAmAge name	Tissue	r with age (n=339)	r with entropy (n=339)
Standard	unbiased	DNAmAge	pan	0.89	0.50
Standard	unbiased	DNAmAge	liver	0.92	0.51
Developmental	unbiased	dev.DNAmAge	pan	0.87	0.46
Developmental	unbiased	dev.DNAmAge	liver	0.91	0.45
Interventional	unbiased	int.DNAmAge	pan	0.85	0.38
Interventional	unbiased	int.DNAmAge	liver	0.86	0.49
Standard	BXD	DNAmAge	pan	0.93	0.55
Standard	BXD	DNAmAge	liver	0.97	0.54
Developmental	BXD	dev.DNAmAge	pan	0.95	0.51
Developmental	BXD	dev.DNAmAge	liver	0.96	0.51
Interventional	BXD	int.DNAmAge	pan	0.89	0.43
Interventional	BXD	int.DNAmAge	liver	0.94	0.53
Universal	unbiased	univ.DNAmAge	pan	0.92	0.59
Lifespan	unbiased	Predicted-maxLS	pan	-0.07 (ns)	-0.13 (p=0.02)

[1] *Unbiased* denotes that the clocks were trained in a completely independent dataset.

clocks). In addition to the mouse clocks, we estimated DNAmAge using the universal mammalian clock (univ.DNAmAge)[15]. The clocks performed well in age estimation (**Table 1; Fig 1a**). The EAA derived from these clocks showed wide individual variation (**Fig 1b**), but the EAA values are uncorrelated with chronological age.

We used the universal maximum lifespan predictor[17] to estimate the potential maximum lifespan (predicted-maxLS) of mice. Predicted-maxLS was uncorrelated with chronological age (**Table 1**), and this is expected since the chronological age represents the time when the biospecimens were collected; not the time of natural demise. Instead, the predicted-maxLS showed an overall inverse correlation with EAA from the different clocks, and this suggests higher age-acceleration for mice with lower predicted-maxLS (**Data S2**).

Association with methylome entropy

The methylome-wide entropy provides a measure of randomness and information loss, and this increased with chronological age (**Fig 1c**)[7]. As direct correlates of chronological age, all the DNAmAge were positively correlated with entropy (**Table 1**).

We hypothesized that higher entropy levels will be associated with (a) higher EAA, and (b) lower predicted-maxLS. Indeed, the univ.EAA had a significant positive correlation with entropy that was significant regardless of diet (**Fig 1d**). However, the EAA from the unbiased mouse clocks showed only weak correlations with entropy (**Data S2**). Entropy had a modest negative correlation with predicted-maxLS primarily in the CD group (**Fig 1e**). Taken together, our results indicate that discordance in the methylome increases with age, and is higher with higher univ.EAA. Mice with shorter predicted-maxLS may also had slightly higher entropy.

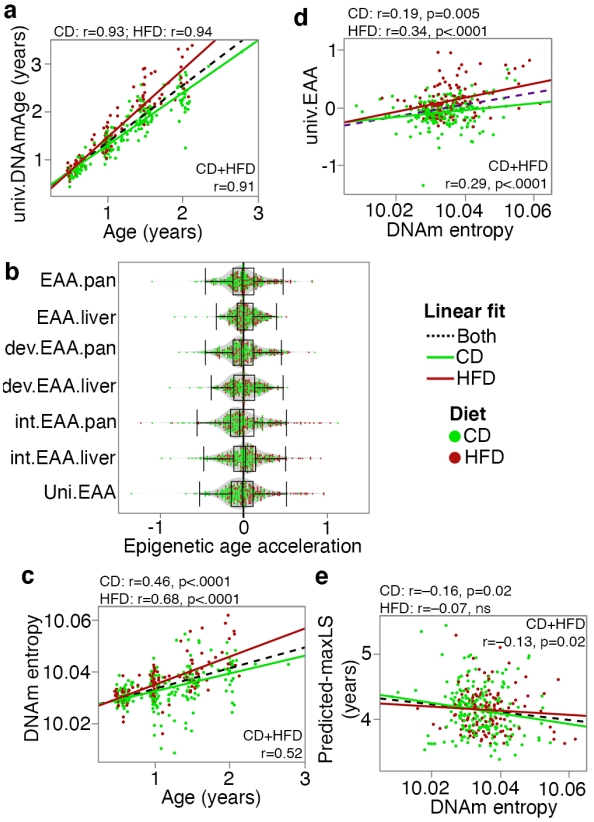


Fig 1. DNA methylation readouts and inter-correlations

(a) Correlation between predicted age (universal clock), and known chronological age. Mice on control diet depicted by green dots (CD; normal lab chow; n = 210); mice on high fat diet (HFD; n = 129) by red dots. **(b)** Violin plots of epigenetic age acceleration (EAA) derived from different DNAmAge clocks. **(c)** Shannon entropy, calculated from the full set of high quality CpGs, increases with age. **(d)** DNA methylation entropy has a direct correlation with EAA derived from the universal clock. **(e)** There is a slight inverse correlation between the entropy and DNA methylation based predicted maximum lifespan (predicted-maxLS).

How the epigenetic readouts relate to diet, body weight, and sex

Diet. EAA from most of the clocks, including the universal clock, were significantly higher in the HFD (**Table 2**). Entropy was also significantly higher in the HFD group. The maxLS did not differentiate between diets **(Table 2**).

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]
Unbiased DNAm 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	
BXD biased DNAm clocks	EAA.BXD, pan	CD	-0.06 ± 0.15	<.0001	0.08	ns	0.17	0.01	0.26	0.22
		HFD	0.09 ± 0.19		0.17	0.05	0.37	<.0001	0.42	
	EAA.BXD, liver	CD	0 ± 0.11	0.01	-0.04	ns	0.06	ns	0.18	0.37
		HFD	0.03 ± 0.11		0.18	0.04	0.41	<.0001	0.34	
	dev.EAA.BXD, pan	CD	-0.03 ± 0.13	<.0001	-0.08	ns	0	ns	0.25	0.43
		HFD	0.05 ± 0.16		0.13	ns	0.40	<.0001	0.49	
	dev.EAA.BXD, liver	CD	-0.01 ± 0.13	0.002	0.07	ns	0.11	ns	0.26	0.45
		HFD	0.03 ± 0.12		0.21	0.02	0.47	<.0001	0.40	
	int.EAA.BXD, pan	CD	-0.05 ± 0.19	<.0001	0	ns	0.18	0.01	0.19	0.72
		HFD	0.07 ± 0.3		0.20	0.03	0.42	<.0001	0.41	
	int.EAA.BXD, liver	CD	-0.03 ± 0.16	<.0001	-0.04	ns	0.06	ns	0.39	0.78
		HFD	0.09 ± 0.16		0.23	0.01	0.60	<.0001	0.40	
Universal clock	univ.EAA	CD	-0.08± 0.22	<.0001	0	ns	0.11	ns	0.37	0.67
		HFD	0.13 ± 0.27		0.35	<.0001	0.50	<.0001	0.43	
Entropy	-	CD	10.034 ± 0.007	0.004	-0.15	0.03	-0.34	<.0001	0.39	0.20 (ns)
		HFD	10.036 ± 0.007		-0.11	0.21	0	ns	0.23	
Pred-maxLS	-	CD	4.14 ± 0.37	ns	0.03	ns	0.05	ns	0.66	0.89
		HFD	4.15 ± 0.32		-0.06	ns	-0.11	0.20	0.70	

[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, with the exception of 48 cases that were first weighed 1 or 3 days after HFD (**Data S1**). In the CD group, only the unbiased EAA (pan-tissue) and dev.EAA (liver) showed 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 (**Data S2**). This was particularly strong for the unbiased interventional clocks (r = 0.45, p = 0.001 for int.EAA, pan-tissue; r = 0.58, p < 0.0001 for int.EAA, liver), and for the universal clock (**Fig 2a**). 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 (**Fig 2b**; **Data S2**). The final body weight (BWF) was measured at the time of tissue harvest, and EAA from all the unbiased clocks were significant correlates of BWF on both diets **(Table 2**).

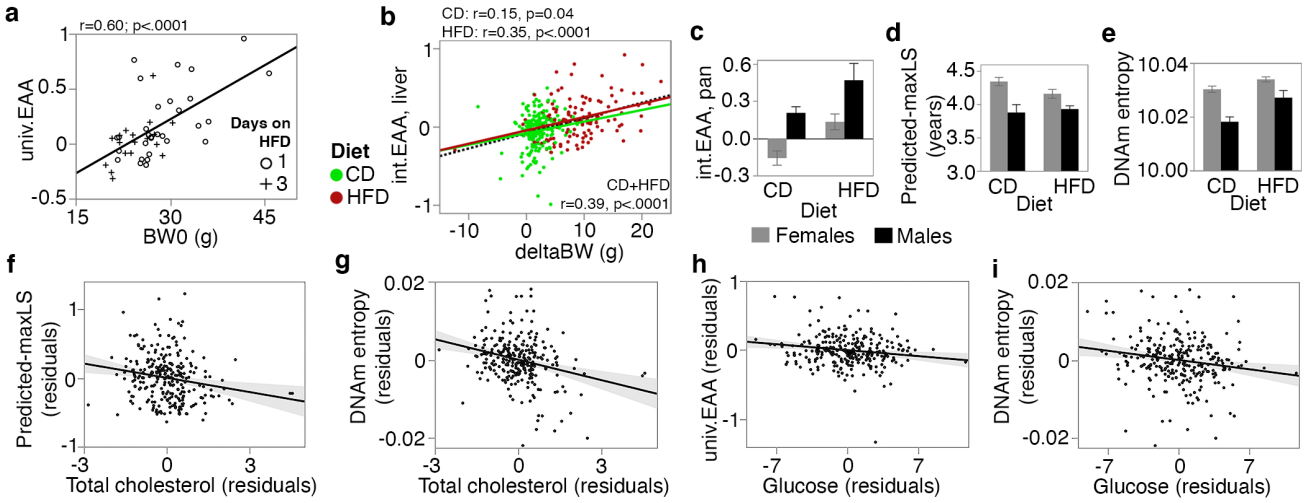


Fig 2. Correlates and modifiers of the epigenetic readouts

(a) For 48 mice, initial body weight (BW0) was measured 1 or 3 days after high fat diet (HFD), and these showed significant correlation between BW0 and epigenetic age acceleration (EAA shown for the universal clock). **(b)** Weight was first measured at mean age of 4.5 ± 2.7 months, and then at 6.3 ± 2.8 months (BW1). Weight gains during this interval (deltaBW = BW1 – BW0) is a direct correlate of EAA derived from the interventional clock in both normal chow (control diet or CD; n = 210) and HFD mice (n = 128). **(c)** For the BXD genotypes with samples from both males and females, males have higher age acceleration and this sex effect is highly significant for the pan-tissue interventional clock (int.EAA). Bars represent mean ± standard error; 40 females (26 CD, 14 HFD) and 18 males (10 CD, 8 HFD). Males had **(d)** lower predicted maximum-lifespan, and **(e)** lower average methylome entropy. **(f–i)** The residual plots display the direction of association between metabolic traits and DNAm readouts (n = 276 cases with metabolic data). After adjusting for chronological age, diet and body weight, serum cholesterol has inverse associations with **(f)** predicted maximum lifespan (p = 0.002), and **(g)** methylome entropy (p = 9.1E-06). Serum glucose level has inverse associations with **(h)** epigenetic age acceleration derived from the universal clock (p = 0.005), and **(i)** methylome entropy (p = 0.003).

Somewhat unexpected, entropy had an inverse correlation with body weight. This effect was primarily in the CD mice **(Table 2)**. We found no association between predicted-maxLS and the body weight traits (**Table 2**).

Sex effect. Four BXD genotypes (B6D2F1, D2B6F1, BXD102, B6) had cases from both males and females. We used these to test for sex effects. All the unbiased mouse clocks showed significant age acceleration in male mice, and this effect was particularly strong for the pan-tissue int.EAA (**Fig 2c; Data S2**). The predicted-maxLS was significantly lower in males (**Fig 2d**). Entropy on the other hand, was significantly higher in females (**Fig 2e**).

Association with metabolic traits

276 cases with DNAm data also had fasted serum glucose and total cholesterol[18,19], and we examined whether these metabolic traits are associated with the DNAm readouts. We applied regression analysis with age, diet and final body weight as covariates, and this showed significant effects of cholesterol on predicted-maxLS (p = 0.002), and entropy (p = 9E-06) (**Table S1**). To visualize how cholesterol levels associate with these, we plotted the residual values after the respective predictor and outcome variables were adjusted for age, diet, and BWF. The residual plot shows an inverse association between cholesterol and predicted-maxLS (**Fig 2f**). For entropy, similar to how it related with weight, higher cholesterol predicted lower entropy. Cholesterol had no significant association with univ.EAA **(Table S1)**.

Glucose had an unexpected inverse association with the univ.EAA that predicts lower age acceleration with higher glucose (p = 0.005) (**Fig 2h; Table S1**). Lower glucose also predicted higher entropy (p = 0.003) **(Fig 2i**). Glucose was not associated with predicted-maxLS **(Table S1)**.

Association with strain longevity

We next obtained longevity data from a parallel cohort of female BXD mice that were allowed to age on CD or HFD [18]. We evaluated whether the DNAm readouts were informative of strain-level lifespan. 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 ≥ 5, we computed the minimum (minLS), 25th quartile (25Q-LS), mean, median lifespan, 75th quartile (75Q-LS), and maximum lifespan (maxLS) (**Data S1**). Specifically, we postulated (a) an accelerated clock for strains with shorter lifespan (i.e., inverse correlation), (b) a direct correlation between predicted-maxLS and observed lifespan, and (c) higher entropy with shorter lifespan.

Overall, the EAA measures showed the expected inverse correlation trend with the lifespan summaries, and this was highly significant for the universal clock (**Table S2**; **Fig 3a**,**b**). For the mouse clocks, this effect was significant for the liver int.EAA (**Table S2**). When separated by diet, these correlations became weaker, but the negative trend remained consistent.

The DNAm entropy had an inverse correlation trend with strain lifespan (**Table S2**). This was nominally significant only for the strain maxLS when CD and HFD groups were combined (r = –0.13, p = 0.02) but became non-significant when separated by diet.

The predicted-maxLS showed a positive correlation trend with the lifespan summaries, and this was significant for the observed strain maxLS (**Fig 3d**). When separated by diet, the predicted-maxLS remained a significant correlate of strain maxLS only in the CD group.

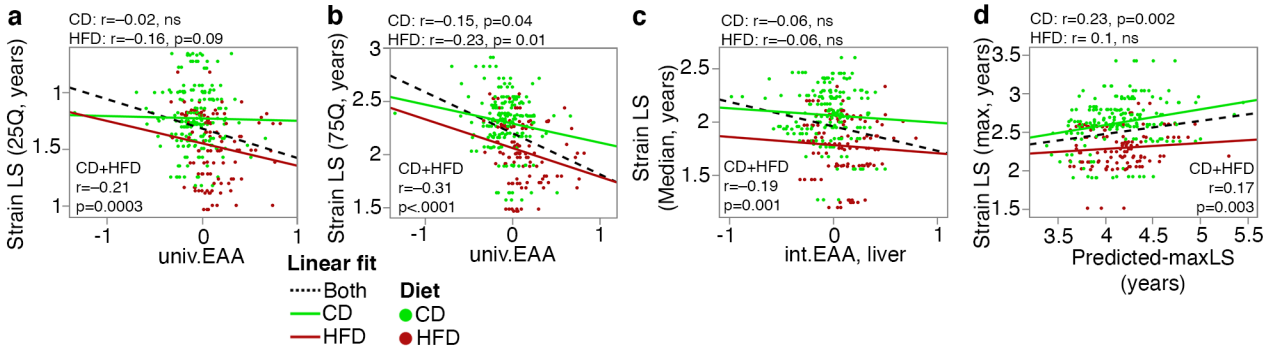


Fig 3. Predictors of strain longevity

BXD genotypes with shorter life expectancy tended to have a more accelerated universal clock (univ.EAA). This inverse relation is depicted for the **(a)** 25th and **(b)** 75th quartile age at natural death for female BXDs kept on either normal chow (CD) or high fat diet (HFD). **(c)** Age acceleration from the liver interventional unbiased mouse clock (int.EAA) also showed a similar inverse correlation with strain longevity, but effect was significant only when both diets were included (here illustrated with median lifespan). **(e)** The predicted maximum lifespan had a significant direct correlation with the observed strain maximum lifespan. Analysis in n = 302 female BXDs; 191 CD and 111 HFD.

Genetic analysis of epigenetic age acceleration and predicted-maxLS

The EAA traits had modest to high heritability, and averaged at 0.50 for the unbiased mouse clocks (**Table 2**). The predicted-maxLS had heritability of 0.66 on CD, and 0.70 on HFD. Another way to gauge level of genetic correlation is to compare between members of strains maintained on different diets. The EAA from the unbiased and universal clocks, and predicted-maxLS had high strain-level correlations between diets that indicates an effect of background genotype that is robust to dietary differences (**Table 2**). The genotype correlations were slightly lower for the BXD-biased clocks.

To uncover genetic loci, we applied QTL mapping using mixed linear modeling that corrects for the BXD kinship structure[31]. First, we performed the QTL mapping for each of the unbiased mouse and universal clocks, with adjustment for diet and body weight. EAA from the two interventional clocks had the strongest QTLs (**Data S3**). The pan-tissue int.EAA had a significant QTL on Chr11 (90–99 Mb) with the highest linkage at ~93 Mb (p = 3.5E-06; equivalent to a LOD score of 4.7) (**Fig 4a**). Taking a genotype marker at the peak interval (BXD variant ID DA0014408.4 at Chr11, 92.750 Mb)[20], we segregated the BXDs homozygous for either the D2 (*DD*) or the B6 (*BB*) alleles. The *DD* genotype had a significantly more accelerated int.EAA (**Fig 4a** inset). The liver int.EAA had the peak 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) (**Fig 4b**). We selected a marker at the peak interval (rs48062674 at Chr19, 38.650 Mb), and the *BB* genotype had significantly higher int.EAA compared to *DD* (**Fig 4b** inset). The QTL map for the univ.EAA did not reach genome-wide significance (**Fig 4c**). However, there were nominally significant peaks at the Chr19 (p = 0.0004), and Chr11 (p = 0.004) intervals.

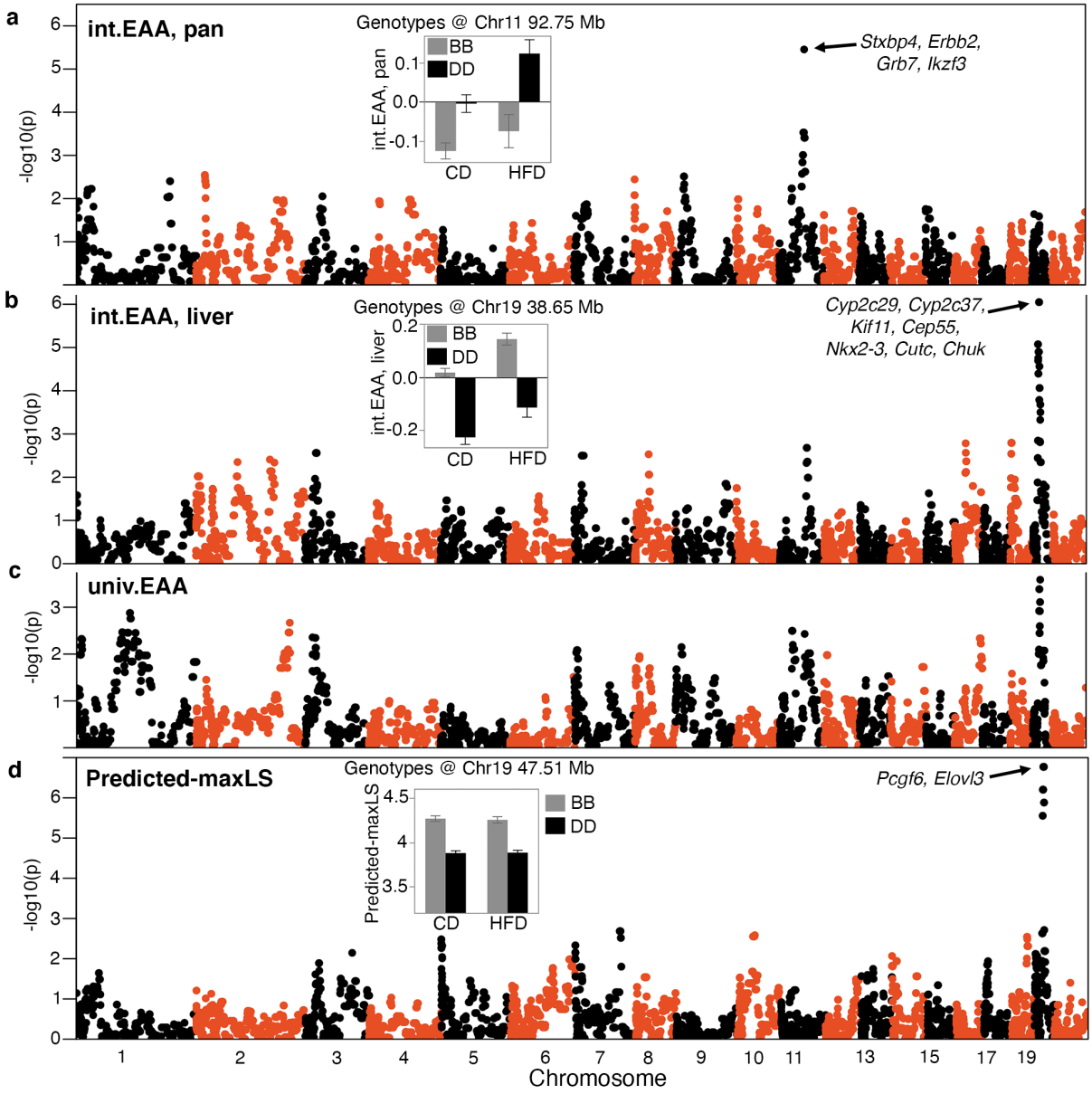


Fig 4. Genetic linkage analysis

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). **(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. **(c)** The linkage statistics are weaker for the EAA derived from the universal clock (univ.EAA). However, there are consistent nominally significant peaks on the Chr11 and Chr19 loci. **(d)** The DNA methylation based predicted maximum lifespan also maps to Chr19, but the peak markers are at ~47.5 Mb.

We next performed QTL mapping for DNAm entropy with adjustment for major covariates (diet,

chronological age, and body weight). No locus reached genome-wide significance (**DataS3**). There were modest QTLs on Chrs11 and 19. However, the Chr11 region is slightly distal to the markers linked to the EAA traits (minimum p = 0.009 at Chr11, ~103.7 Mb). The Chr19 locus somewhat overlapped the QTL for EAA, but the peak marker (minimum p = 0.0009) is slightly distal at ~48 Mb (**Data S3**).

The predicted-maxLS had a significant QTL on Chr19 (**Fig 4d; Data S3**) with the peak markers between 44–48 Mb (p = 2E-07; LOD score of 5.9). This overlaps the EAA QTL, but the peak markers are also distal (rs30567369 at 47.510 Mb). At this locus, mice with the *BB* genotype had significantly higher predicted-maxLS (**Fig 4d** inset).

Consensus QTLs for epigenetic age acceleration

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 for the unbiased mouse and universal EAA traits[32]. The peaks on Chrs 11 and 19 attained the highest consensus p-values (**Fig S1a; Data S3**). Additional consensus peaks (at $-\log_{10}$meta-p > 6) were observed on Chrs 1 (~152 Mb), and 3 (~54 Mb).

We focus on the Chrs 11 and 19 QTLs and refer to these as *EAA QTL on Chr 11* (*Eaaq11*), and *EAA QTL on Chr 19* (*Eaaq19*). *Eaaq11* extends from 90–99 Mb. For *Eaaq19*, we delineated a broader interval from 35–48 Mb that also encompasses the peak markers for the predicted-maxLS, albeit these may be separate loci related to EAA (~39 Mb of *Eaaq19*), and predicted-maxLS (~47 Mb of *Eaaq19*).

We performed marker-specific linkage analyses for each of the unbiased mouse and universal 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 *Eaaq11* 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 (**Fig S1b**). The effect of this locus appeared to be higher for the pan-tissue clocks compared to the corresponding liver-specific clocks. This marker in *Eaaq11* was not associated with either entropy or predicted-maxLS.

For proximal *Eaaq19*, the representative marker (rs48062674) was associated with all the EAA traits and the *BB* mice had higher age acceleration on both diets (**Fig S1c**). This marker was not associated with entropy, and had only a weak effect on predicted-maxLS (**Table 3**). When we performed the same analysis with the marker on distal *Eaaq19* (rs30567369), the association with EAA became weaker, and the association with predicted-maxLS became much stronger (**Table 3**). This suggests that the proximal part of *Eaaq19* is related to EAA while the distal part is related to predicted-maxLS.

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.

Association of EAA QTLs with body weight trajectory

Table 3: Marker specific linkage analyses for epigenetic age acceleration, predicted maximum lifespan, and body weight trajectory

Predictor	Outcome	Estimate	Std Error	t Ratio	p
Eaaq11 DA0014408.4[DD] Chr11, 92.750 Mb (133 *BB* cases, and 173 *DD* cases)	EAA, pan	0.096	0.023	4.184	3.8E-05
	EAA, liver	0.067	0.017	3.880	0.0001
	dev.EAA, pan	0.077	0.025	3.041	0.003
	dev.EAA, liver	0.037	0.020	1.878	0.06
	int.EAA, pan	0.153	0.029	5.278	2.5E-07
	int.EAA, liver	-0.033	0.025	-1.284	0.20
	univ.EAA	0.101	0.025	4.057	6.3E-05
Eaaq19 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
	univ.EAA	-0.145	0.029	-4.932	1.3E-06
	Pred-maxLS	-0.100	0.048	-2.086	0.04
Distal Eaaq19 rs30567369[DD] Chr19, 47.510 Mb (198 *BB* cases, and 106 *DD* cases)	int.EAA, liver	-0.079	0.026	-2.995	0.003
	univ.EAA	-0.053	0.026	-2.012	0.05
	Pred-maxLS	-0.383	0.036	-10.781	3.9E-23
Mixed model for longitudinal change in body weight					
Predictor	**Outcome**	**Estimate**	**Std Error**	**t Ratio**	**p**
Eaaq11 DA0014408.4[DD] Number of observations = 6885; number of individuals = 2112	Body weight	0.619	0.345	1.794	0.07
Eaaq19 rs48062674[DD] Number of observations = 6132; number of individuals = 1852	Body weight	-1.847	0.374	-4.945	7.6E-07
Distal Eaaq19 rs30567369[DD] Number of observations = 6059; number of individuals = 1802	Body weight	-1.619	0.363	-4.458	8.3-06

Since body weight gains was an accelerator of the clocks, we examined whether the selected markers in *Eaaq11* and *Eaaq19* 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 *Eaaq11* in the CD group, the *DD* genotype in the CD group also had slightly higher mean weight at older adulthood (12 and 18 months; **Fig 5a**). However, this marker had no significant association with body weight when tested using a mixed effects model (p = 0.07; **Table 3**). In proximal *Eaaq19*, it was the *BB* genotype that exhibited consistently accelerated clock on both diets, and the *BB* genotype also had higher average body weight by 6 months of age (**Fig 5b**), and this locus had a significant influence on the body weight trajectory (p = 7.6E-07; **Table 3**). The nearby marker on distal *Eaaq19* also showed a similar pattern of association with body weight (**Table 3**).

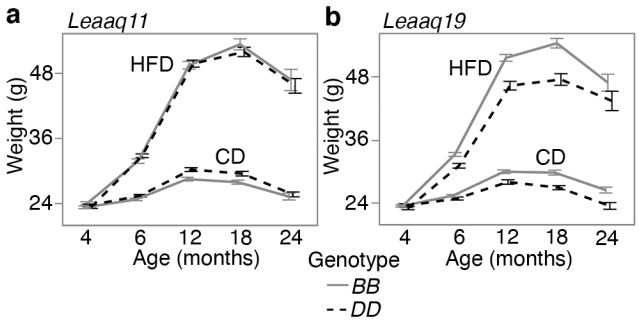


Fig 5. Body weight trajectory by diet and genotype

Body weight was measured at regular age intervals (x-axis) from **(a)** 2112 BXD mice that were homozygous at the *Leaaq11* marker (DA0014408.4; 842 *BB*, 1279 *DD*), and **(b)** 1852 BXD mice that were homozygous at the proximal *Leaaq19* marker (rs48062674; 1252 *BB*, 600 *DD*). Mice were maintained on either control diet (CD) or high fat diet (HFD). The graphs show the segregation of body weight over time by diet and genotype. Mean ± standard error; heterozygotes were excluded.

Candidate genes for epigenetic age acceleration

There are several positional candidate genes in *Eaaq11* and *Eaaq19*. 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 a cis-acting expression QTL (eQTL). For the eQTL analysis, we utilized an existing liver transcriptome data from the same aging cohort[19]. We identified 24 positional candidates in *Eaaq11* that includes *Stxbp4*, *Erbb2* (*Her-2* oncogenic gene), and *Grb7* (growth factor receptor binding) (**Data S4**). *Eaaq19* 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 (**Data S4**).

For further prioritization, we converted the mouse QTL regions to the corresponding syntenic regions in the human genome, and retrieved GWAS annotations for these regions[33]. 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 *Eaaq11*, and 3 genes in *Eaaq19* (**Table S4**). We also identified a GWAS study that found associations between variants near *Myof-Cyp26a1* and human longevity[30], and a meta-GWAS that found gene-level associations between *Nkx2–3* and *Cutc*, and epigenetic aging[28] (**Table S4**).

Gene expression correlates of EAA and predicted max-LS

Liver RNA-seq data was available for 153 of the BXD cases that had DNAm data (94 CD, and 59 HFD)[19]. We used this set to perform transcriptome-wide correlation analysis for the univ.EAA. To gain insights into gene functions, we selected the top 2000 transcriptome correlates ($|r| \geq 0.37$, $p \leq 2.8E\text{-}06$; **Data S5**) for functional enrichment analysis. These top correlates represented transcript variants from 1052 unique genes and included a few positional candidates (e.g., *Ikzf3*, *Kif11*, *Cep55*, *Cyp2c29*, *Cyp2c37*). Only 62 transcripts from 36 unique genes were negatively correlated with univ.EAA, and this set was significantly enriched (Bonferroni correct $p < 0.05$) in oxidation-reduction, and metabolic pathways (**Data S6**; **Fig 6a**). These functional categories included the cytochrome genes, *Cyp2c29* and *Cyp2c37*, located in *Eaaq19*. This set was also highly liver specific. The positive correlates were enriched in a variety of gene functions, and

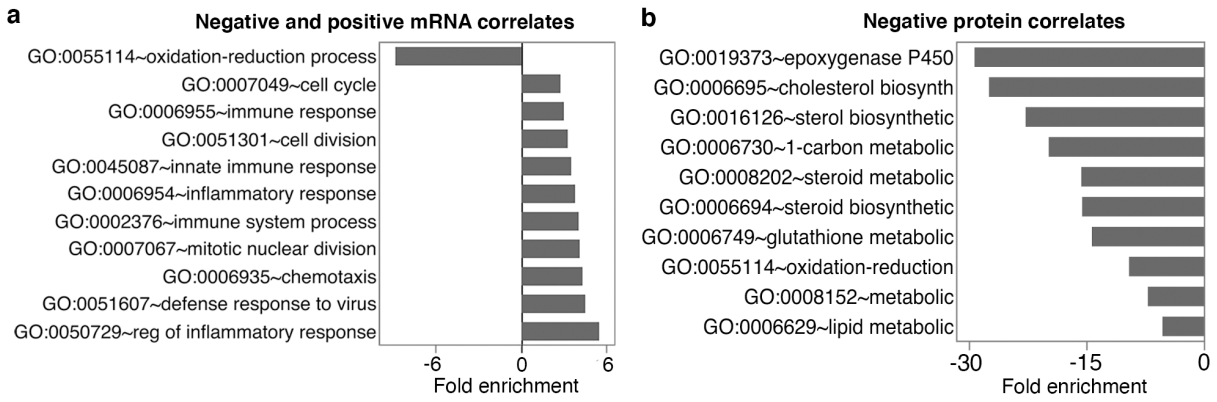


Fig 6. Gene expression correlates of the universal clock
(a) Graph displays the enriched gene ontologies (GO) for the top 2000 transcriptome correlates of epigenetic age acceleration (univ.EAA). The y-axis shows the fold enrichment; positive values are for the positive correlates, and negative values are for the negative correlates of univ.EAA. The smaller set of negative mRNA correlates is enriched in oxidation-reduction process. For positive mRNA correlates, the top 10 enriched GOs highlight immune response, cell cycle, and mitosis. **(b)** The top 10 negative proteome correlates of univ.EAA is also enriched in oxidation-reduction and related metabolic pathways, and these are populated by several cytochrome P450 genes.

was not a liver-specific gene set (**Data S6**). Taking the top 10 GO categories, we can broadly discern two functional domains: immune and inflammatory response, and mitosis and cell cycle (**Fig 6a**). To verify that these associations are robust to the effect of diet, we repeated the correlation and enrichment analysis in the CD group only (n = 94). Again, taking the top 2000 correlates ($|r| \leq 0.30$; $p \leq 0.003$), we found the same enrichment profile for the positive and negative correlates (**Data S6**).

Next, we performed the correlational analysis using liver proteomic data that was available for 164 of the BXDs. The proteome data quantifies over 32000 protein variants from only 3940 unique genes[19]. We took the top 2000 protein correlates of univ.EAA ($|r| \geq 0.27$, $p \leq 6.0E\text{-}04$) (**Data S7**). This represented protein levels from 563 unique genes. 1139 protein variants (215 genes) had negative correlations, and similar to the mRNA correlates, there was enrichment in oxidation-reduction and metabolic processes. This set was also enriched in liver genes, and

included pathways related to lipid and steroid metabolism, epoxygenase p450 pathway, and xenobiotics (**Fig 6b**; **Data S8**). These categories were populated by the cytochrome genes including candidates in *Eaaq19* (e.g., *Cyp2c29*, *Cyp2c37*). The positive proteome correlates showed a different functional profile than the transcriptomic set. These were enriched in genes related to transport (includes apolipoprotein such as APOE), cell adhesion, protein translation, protein folding, and metabolic pathways related to glycolysis and gluconeogenesis (**Data S8**).

We performed a similar transcriptome and proteome analysis for the predicted-maxLS. For mRNA, both the negative and positive correlates were enriched in metabolic pathways including glucose and lipid metabolism (**Data S9, S10**). Similarly, the positive and negative protein correlates of predicted-maxLS converged on oxidation-reduction processes (included cytochrome genes located in proximal *Eaaq19*) and metabolic pathways (**Data S11, S12**).

Discussion

The goal of this study was to examine the aging methylome, its correlates and modifiers, and potential genetic drivers. HFD had a strong age-accelerating effect that concurs with the association between EAA and obesity in humans[25,34,35]. Age-acceleration due to diet manifested within the first 1 to 3 days of transitioning from normal lab chow to HFD. Even among the CD mice, higher weight gain at a younger age was associated with an accelerated clock.

Somewhat surprising was how entropy related to the metabolic traits. Epigenetic entropy increases with age, and is likely an indicator of the level of stochastic noise that increases with time[7,36]. In biological systems, entropy is kept at bay by the uptake of energy, and investment in maintenance and repair[37]. As HFD increased entropy (possibly due to higher cellular heterogeneity and adiposity of liver tissue), we expected entropy to be higher with higher body weight. But instead, entropy had an inverse correlation with weight, an effect that was primarily in the CD mice. Higher levels of serum glucose and total cholesterol were also associated with lower entropy. The reason for this is unclear, and we can only speculate that the enhanced energy consumption in mice that had higher metabolic substrates may have kept the methylome in a more ordered state. Despite this, mice with higher entropy also tended to have higher EAA. Entropy had a modest negative correlation with not only the DNAm based predicted-maxLS, but also with the known strain-level maxLS. The predicted-maxLS on the other hand, showed no direct association with diet or body weight, but higher total cholesterol and EAA predicted shorter predicted-maxLS.

For the BXDs, life expectancy is highly dependent on the background genotype[18,22,24]. Similarly, the universal and interventional clocks were more accelerated in mice belonging to strains with shorter lifespan, and the predicted-maxLS also concurred with the observed strain maxLS. We note that the predicted-maxLS overestimated the strain max-LS by 0.7 to 3 years (median error of +1.6 years). Nonetheless, the correlation between individual-level predicted-, and strain-level observed maxLS is remarkable considering that both the universal clock and max-LS predictor are pan-mammalian, and species- and tissue-agnostic[17,38]. Our results suggest that these universal epigenetic predictors of biological aging, and lifespan are informative of the subtle and normative lifespan variation in a family of inbred mice. The analysis between the epigenetic readouts and lifespan was also 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. Nonetheless, this indirect comparison demonstrates that these epigenetic predictors capture genotype-dependent effects.

We tested different versions of the mouse DNAmAge clocks, and these appeared to capture slightly different aspects of epigenetic aging. For instance, the interventional clocks were sensitive to diet and early weight change, but not related to BW0 in the CD mice. Instead, BW0 had a significant accelerating effect on the liver specific developmental clock (dev.EAA).

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 *Eaaq11* 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[26-28]. 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[39], and has oncogenic activity and implicated in different cancers[40,41]. Furthermore, GWAS have also associated *STXBP4* with age of menarche[42,43]. *Eaaq11* corresponds to the 17q12-21 region in humans, and the location of additional oncogenic genes, e.g., *ERBB2/HER2*, *GRB7*, and *BRCA1*[44]. 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 *Eaaq11*, and the co-activation of *Erbb1*, *Grb7*, and *Nr1d1* has been linked to breast and other cancers[45,46].

Eaaq19 was consistently associated with EAA from all the clocks we evaluated, and also with body weight gains, irrespective of diet. The predicted-maxLS also maps to this region, and DNAm entropy may also have a weak association with markers at this interval. The EAA traits have peak markers in the proximal part of *Eaaq19* (around the cytochrome cluster), and the predicted-maxLS peaks in the distal portion (over candidates like *Elovl3*, *Pcgf3*). Two candidates in *Eaaq19* 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*)[28]. *Eaaq19* is also the location of the *Cyp26a1-Myof* genes, and the human syntenic region is associated with longevity, metabolic traits, and lipid profiles [30,47,48]. Another noteworthy candidate in *Eaaq19* is *Chuk*, a regulator of mTORC2, that has been associated with age at menopause[42,49]. Clearly, *Eaaq19* presents a complex and intriguing QTL related to the different DNAm readouts, and potentially metabolic traits. Both *Eaaq19* and *Eaaq11* exemplify the major challenge that follows when a genetic mapping approach leads to gene- and variant-dense regions [50,51]. 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 related to lipids, glucose, and proteins for both the univ.EAA and predicted-maxLS. Other enriched pathways were mitosis and cell division, and immune processes, but this was specific to the positive transcriptomic

correlates. The more compelling evidence is for the cytochrome P450 genes, which are both positional candidates, as well as expression correlates at the transcriptomic and proteomic levels. These genes have high expression in liver, and have major downstream impact on metabolism[52-54]. One caveat is that these CYP genes are part of a gene cluster in *Eaaq19* 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 [55]. Nonetheless, the cytochrome genes in *Eaaq19* are strong candidate modulators of EAA that calls for further investigation.

Aside from *Eaaq11* and *Eaaq19*, loci with evidence of consensus QTLs were also detected on Chrs 1 and 3. We do not delve into these in the present work, but the Chr3 interval is near genes associated with human epigenetic aging (*Ift80*, *Trim59*, *Kpna4*)[26,28]. 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 *Eaaq11* and *Eaaq19*, these other loci also present potentially important regions for EAA.

In summary, we have identified two main QTLs—*Eaaq11* and *Eaaq19*—that contribute to variation in two DNAm readouts: EAA, and predicted-maxLS. *Eaaq11* contains several genes with oncogenic properties (e.g., *Stxbp4*, *Erbb2*), while *Eaaq19* 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. The convergence of evidence from genetic and gene expression analyses suggests that genes involved in metabolism and energy balance may modulate the age-dependent restructuring of the methylome, and this may in turn, have an impact on the epigenetic predictors of aging and lifespan.

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 [18,19]. 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 [19]. 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, quality check, and entropy calculation

DNA samples from ~350 BXD mice were profiled on the Illumina HorvathHumanMethylChip40 array. Details of this array are described here[15].The array contains probes that target ~36K highly conserved CpGs in mammals. Over 33K probes map to homologous regions in the mouse genome, and data from these were normalized using the SeSame method [56]. Unsupervised hierarchical clustering was performed to identify outliers and failed arrays, and these 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. The top principal component (PC1 and PC2) segregated the samples by strain identity, and samples that did not cluster with 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.

For entropy calculation, we used 27966 probes that have been validated for the mouse genome using calibration data generated from synthetic mouse DNA[57]. Shannon entropy was calculated for each sample using the R package, "entropy" (v1.2.1) with method = "ML": maximum likelihood[58].

Clock estimation and maximum lifespan predictor

The development of the universal pan-tissue epigenetic clocks of age, and the universal maximum lifespan predictor are described in Lu et al[38], and Li et al.[17], respectively. For the present work, we utilized the universal clock that predicts relative age, defined as individual age relative to the maximum lifespan of its species, followed by inverse transformation to estimate DNAmAge[38]. The mouse specific clock were built using subsets of CpGs, and these will be described in a companion paper. Age acceleration (EAA) measures were defined as the residuals from regression of DNAm age on chronological age. By definition, EAA measures are independent of age.

Statistics

Statistical analyses between the epigenetic predictors and continuous variables (body weight, strain lifespan) were based on Pearson correlations, and t-test was used to evaluate the effect of categorical predictors (sex, diet).

Two metabolic traits were downloaded from the bioinformatics platform GeneNetwork 2 (GN2) [59]: (1) fasted serum glucose, and (2) fasted serum total cholesterol (more information on how to retrieve these data directly from GN2 are provided in **Data S13**). Association with metabolic traits was examined using multivariable linear regression (the R equations are provided in **Table S1**). For visualization, residuals for both the predictor and outcome variables were extracted after regressing on age, diet, and BWF using the R code: residuals(lm(~ age + diet + BWF)).

Longevity data (defined as age at natural death) was also downloaded from GN2 (**Data S13**)[18]. 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.

Genetic analyses

The broad sense heritability within diet was estimated as the fraction of variability that was explained by background genotype[20,60,61]. For this, we applied a simple 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, and these traits can be accessed from this website[59] (trait accession IDs provided in **Data S13**). 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[31], 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 (3720 markers as of early 2021). For the combined p-values, QTL mapping was done separately using GEMMA for each EAA derived from all the unbiased mouse and universal clocks. Fisher's p-value combination was then applied to get the meta-p-value[32]. We used this method to simply highlight loci that had consistent linkage across the different EAA measures. QTL mapping for entropy, major covariates—age, diet, BW1, and BWF—were included as co-factors. QTL mapping for predicted-maxLS was done without co-factors as age, weight, and diet were not significant covariates (including these do not change the results).

For marker specific linkage, we selected SNPs located at the peak QTL regions (DA0014408, rs48062674, rs30567369), and grouped the BXDs by their genotypes (F1 hybrids and other heterozygotes were excluded from this), and marker specific linkage was tested using ANOVA. rs48062674 and rs30567369 are reference variants that is already catalogued in dbSNP[62], 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[20]. Genotypes at these markers for individual BXD samples are in **Data S1**.

For marker specific QTL analysis for EAA, we performed linear regression using the data in **Data S1**. Heterozygotes at the respective markers were excluded, and we applied the following regression model for each of the unbiased mouse and universal EAA separately: lm(EAA ~ genotype + diet). To test the effect on body weight change, body weight data measured at approximately 4 (baseline), 6, 12, 18, and 24 months were downloaded from GN2 (**Data S13**). Detailed description of these weight data are in Roy et al[18]. We then applied a mixed effects regression model using the lme4 R package[63]: lmer(weight ~ age + diet + genotype + (1|ID)), where ID is the identifier for individual mouse.

Bioinformatic tools for candidate gene 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)[64-66]. 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[19]. 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*. This data was also used for the transcriptome-wide correlations analysis for univ.EAA in the 153 cases that had both DNAm and RNA-seq data. We considered the top 2000 highest correlated transcripts, 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 6.8) for GO enrichment analysis[67,68]. Similarly, 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)* [19].

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 catalogue[33], and retained the GWAS hits that were related to aging.

Data availability

The full microarray data will be released via NCBI's Gene Expression Omnibus upon official publication. Genome annotations of the CpGs can be found on Github https://github.com/shorvath/MammalianMethylationConsortium. Individual level BXD data are available on www.genenetwork.org on FAIR+ compliant format; data identifiers, and way to retrieve data are described in **Data S13**.

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

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

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

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

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

Reference:

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

2 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).

3 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).

4 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).

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

6 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).

7 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).

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

9 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).

10 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).

11 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).

12 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).

13 Levine, M. E. *et al.* An epigenetic biomarker of aging for lifespan and healthspan. *Aging* **10**, 573-591, doi:10.18632/aging.101414 (2018).

14 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).

15 Arneson, A. *et al.* A mammalian methylation array for profiling methylation levels at conserved sequences. *bioRxiv*, 2021.2001.2007.425637, doi:10.1101/2021.01.07.425637 (2021).

16 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).

17 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).

18 Roy, S. *et al.* Gene-by-environmental modulation of longevity and weight gain in the murine BXD family. *bioRxiv*, 776559, doi:10.1101/776559 (2020).

19 Williams, E. G. *et al.* Multi-Omic Profiling of the Liver Across Diets and Age in a Diverse Mouse Population. *bioRxiv*, 2020.2008.2020.222968, doi:10.1101/2020.08.20.222968 (2020).

20 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).

21 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).

22 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).

23 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).

24 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).

25 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).

26 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).

27 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).

28 McCartney, D. L. *et al.* Genome-wide association studies identify 137 loci for DNA methylation biomarkers of ageing. *bioRxiv*, 2020.2006.2029.133702, doi:10.1101/2020.06.29.133702 (2020).

29 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).

30 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).

31 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).

32 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).

33 Catalog, G. <https://www.ebi.ac.uk/gwas/> (

34 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).

35 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).

36 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).

37 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).

38 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).

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

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

41 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).

42 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).

43 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).

44 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).

45 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).

46 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).

47 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).

48 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).

49 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).

50 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).

51 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).

52 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).

53 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).

54 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).

55 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).

56 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).

57 Haghani, A. *et al.* DNA Methylation Networks Underlying Mammalian Traits. *bioRxiv*, 2021.2003.2016.435708, doi:10.1101/2021.03.16.435708 (2021).

58 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).

59 *GeneNetwork 2*, <http://genenetwork.org> (

60 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).

61 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).

62 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).

63 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*.

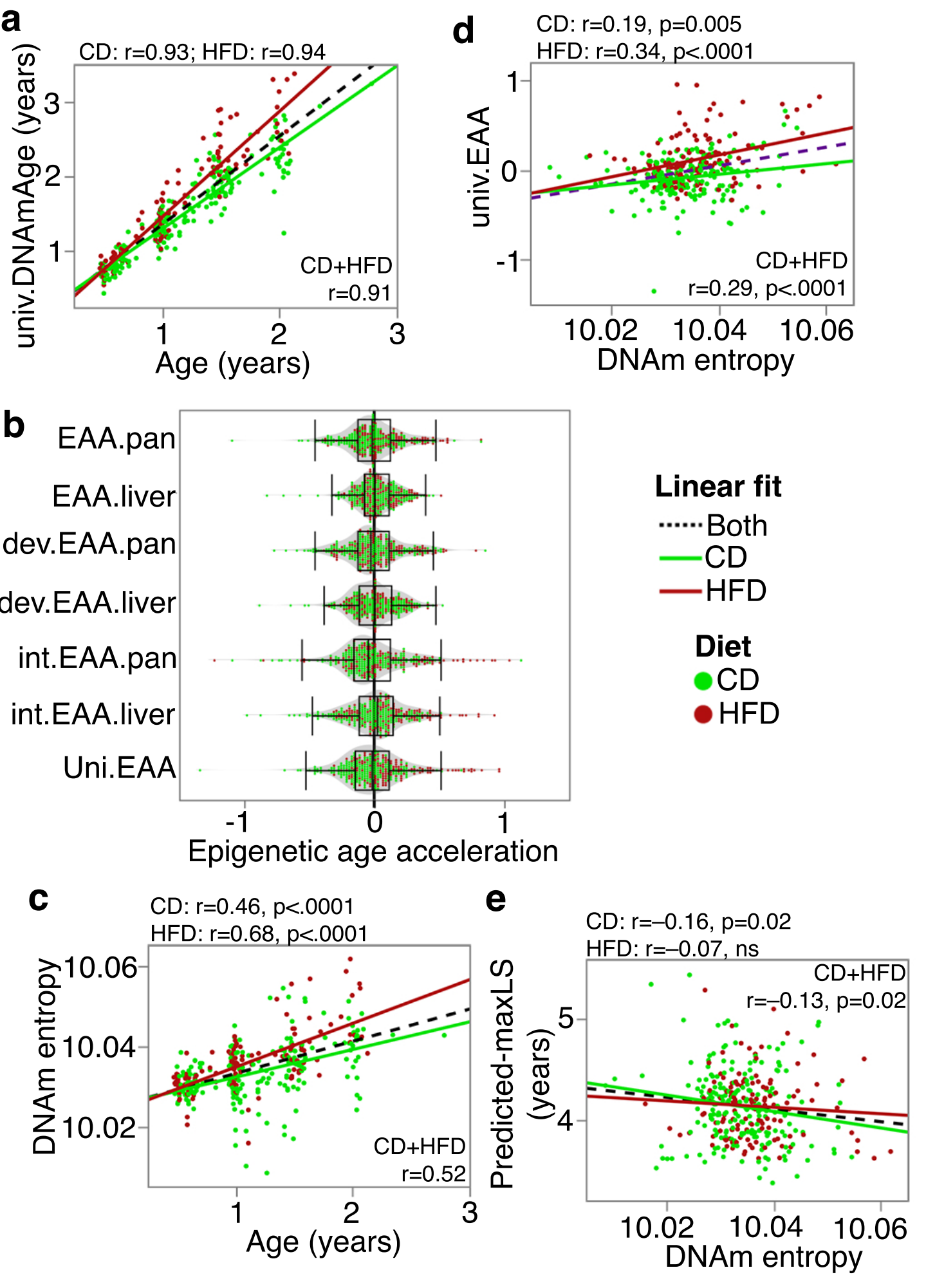
64 *Wellcome Sanger Institute Mouse Genome Project*, <https://www.sanger.ac.uk/sanger/Mouse_SnpViewer/rel-1505> (

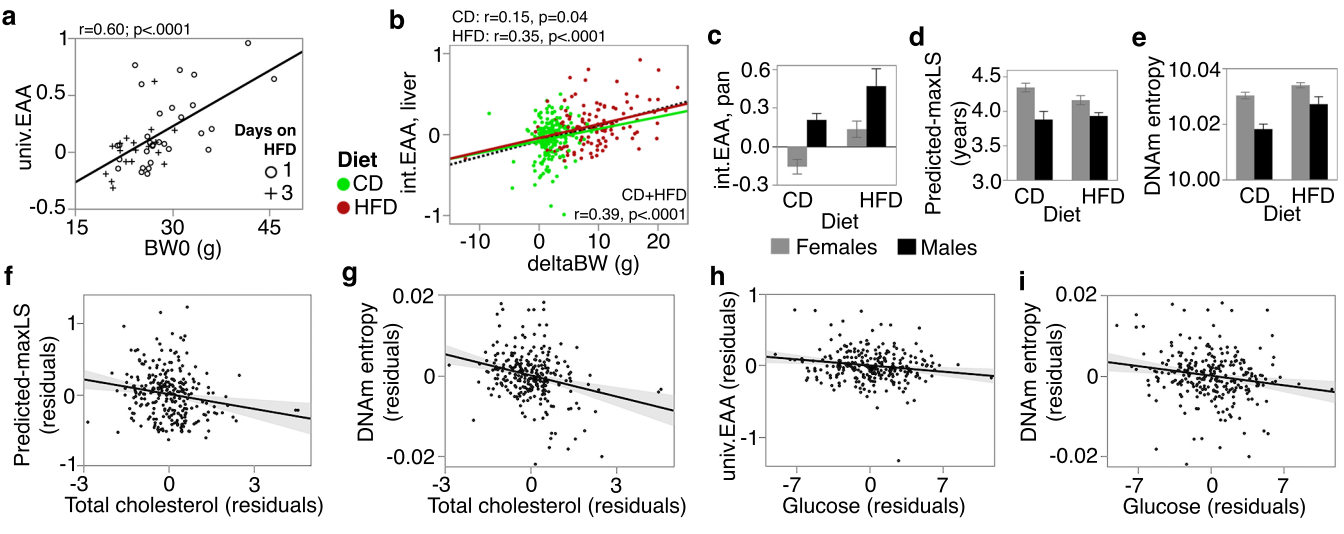
65 Keane, T. M. *et al.* Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature* **477**, 289-294, doi:10.1038/nature10413 (2011).

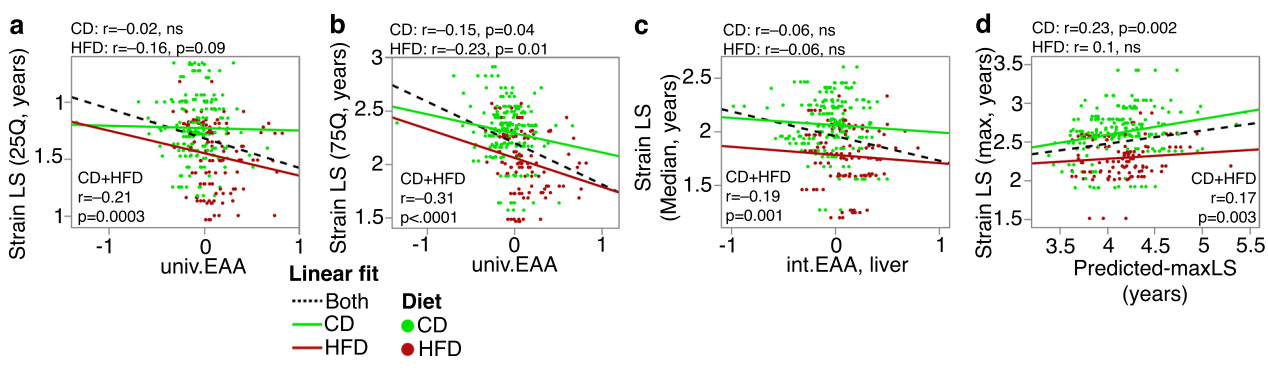
66 Yalcin, B. *et al.* Sequence-based characterization of structural variation in the mouse genome. *Nature* **477**, 326-329, doi:10.1038/nature10432 (2011).

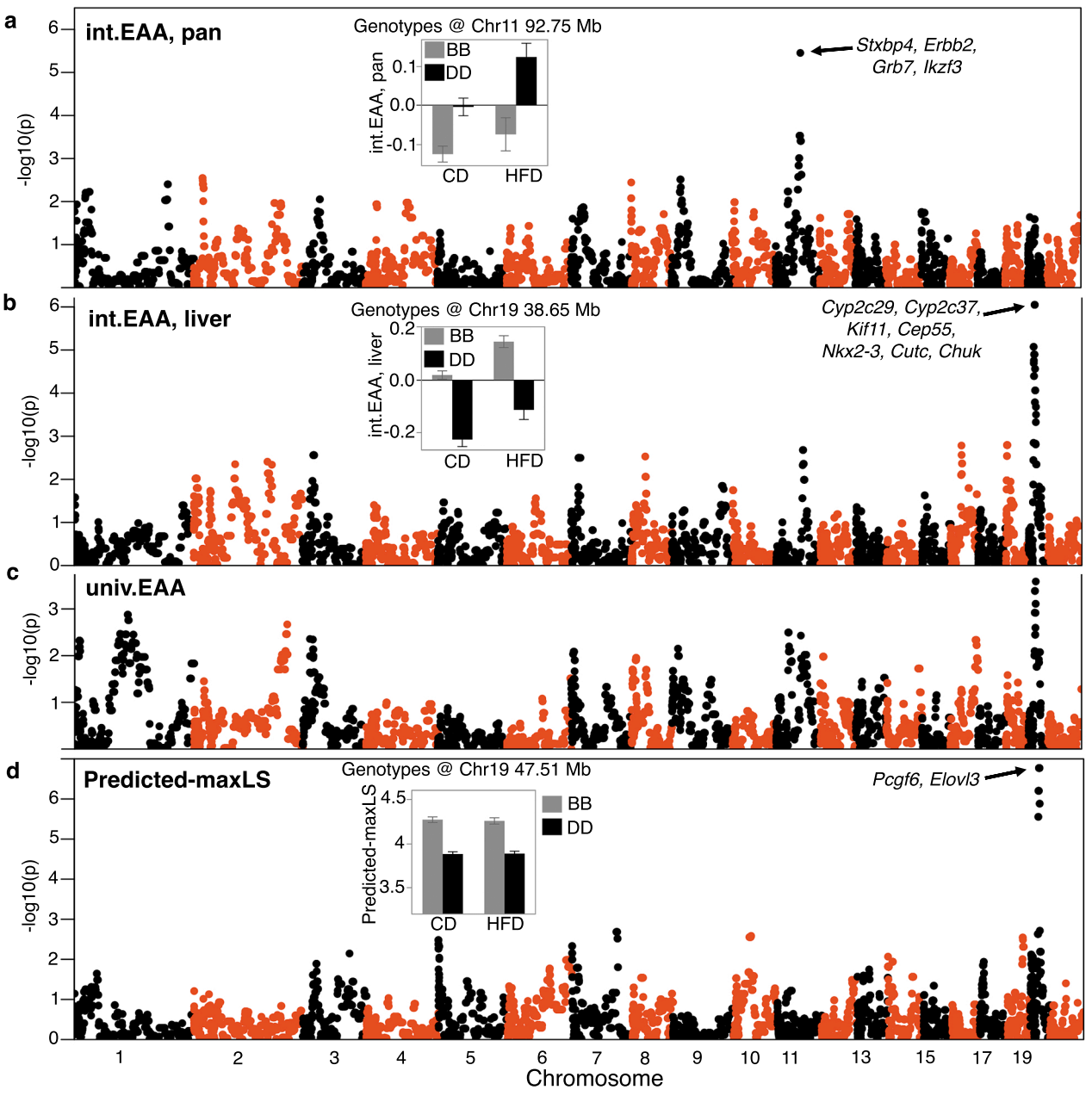
67 Huang da, W., Sherman, B. T. & Lempicki, R. A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic acids research* **37**, 1-13, doi:10.1093/nar/gkn923 (2009).

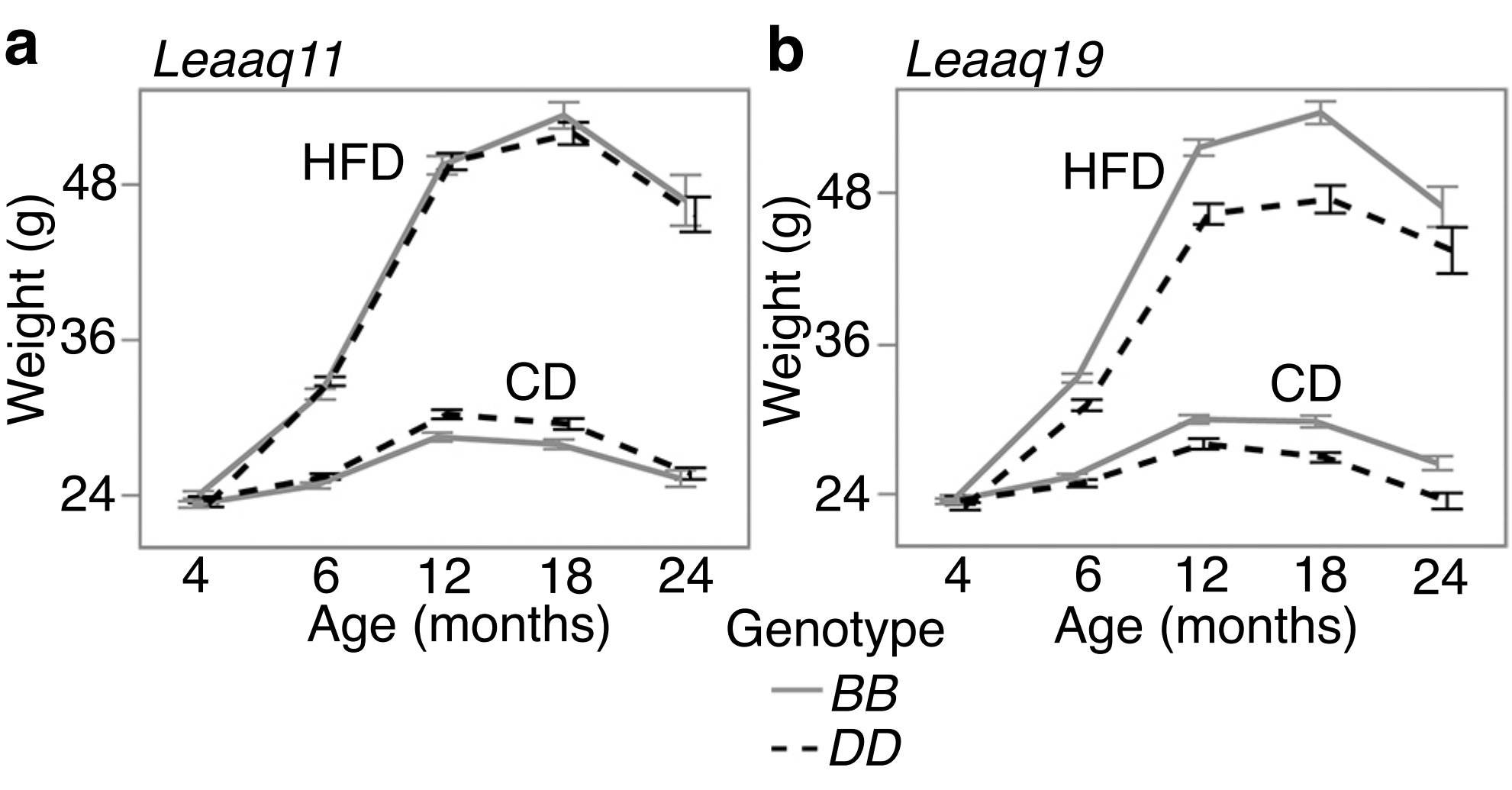
68 Huang da, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols* **4**, 44-57, doi:10.1038/nprot.2008.211 (2009).

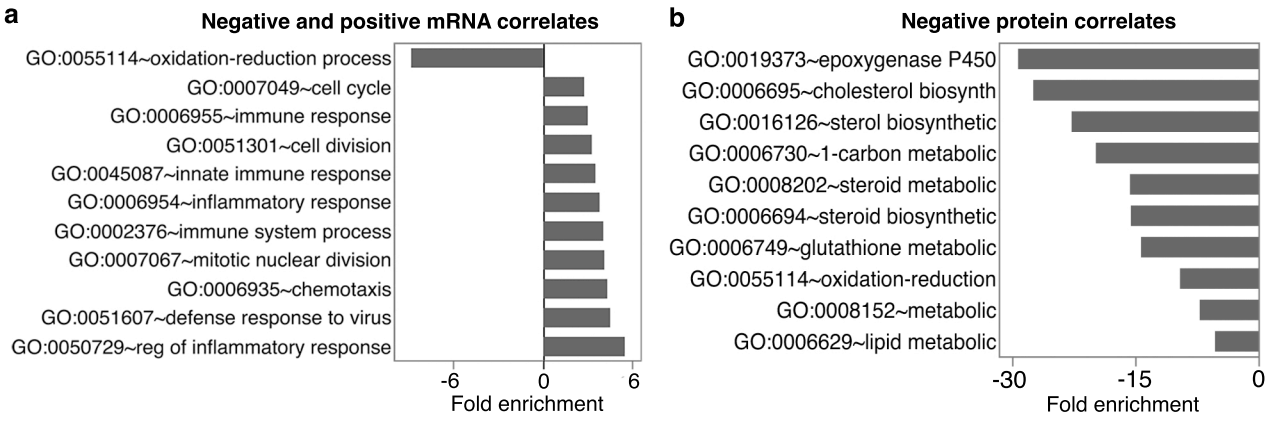

a
CD: r=0.93; HFD: r=0.94
univ.DNAmAge (years)
Age (years)
CD+HFD
r=0.91
d
CD: r=0.19, p=0.005
HFD: r=0.34, p<.0001
univ.EAA
DNAm entropy
CD+HFD
r=0.29, p<.0001
b
EAA.pan
EAA.liver
dev.EAA.pan
dev.EAA.liver
int.EAA.pan
int.EAA.liver
Uni.EAA
Epigenetic age acceleration
Linear fit
Both
CD
HFD
Diet
CD
HFD
c
CD: r=0.46, p<.0001
HFD: r=0.68, p<.0001
DNAm entropy
Age (years)
CD+HFD
r=0.52
e
CD: r=–0.16, p=0.02
HFD: r=–0.07, ns
Predicted-maxLS (years)
DNAm entropy
CD+HFD
r=–0.13, p=0.02


a
r=0.60; p<.0001
univ.EAA
BW0 (g)
Days on HFD
1
3
b
CD: r=0.15, p=0.04
HFD: r=0.35, p<.0001
int.EAA, liver
deltaBW (g)
Diet
CD
HFD
CD+HFD
r=0.39, p<.0001
c
int.EAA, pan
CD
HFD
Diet
d
Predicted-maxLS (years)
CD
HFD
Diet
e
DNAm entropy
CD
HFD
Diet
Females
Males
f
Predicted-maxLS (residuals)
Total cholesterol (residuals)
g
DNAm entropy (residuals)
Total cholesterol (residuals)
h
univ.EAA (residuals)
Glucose (residuals)
i
DNAm entropy (residuals)
Glucose (residuals)


a
CD: r=−0.02, ns
HFD: r=−0.16, p=0.09
Strain LS (25Q, years)
CD+HFD
r=−0.21
p=0.0003
univ.EAA
b
CD: r=−0.15, p=0.04
HFD: r=−0.23, p= 0.01
Strain LS (75Q, years)
CD+HFD
r=−0.31
p<.0001
univ.EAA
c
CD: r=−0.06, ns
HFD: r=−0.06, ns
Strain LS (Median, years)
CD+HFD
r=−0.19
p=0.001
int.EAA, liver
d
CD: r=0.23, p=0.002
HFD: r= 0.1, ns
Strain LS (max, years)
CD+HFD
r=0.17
p=0.003
Predicted-maxLS (years)
Linear fit
Both
CD
HFD
Diet
CD
HFD


a
int.EAA, pan
Genotypes @ Chr11 92.75 Mb
BB
DD
int.EAA, pan
0.1
0.0
-0.1
CD
HFD
Stxbp4, Erbb2,
Grb7, Ikzf3
-log10(p)
b
int.EAA, liver
Genotypes @ Chr19 38.65 Mb
BB
DD
int.EAA, liver
0.2
0.0
-0.2
CD
HFD
Cyp2c29, Cyp2c37,
Kif11, Cep55,
Nkx2-3, Cutc, Chuk
-log10(p)
c
univ.EAA
-log10(p)
d
Predicted-maxLS
Genotypes @ Chr19 47.51 Mb
BB
DD
Predicted-maxLS
4.5
4
3.5
CD
HFD
Pcgf6, Elovl3
-log10(p)
1
2
3
4
5
6
7
8
9
10
11
13
15
17
19
Chromosome


a
Leaaq11
HFD
CD
Weight (g)
48
36
24
4
6
12
18
24
Age (months)
b
Leaaq19
HFD
CD
Weight (g)
48
36
24
4
6
12
18
24
Age (months)
Genotype
BB
DD


a
Negative and positive mRNA correlates
GO:0055114~oxidation-reduction process
GO:0007049~cell cycle
GO:0006955~immune response
GO:0051301~cell division
GO:0045087~innate immune response
GO:0006954~inflammatory response
GO:0002376~immune system process
GO:0007067~mitotic nuclear division
GO:0006935~chemotaxis
GO:0051607~defense response to virus
GO:0050729~reg of inflammatory response
-6
0
6
Fold enrichment
b
Negative protein correlates
GO:0019373~epoxygenase P450
GO:0006695~cholesterol biosynth
GO:0016126~sterol biosynthetic
GO:0006730~1-carbon metabolic
GO:0008202~steroid metabolic
GO:0006694~steroid biosynthetic
GO:0006749~glutathione metabolic
GO:0055114~oxidation-reduction
GO:0008152~metabolic
GO:0006629~lipid metabolic
-30
-15
0
Fold enrichment