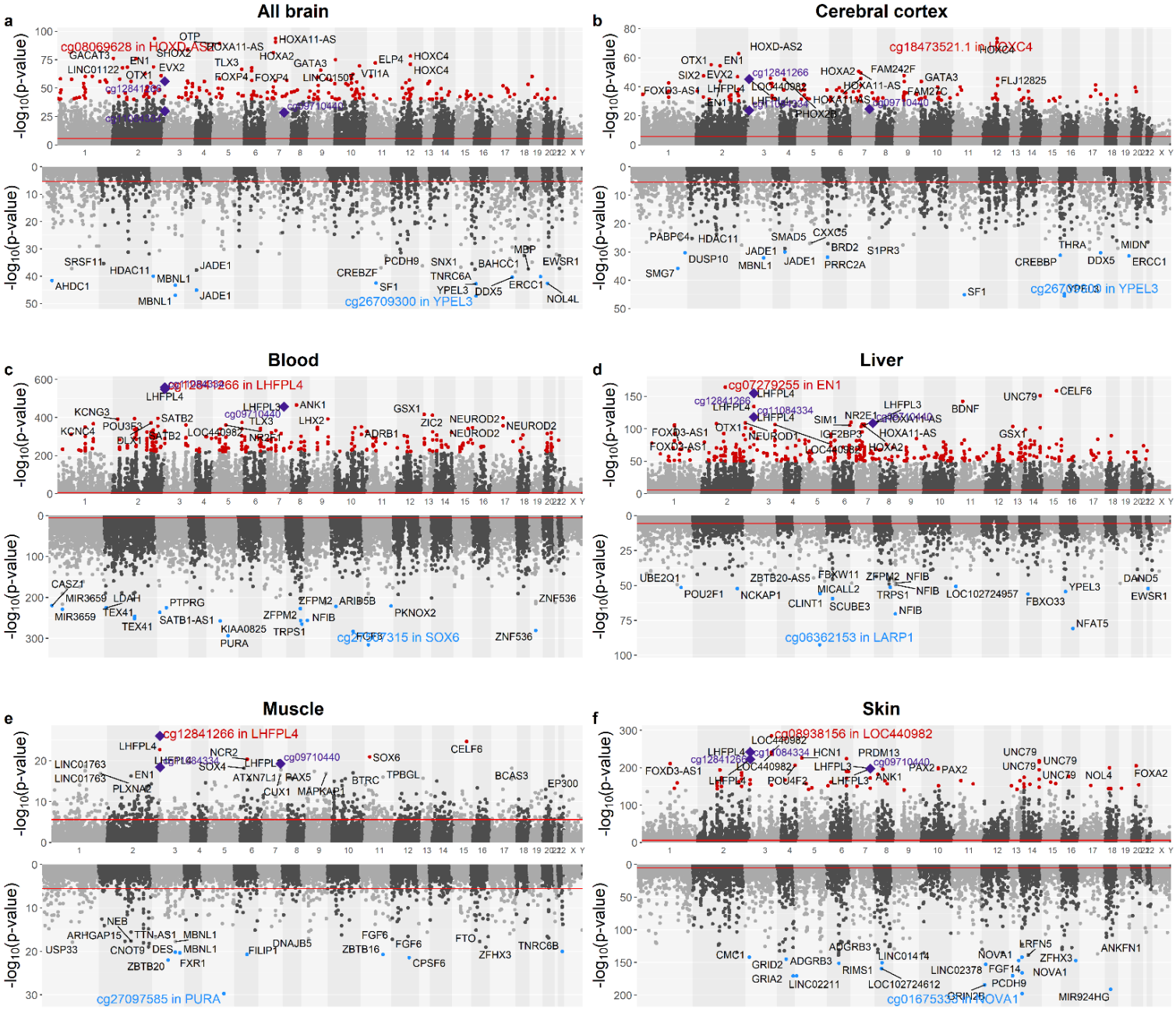
Extended Data Figures and Tables

for the article entitled “Universal DNA methylation age across mammalian tissues"

AT Lu *et al.*

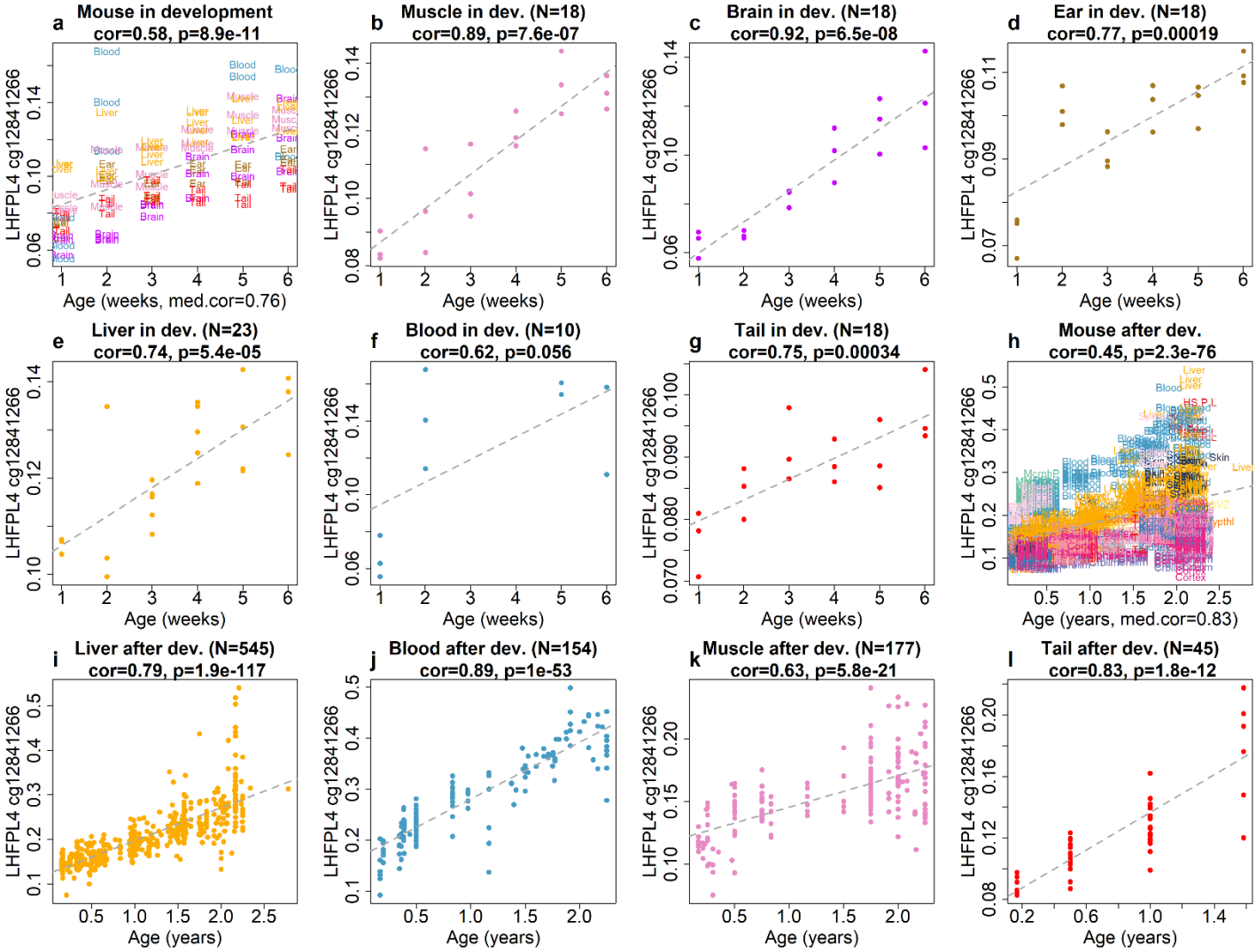
Extended Data Figure 1. Meta analysis of chronological age in mammalian samples across specific tissue types.

Meta-analysis p-value (-log base 10 transformed) versus chromosomal location (x-axis) according to human genome assembly 38 (Hg38) in a, brain tissues (across multiple regions), b, cerebral cortex in bottom, c, blood, d, liver, e, muscle and f, skin tissues. The upper and lower panels of the Manhattan plot depict CpGs that gain/lose methylation with age. In panel a, P values were calculated via two-stage meta-analysis that combined EWAS results across strata formed by species/brain-tissue (with n>= 15 samples). CpGs are colored in red and blue if they exhibit highly significant positive and negative age correlations according to P $<1.0x10^{-40}$, $1.0x10^{-30}$, $1.0x10^{-220}$, $1.0x10^{-50}$, $1.0x10^{-20}$ and $1.0x10^{-140}$ for a -f, respectively. Red lines denote Bonferroni correction. Gene names are annotated for the top 20 CpGs with positive and negative associations, respectively. CpG-gene names are annotated for the best CpG with positive/negative association; additionally, cg12841266 and cg11084334 in LHFPL4 and cg09710440 in LHFPL3 are marked in purple and diamond shape.



Extended Data Figure 2. Association between methylation levels in LHFPL4 cg12841266 and chronological age in and after development stage across mouse tissues.

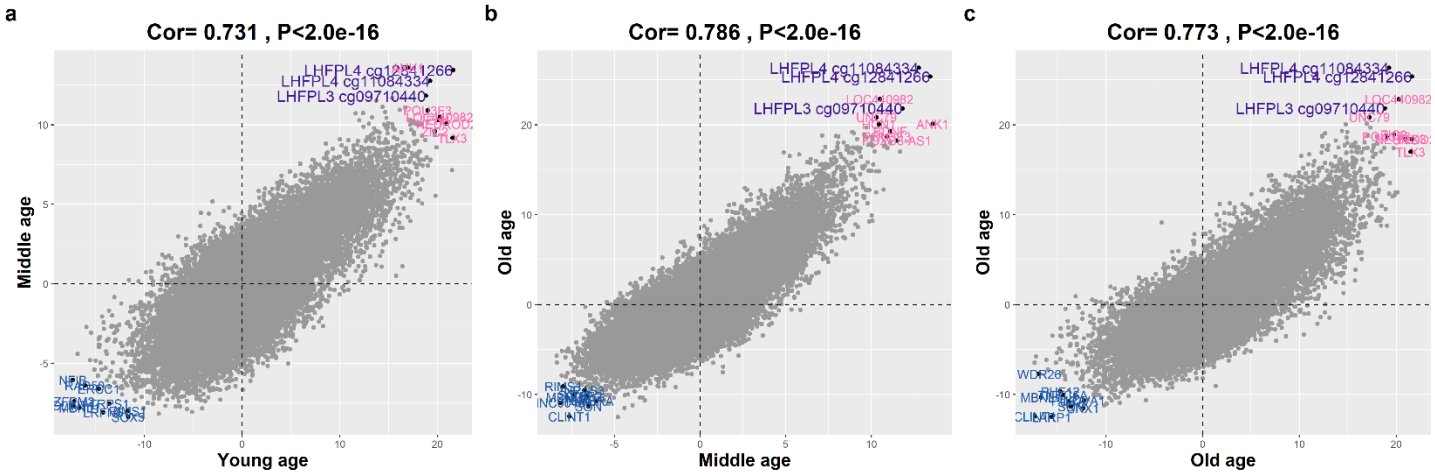
Pearson correlation coefficients and P values are listed in titles. Mouse tissues are grouped into development stage, before sexual maturity, **a-g** (n=105) and post-development stage, after sexual maturity, **h-m** (n=1514). The mean±SD [range] of age is 3.5±1.7 [1.0, 6.0] weeks in the development group and 1.10±0.72 [0.15, 2.78] years in the post-development group. **a,h,** correlation between age (in units of year) and methylation levels of LHFPL4 cg12841266, in development and post-development stage, respectively. Each sample is marked by its tissue name and tissue-specific color. **b-g** display age-related methylation of cg12841266 in different tissues in the development (dev.) group. **i-l** display age-related methylation of cg12841266 indifferent tissues in the post-development group. Table m lists additional tissue types in the post-development group. P values are reported for sample size greater than 2.



m

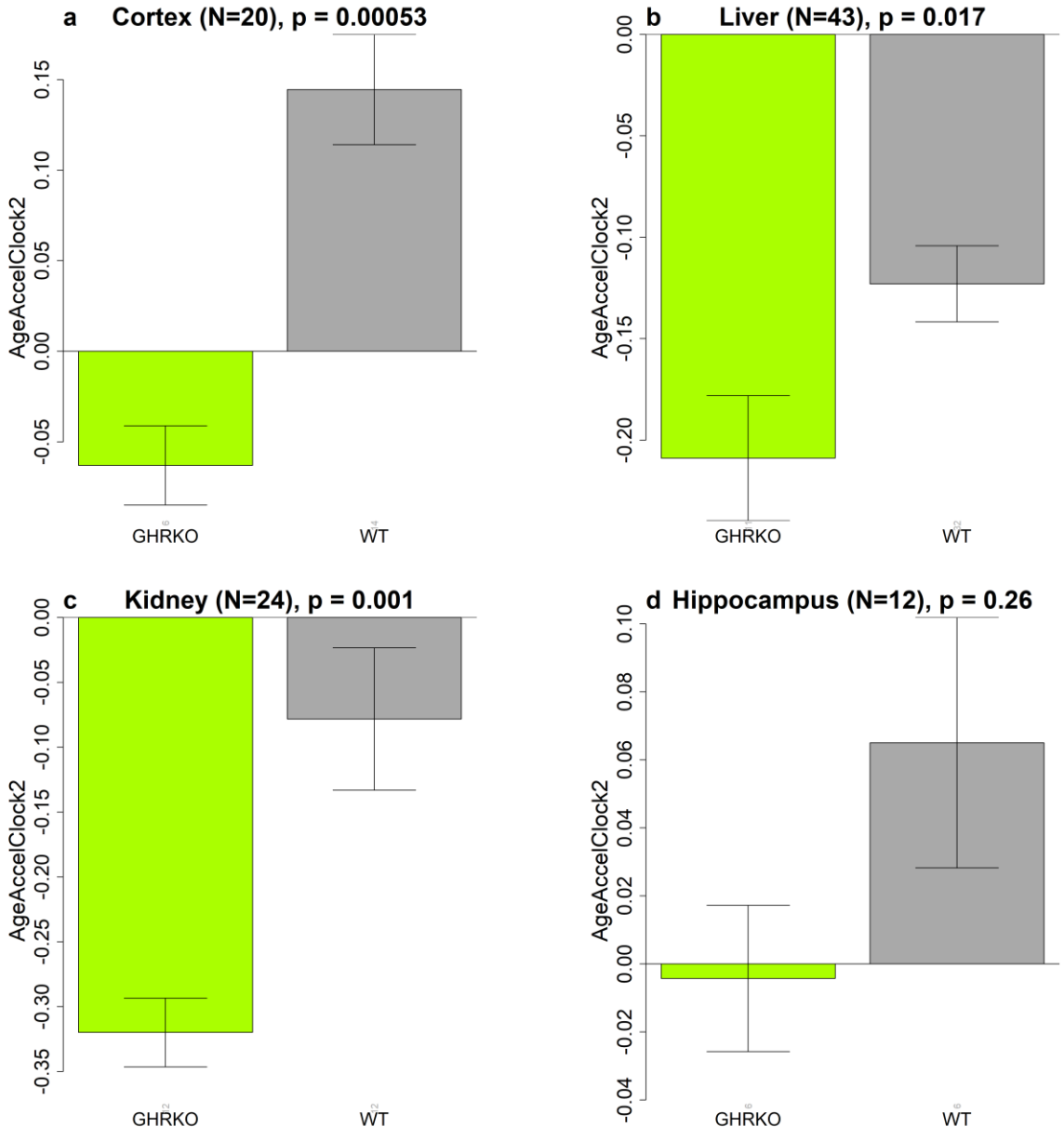
Tissue	N	Cor	P
Hypothalamus	2	1	--
Hemat.Prog.LSK	12	0.95	1.50E-06
Spleen	31	0.94	4.70E-15
Skin	34	0.89	1.40E-12
Kidney	45	0.88	1.20E-15
SVZ	8	0.88	4.10E-03
Heart	89	0.85	1.30E-25
Lung	15	0.83	1.10E-04
Cortex	151	0.52	5.30E-12
Striatum	50	0.44	1.20E-03
MacrophagePeritoneal	42	0.39	1.00E-02
Brain	19	0.38	1.10E 01
Cerebellum	83	0.36	7.00E-04
Fibroblast	12	0.21	5.20E-01

Extended Data Figure 3. EWAS meta analysis of chronological age in mammalian samples stratified by age group.
Each axis reports a Z scores from two-stage EWAS meta-analysis of age across all species and tissues: young age (age < 1.5* age at sexual maturity [ASM]), middle age (age between 1.5 and 3.5 ASM), or old age (age >= 3.5 ASM). Each dot corresponds to a CpG. Given by the product of Z scores in x and y-axis, the top 10 hypermethylated/hypomethylated CpGs are marked in pink/blue or especially marked in purple if they are located in LHFPL4/LHFPL3 genes. The Pearson correlation coefficient and corresponding nominal (unadjusted) two-sided correlation test p-value can be found in the title.



Extended Data Figure 4. Universal clocks applied to GHRKO mouse experiment.

We applied clock 2 to mouse dataset (N=99 tissues) comprising two genotypes: growth hormone receptor/binding protein knockout (GHRKO,n=35) and wild type (WT, n=64). The mean±SD [range] of age is 0.66±0.06 [0.58, 0.79] years in the GHRKO group and is 0.59±0.09 [0.50, 0.79] years in the WT group. To measure epigenetic age acceleration we fit a regression model of DNAm age on chronological age and defined age acceleration (AgeAccel) as the resulting raw residual. Each panel depicts epigenetic age acceleration (y-axis) versus mouse genotypes. Each panel report test tissue type, sample size and the p-value of a non-parametric group comparison test (Kruskal-Wallis). The y-axis of the bar plots depicts the mean and one standard error. The numbers of individuals per genotype are reported as grey numbers in each bar. To combine the results across tissues, we applied Stouffer's method: meta p-value =2.82x10^{-7} for the age acceleration from clock 2.



Extended Data Table 1. The overlap of age-associated genes across different EWAS results. For EWAS of age in brain, blood, muscle and skin, we identified a total of 51 overlapping CpGs across the top 1000 CpGs with positive age association. Analogously, we identified 3 overlapping CpGs with negative age association. We identified 33 unique genes proximal to the 51 hypermethylation CpGs and 3 genes proximal to the three hypomethylation CpGs. For genes with multiple overlapping CpGs, we report the CpG with the best P value across all tissues and ranked the genes accordingly. The table lists the columns: rank of the gene base on the EWAS meta P value from all tissues, HGNC gene symbol, the best CpG, protein domain, activity and function of gene expression.

Rank	Gene	CpG	Protein domain	Activity	Function
Hypermethylation					
1	LHFPL4	cg12841266	Tetraspan transmembrane	Clustering of GABA receptors	Dev. of inhibitory synapse
2	LHFPL3	cg09710440	Tetraspan transmembrane	Clustering of GABA receptors	Dev. of inhibitory synapse
3	TLX3	cg26844246	Homeobox	TF	Dev.
4	EVX2	cg09227056	Homeobox	TF	Limb Dev.
5	NEUROD2	cg03679521	bHLH	TF	Neural Dev.
6	ZIC2	cg15682828	Zinc-finger	TF	Early Dev.
7	CELF6	cg23087015	RNA recognition motif (RRM)	RNA-binding	RNA processing
8	PAX2	cg01486146	Paired box	TF	Early Dev.
9	BDNF	cg27201382	neurotrophin family of growth factors		Brain Dev.
10	PAX5	cg20766695	Paired box	TF	Early Dev.
11	NRN1	cg16356803	neuritin family		Nervous System Dev.
12	PRDM13	cg05797975	Zinc finger		transcriptional regulation
13	OTP	cg24352905	Homoebox	TF	Dev./Cell fate specification
14	NEUROD1	cg09942248	bHLH	TF	Beural Dev.
15	SALL1	cg13909487	Zinc finger	TF	Stem cell regulation and Dev.
16	NR2E1	cg27496468		Nuclear receptor	Retinal Dev.
17	TBX18	cg09817427	T-Box	TF	Stem Cell Differentiation
18	TWIST1	cg20477718	bHLH	TF	Embryonic Dev.
19	SIX2	cg03557129	Homoebox	TF	Limb & Eye Dev.
20	TFAP2D	cg26834803		TF	Embryonic Dev.
21	OBI1-AS1	cg23985931		Anti-sense RNA	
22	NKX2-2	cg16524928	Homoebox	TF	Spinal cord Dev.
23	PHOX2B	cg13278722	Homeobox	TF	Neural Dev.
24	DBX1	cg21452781	Homeobox	TF	Internertonal Differentiation
25	POU3F1	cg13193455	Homeobox	TF	Neuronal Dev.
26	OTX1	cg03957108	Homeobox	TF	Brain and sensory Dev.
27	ZIC5	cg24255409	Zinc finger	TF	Early Dev.
28	LOC642366	cg16744551		Anti-sense RNA	
29	VSX2	cg10707128	Homeobox	TF	Specification of sensory retina
30	TBR1	cg04879029	T-Box	TF	Numerous Dev. processes
31	EGR3	cg02040024	Zinc finger	Transcription Regulator	Numerous Dev. processes
32	POU3F2	cg08921975	Homeobox	TF	Neuron Dev.
33	IRX1	cg02926165	Homeobox	TF	Pattern formation
Hypomethylation					
1	LARP1	cg12880090	mTORC1 downstream targets	RNA-binding	Regulation of translation
2	SON	cg24466972		RNA-binding	RNA splicing
3	SNX1	cg27626343	Phox (PX)	Endosomal protein	Regulates expression of EGF receptor

Dev=development; TF=transcriptional factor.

Extended Data Table 2. Top results of GREAT enrichment analysis. The table presents the top results of GREAT analysis of genes proximal to positive/negative age-related CpGs from EWAS across all tissues (All), brain, cortex, blood, liver, muscle and skin, respectively. Column 1 lists tissue type in EWAS. Columns 2-5 lists the name of top gene set and unadjusted hypergeometric P values. For MSigDB Perturbation based on the positive age-related genes, we report the results from BENPORATH_ES_WITH_H3K27ME3, BENPORATH_EED_TARGETS and BENPORATH_SUZ12_TARGETS, respectively. For the other ontogenies, we report the most significant pathway/gene set, provided hypergeometric $P < 0.001$.

Tissue	GO Biological Process	MSigDB Pathway	MSigDB Perturbation	Mouse Phenotype
Positive age-related CpGs				
All	nervous system development(P=3.7e-207)	Reg. of gene exp. of beta cells (P= 4e-38)	H3K27, Eed, Suz12 PRC2 targets.(P values=2.8e-266, 1.7e-262, 7.1e-225)	abnormal nervous system morphology(P=8e-199)
Brain	pattern specification process(P=3e-64)	--	H3K27, Eed, Suz12 PRC2 targets(P values=3.3e-68, 5.5e-64,8.7e-54)	lethality during fetal growth through weaning(P=2e-45)
Cortex	pattern specification process(P=3e-115)	--	H3K27, Eed, Suz12 PRC2 targets(P values=5.1e-116,3.5e-111,6.1e-87)	abnormal skeleton morphology(P= 2e-75)
Blood	anatomical structure development(P=2e-249)	Reg. of gene exp. of beta cells (P= 5e-30)	H3K27, Eed, Suz12 PRC2 targets(P values=3.9e-283, 1.9e-294, 3.9 e-259)	abnormal nervous system morphology(P=2e-231)
Liver	anatomical structure development(P=1e-157)	--	H3K27, Eed, Suz12 PRC2 targets(P values =3.3e-189, 4.9e-190, 1.7e-149)	nervous system phenotype(P=1e-139)
Muscle	central nervous system development(P=7e-17)	--	H3K27, Eed, Suz12 PRC2 targets(P values =8.7e-18, 1.4e-17, 8.2 e-16)	abnormal midbrain morphology(P=6e-13)
Skin	nervous system development(P=5.4e-141)	Maturity onset diabetes of the young(P= 3e-38)	H3K27, Eed, Suz12 PRC2 targets(P values =3.3e-189, 6.2e-191, 2.6e-150)	nervous system phenotype(P=1e-144)
Negative age-related CpGs				
All	--	--	--	--
Brain	RNA splicing (P=6e-42)	Circadian rhythm – mammal (P=9e-18)	Genes up-regulated in Alzheimer's disease. (P=2.1e-30)*	increased circulating glucagon level (P=3e-18)
Cortex	RNA processing (P= 5e-34)	Circadian rhythm – mammal (P= 4e-19)	Genes related to number of oligodendrocytes (P= 5.9e-22)	increased circulating glucagon level (P= 3e-17)
Blood	anterior compartment pattern formation (P= 3e-08)	--	Gene set associated with MLL fusions (P= 1e-09)	improved righting response (P= 2e-06)
Liver	lung ciliated cell differentiation (P= 4e-10)	--	Genes up-regulated under low serum(P= 2e-08)	absent lung saccules (P= 4e-10)
Muscle	cellular component assembly (P=3e-15)	--	Genes up-regulated in TLX KO neural stem cells(P=1e-28)	increased variability of skeletal muscle fiber size (P=6e-22)
Skin	--	--	--	abnormal circulating alanine transaminase level (P= 2e-06)

-- denotes not available.

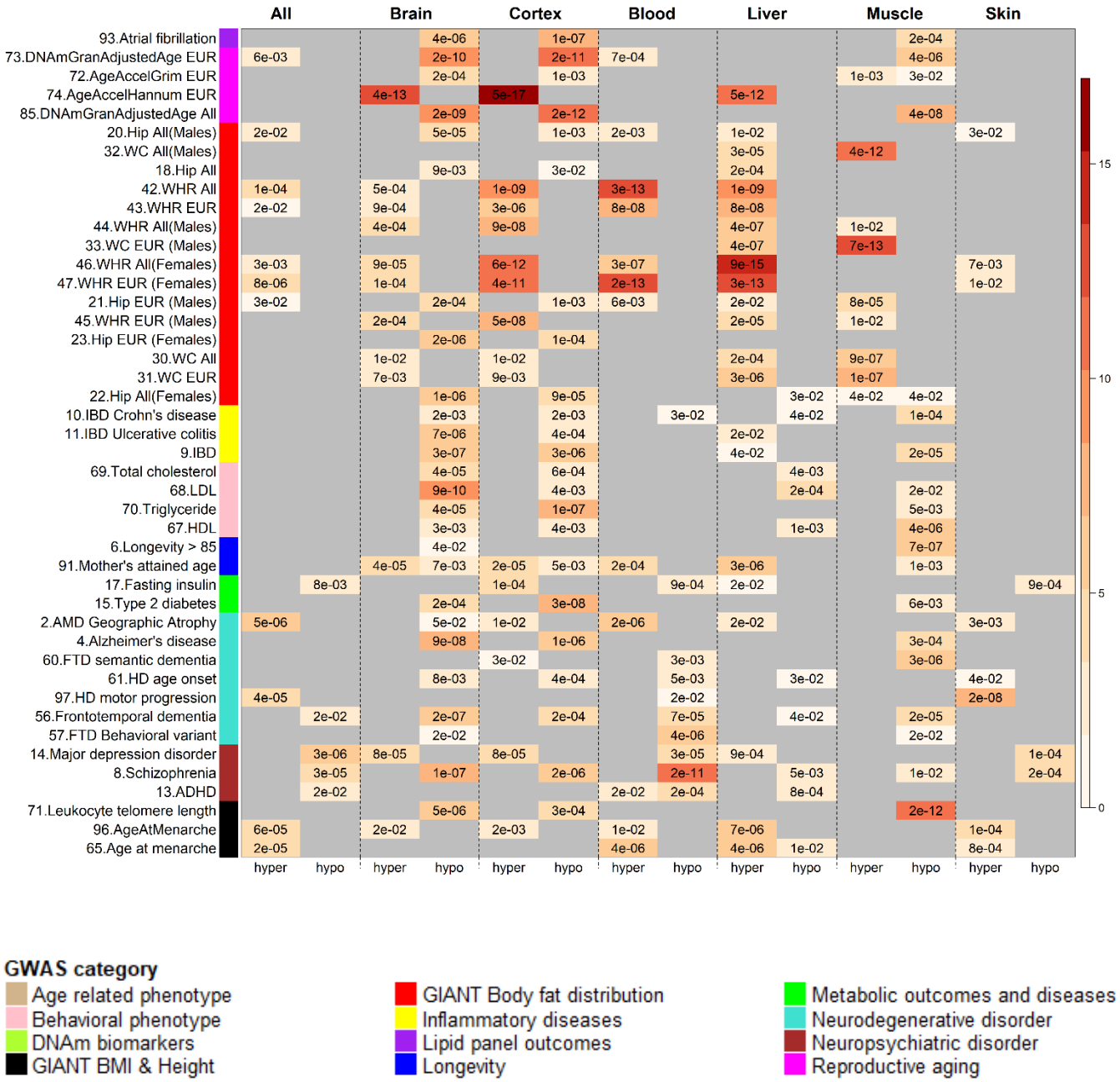
* ID of the gene set is BLALOCK_ALZHEIMERS_DISEASE_UP. The results of BLALOCK_ALZHEIMERS_DISEASE_UP for other tissues are listed in Supplementary Table 5.2 (all tissues), 5.7 (cortex), 5.8 (blood) and 5.13 (muscle).

Extended Data Table 3 Select gene sets from GREAT enrichment analysis. Results from the GREAT enrichment analysis of top (+/-) 1,000 age-related CpGs in all tissues (All), brain, blood, cortex, liver, muscle and skin EWAS. The top panel highlights the results related to mitochondrial gene sets from GO Cellular component ontology. The lower panel highlights the gene sets from HGNC Gene Families ontology – gene sets based on sequence similarity. Columns 1 to 4 lists EWAS tissue type, sign of age-related CpGs, gene set ID, and hypergeometric P value. Column 5 lists the overlapping genes between the genes proximal to the 1,000 age-related CpGs and the genes in the corresponding test gene set. Abbreviations: mit.ATP denotes mitochondrial proton-transporting ATP synthase complex (mit.ATP) and mit.part denotes mitochondrial part.

Mitochondrial				
Tissue	**Sign**	**ID**	**P**	**Overlap gene**
All	+	mit.ATP	4.56E-10	ATP5EP2,ATP5G2
Blood	+	mit.ATP	3.24E-11	ATP5EP2,ATP5G2
Brain	-	mit.part	2.20E-08	ATP5G2,ATP5L,BAD,BCL2L1,BMF,CHPF,CNP,COQ3,COX20,CSDE1,DDX3X,ERAL1,GK2,IDH3A,KIAA1967,MALSU1,MAPK8IP1,MFN2,NDUFB11,NFS1,PI4KB,PINK1,PLSCR3,POLG2,PPOX,RAB11FIP5,SFXN5,SLC25A36,TIMM10B,TMEM126A,TUFM,USMG5
Cortex	-	mit.part	3.40E-06	ATP5A1,BCL2L1,CHPF,COQ3,ERAL1,GRSF1,IDH3A,KIAA1967,MALSU1,NFS1,PCCB,POLG2,TUFM
HGNC Gene Families				
All	+	NKL	2.4E-49	BARHL1,BARHL2,DBX1,DLX1,DLX2,DLX6,EN1,EN2,HMX2,LBX1,MSX1,NKX2-1,NKX2-2,NKX2-5,NKX2-8,NKX3-2,TLX1,TLX3,VAX1
	+	PRD	1.1E-34	DMBX1,DRGX,OTP,OTX1,OTX2,PHOX2B,PITX2,PRRX1,SHOX2,UNCX,VSX2
	+	FOX	3.50E-18	FOXA1,FOXA2,FOXB1,FOXD2,FOXD3,FOXG1,FOXQ1
Brain	+	PRD	2.00E-18	DMBX1,DRGX,OTP,OTX1,PHOX2B,PITX2,SHOX2,UNCX,VSX2
	+	NKL	2.80E-14	BARHL1,DBX1,DLX1,DLX2,DLX5,DLX6,EMX1,EMX2,EN1,HHEX,HMX3,LBX1,MSX1,NKX2-1,NKX2-2,NKX3-2,NKX6-3,TLX3
	+	FOX	1.50E-13	FOXA1,FOXA2,FOXD3,FOXG1,FOXL1,FOXN1,FOXO6,FOXP2,FOXP4
Cortex	+	PRD	1.70E-36	DMBX1,DRGX,OTP,OTX1,PHOX2B,PITX2,SHOX2,UNCX,VSX2
	+	NKL	7.00E-21	BARHL2,DBX1,DLX1,DLX2,DLX5,DLX6,EN1,HMX3,LBX1,MSX1,NKX2-1,NKX2-2,NKX3-2,TLX3
	+	FOX	1.40E-16	FOXA1,FOXA2,FOXB1,FOXC1,FOXD2,FOXD3,FOXE3,FOXG1,FOXL1,FOXO6,FOXP2,FOXP4,FOXQ1
Blood	+	NKL	5.70E-74	BARHL1,BARHL2,DBX1,DLX1,DLX2,DLX5,DLX6,EMX2,EN1,EN2,HMX2,LBX1,MSX1,NKX2-1,NKX2-2,NKX2-5,NKX2-8,NKX3-2,TLX1,TLX3,VAX1
	+	PRD	4.70E-45	DMBX1,DRGX,OTP,OTX1,OTX2,PHOX2B,PITX1,PITX2,PRRX1,SHOX2,UNCX,VSX2
	+	FOX	3.50E-18	FOXA1,FOXA2,FOXB1,FOXD2,FOXD3,FOXF2,FOXG1,FOXP4,FOXQ1
Liver	+	NKL	1.40E-45	BARHL1,BARHL2,DBX1,DLX1,DLX2,DLX5,DLX6,EMX2,EN1,EN2,HMX3,LBX1,MSX1,NKX2-1,NKX2-2,NKX2-3,TLX1,TLX3,VAX1
	+	PRD	1.40E-10	DMBX1,OTP,OTX1,PHOX2B,PITX2,PRRX1,SHOX2,UNCX,VSX2
Muscle	+	PRD	1.20E-04	DMBX1,OTP,OTX1,PHOX2B,PITX2,UNCX,VSX2
	+	NKL	2.10E-04	BARHL2,DBX1,DLX1,DLX2,DLX6,EMX1,EMX2,EN1,EN2,HMX3,LBX1,NKX2-1,NKX2-2,TLX3
Skin	+	NKL	5.00E-26	BARHL1,BARHL2,DBX1,DLX6,EN1,EN2,HHEX,HMX2,LBX1,MSX1,NKX2-1,NKX2-2,NKX2-8,NKX6-1,TLX1,TLX3,VAX1
	+	PRD	1.70E-24	DMBX1,DRGX,OTP,OTX1,OTX2,PHOX2B,PITX2,PRRX1,UNCX,VSX2
	+	FOX	5.30E-19	FOXA1,FOXA2,FOXB1,FOXD3,FOXG1
Brain	-	RBM	6.50E-10	CIRBP,CPEB2,CPEB4,CPSF6,DAZAP1,EWSR1,HNRNPA2B1,HNRNPC,HNRNPDL,HNRNPLL,PABPC4,PABPN1,PUF60,RBM39,SFPQ,SNRNP70,TRA2A,TRA2B,U2AF2
	-	ZC3H	8.00E-07	MBNL1,ZC3H14
Cortex	-	RBM	6.50E-10	CIRBP,DAZAP1,EWSR1,GRSF1,HNRNPA2B1,HNRNPC,HNRNPLL,PABPC4,PUF60,RBM39,SFPQ,SNRNP70,TRA2A,TRA2B,U2AF2
muscle	-	RBM	6.30E-06	CELF1,CIRBP,CPEB2,CPEB4,CPSF6,DAZAP1,HNRNPC,HNRNPD,PABPC4,RBM39,SPEN,TRA2B
	-	ZC3H	1.00E-04	MBNL1,ZC3H11A

Extended Data Figure 5 Enrichment analysis EWAS-GWAS associated genes.

The plot highlights the significant results from the genomic-region based enrichment analysis between (1) the top 5% genomic regions involve in GWAS of complex traits-associated genes and (2) the top 1000 hypermethylated/hypomethylated CpGs from EWAS of age-associated ages across all, blood, brain, cerebral cortex, liver, muscle, and skin tissues, respectively. We list the GWAS study (x-axis) if that at least one enrichment P value (columns) is significant at a nominal significance level of $5.0x10^{-4}$. Cells are colored in grey if nominal P>0.05. The heatmap color gradient is based on -log10 (hypergeometric P value). The x-axis indicates whether the EWAS list for age is based on hypermethylation (hyper, positive correlation with age) or hypomethylation (hypo, negative correlation with age). The y-axis lists GWAS index number (listed in **Supplementary Info**), trait name. The color band next to the trait encodes the GWAS category. Abbreviations: All=All ancestries, HD=Huntington's disease, WC=waist circumference, WHR= waist to hip ratio, IBD=inflammatory bowl disease, EUR=European ancestry. FTD=frontal temporal dementia.



GWAS category

- Age related phenotype
- Behavioral phenotype
- DNAm biomarkers
- GIANT BMI & Height
- GIANT Body fat distribution
- Inflammatory diseases
- Lipid panel outcomes
- Longevity
- Metabolic outcomes and diseases
- Neurodegenerative disorder
- Neuropsychiatric disorder
- Reproductive aging

Extended Data Table 4 Predictive accuracy of three universal mammalian clocks.
The rows correspond to 3 universal clocks based on penalized regression models for I) log-transformed age, log(Age+2), II) relative age, and III) log-linear age. The inverse transformations were used to arrive at age estimates, DNA methylation age, in units of years. Each clock was evaluated two validation schemes: leave-one-fraction-out cross validation (LOFO) and leave-one-species-out cross validation (LOSO). Columns 3 and 4 report the Pearson correlation coefficient (Cor) and median absolute error (MAE, in units of years) across all samples, i.e. species and tissue type are being ignored. Columns 5 and 6 report the median value across species, i.e. median Cor (med.Cor) and median MAE (med.MAE), for species with at least 15 observations. Columns 7-9 repeated the analysis based on relative age scales except med.Cor, which is not changed by using relative age scale. In Clocks 1&3, DNAm relative age was computed based on formula (1) with the variable Age replaced by DNAm Age.

		DNAm Age vs age				DNAm relative age vs relative age		
Universal Clock	**Validation scheme**	**Cor**	**MAE**	**med. Cor**	**med. MAE**	**Cor**	**MAE**	**med. MAE**
Clock 1: Naïve, log(Age+2)	LOFO	0.964	0.768	0.909	0.919	0.899	0.032	0.032
Clock 2: RelativeAge	LOFO	0.986	0.521	0.940	0.675	0.960	0.020	0.022
Clock 3: log-linear age	LOFO	0.982	0.740	0.914	1.023	0.925	0.03	0.033
Clock 1: Naïve, log(Age+2)	LOSO	0.729	1.729	0.830	1.837	0.493	0.084	0.058
Clock 2: RelativeAge	LOSO	0.939	1.416	0.893	1.217	0.757	0.061	0.044
Clock 3: log-linear age	LOSO	0.920	1.822	0.883	1.912	0.682	0.082	0.053

Extended Data Figure 6 Universal clocks applied to species with fewer than 15 methylation samples.

The title of each panel lists the type of universal clock: a, Clock 1=naïve universal clock based on log(Age+2), b,d, Clock 2=universal clock for relative age, c, Clock3 =universal clock for log-linear age. Leave one folder out (LOFO) methylation estimates versus chronological age (a-c) or relative age (d). The respective inverse transformations were applied to arrive at DNA methylation-based estimates of chronological age in years or relative age (y-axis).