Supplementary Material

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

AT Lu *et al.*

Supplementary Note

Supplementary Note 1: Data description.

The combined data was generated by the Mammalian Methylation Consortium. The dataset is composed of individual data sets for different species that are described in separate articles [1-15].

Supplementary Note 2: Sensitivity analysis of enrichment results

It critical to use a suitable background when it comes to any gene/pathway enrichment study. The wrong choice of background could easily lead to erroneous but highly significant associations due to hidden biases. When it comes to the mammalian array the choice of the proper background must reflect the following sources of bias. First, limited genome coverage provided by the 37k CpGs on the array. For example, the CpGs on the mammalian array cover 6871 human and 5659 mouse genes when each CpGs is assigned uniquely to its closest gene neighbor. Second, by design, the mammalian array is biased toward highly conserved genomic regions. To address these biases, we evaluated the GREAT analysis software tool. As illustrated below, we find that GREAT analysis effectively deals with these biases and leads to biologically meaningful insights. In the following, we will report results from two different sensitivity analyses that were inspired by our GREAT enrichment analysis of the top 1 thousand age related CpGs (EWAS of age). Our first sensitivity analysis involved a random set of 1000 CpG mammalian CpGs. In essence, this evaluates the null hypothesis of no relationship between chronological age and methylation. The most significant (nominal) enrichment p value was $p=3.9x10^{-4}$. Note that this p-value is far less significant than the enrichment p values for age-related CpGs in our article: top 1k negative CpGs lead to $p=2.7x10^{-8}$; top 1k positive age-related CpGs lead to $p=2.7x10^{-266}$. We repeated this analysis with several sets of random 1k CpGs and obtained similar results.

Second, we also evaluated the enrichment of the top 1087 most highly conserved CpGs across 158 mammalian genomes. This sensitivity analysis addresses the concern that highly conserved CpGs could have an increased chance of correlating strongly with chronological age or, conversely, non-conserved (noise) CpGs are expected to have no signal for age and will therefore not be selected in an EWAS of age. This hidden bias would manifest itself as follows: the enrichment analysis of our meta analysis EWAS for age would be equivalent to the EWAS

of highly conserved CpGs. In the following, we provide details that demonstrate that this is not the case. This biologically meaningful set of 1087 highly conserved CpGs led to highly significant enrichment p-values for gene sets involved in RNA processing, and RNA splicing, and lipoprotein particle biosynthesis. Some of the top gene families of these conserved probes include RBM and LDLR. For example, for ontology class "MSigDB Cancer Neighborhood" we find $p=5.2x10^{-19}$ for "Neighborhood of SMC1L1", $p=2.67\mathbf{x10^{-18}}$ for "Neighborhood of TDG", $p=1.57x10^{-16}$ for "Neighborhood of XRCC5". Highly significant GO Biological Processes include RNA processing ($p=1.56x10^{-17}$), RNA binding ($p=5.90x10^{-16}$), mRNA processing ($p=1.15x10^{-14}$), and RNA splicing ($p=3.9x10^{-11}$). However, these enrichments are quite distinct from those observed for the EWAS of age. RNA splicing and processing only showed a weak significance ($p = 0.05$ to $1.4x10^{-3}$) in hypomethylated age-related CpGs. In summary, we did not observe any overlap between the top enrichment terms for the age-related CpGs with those from highly conserved regions (or those from a random set of CpGs). A detailed enrichment analysis of all the CpGs on the mammalian array can be found in [1].

GREAT was not explicitly designed to adjust for the issue of certain CpG's having more power to detect association based on working in more species, but it appears not to be driving categories of enrichment for age. Overall, our sensitivity analysis of the enrichment study demonstrates that GREAT analysis adjusted for potential biases arising from the design of the mammalian array and protected us against spurious associations.

Supplementary Note 3: Enrichment analysis for overlap between EWAS of mammalian age associated genes and large-scale GWAS associated ages of complex traits

We investigate the overlap genomic regions using our EWAS results for age (all tissues and stratified by tissue types) and a total of 69 large-scale GWAS studies for anthropometric traits, behavioral phenotypes, cognitive related traits, inflammatory diseases, lipid panel outcomes, metabolic outcomes and diseases, neurodegenerative and neuropsychiatric disorders, longevity, reproductive aging and other age related phenotypes including DNA methylation based biomarkers. The GWAS results are corresponding to previously published large-scale studies. For instance, GWAS of anthropometric traits are based on the studies conducted by the GIANT consortium,

https://portals.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium. The GWAS articles are summarized below.

Index	Hg	Category	Trait	Ethnicity	Sex	PMID
1	hg19	Neurodegenerative disorder	Age-related Macular degeneration (AMD)	EUR+ASN	All	23455636
2	hg19	Neurodegenerative disorder	AMD Geographic Atrophy	EUR+ASN	All	23455636
3	hg19	Neurodegenerative disorder	AMD Neovascular	EUR+ASN	All	23455636
4	hg19	Neurodegenerative disorder	Alzheimer's disease	EUR	All	24162737
5	hg18	Longevity	Longevity > 90	EUR	All	24688116
6	hg18	Longevity	Longevity > 85	EUR	All	24688116
7	hg19	Neurodegenerative disorder	Parkinson's disease	EUR	All	19915575
8	hg19	Neuropsychiatric disorder	Schizophrenia	All	All	25056061
9	hg19	Inflammatory diseases	IBD	EUR	All	26192919
10	hg19	Inflammatory diseases	IBD Crohn's disease	EUR	All	26192919
11	hg19	Inflammatory diseases	IBD Ulcerative colitis	EUR	All	26192919
12	hg18	Neuropsychiatric disorder	Bipolar disorder	All	All	21926972
13	hg18	Neuropsychiatric disorder	ADHD	All	All	20732625
14	hg18	Neuropsychiatric disorder	Major depression disorder	EUR	All	22472876
15	hg18	Metabolic outcomes and diseases	Type 2 diabetes	EUR	All	22885922
16	hg18	Metabolic outcomes and diseases	Fasting glucose	EUR	All	22581228
17	hg18	Metabolic outcomes and diseases	Fasting insulin	EUR	All	22581228
18	hg18	GIANT Body fat distribution	Hip AllAncestries	ALL	M&F	25673412
19	hg18	GIANT Body fat distribution	Hip EUR	EUR	M&F	25673412
20	hg18	GIANT Body fat distribution	Hip AllAncestries(Males)	ALL	M	25673412
21	hg18	GIANT Body fat distribution	Hip EUR (Males)	EUR	M	25673412

22	hg18	GIANT Body fat distribution	Hip AllAncestries(Females)	ALL	F	25673412
23	hg18	GIANT Body fat distribution	Hip EUR (Females)	EUR	F	25673412
30	hg18	GIANT Body fat distribution	Waist circumference AllAncestries	ALL	M&F	25673412
31	hg18	GIANT Body fat distribution	Waist circumference EUR	EUR	M&F	25673412
32	hg18	GIANT Body fat distribution	Waist circumference AllAncestries(Males)	ALL	M	25673412
33	hg18	GIANT Body fat distribution	Waist circumference EUR (Males)	EUR	M	25673412
34	hg18	GIANT Body fat distribution	Waist circumference AllAncestries(Females)	ALL	F	25673412
35	hg18	GIANT Body fat distribution	Waist circumference EUR (Females)	EUR	F	25673412
42	hg18	GIANT Body fat distribution	Waist to hip ratio AllAncestries	ALL	M&F	25673412
43	hg18	GIANT Body fat distribution	Waist to hip ratio EUR	EUR	M&F	25673412
44	hg18	GIANT Body fat distribution	Waist to hip ratio AllAncestries(Males)	ALL	M	25673412
45	hg18	GIANT Body fat distribution	Waist to hip ratio EUR (Males)	EUR	M	25673412
46	hg18	GIANT Body fat distribution	Waist to hip ratio AllAncestries(Females)	ALL	F	25673412
47	hg18	GIANT Body fat distribution	Waist to hip ratio EUR (Females)	EUR	F	25673412
54	hg18	GIANT BMI & Height	BMI	EUR	All	25673413
55	hg18	GIANT BMI & Height	Height	EUR	All	20881960
56	hg19	Neurodegenerative disorder	Frontotemporal dementia	EUR	All	24943344
57	hg19	Neurodegenerative disorder	FTD Behavioral variant	EUR	All	24943344
58	hg19	Neurodegenerative disorder	FTD with motor neuron disease	EUR	All	24943344
59	hg19	Neurodegenerative disorder	FTD progressive non-fluent aphasia	EUR	All	24943344
60	hg19	Neurodegenerative disorder	FTD semantic dementia	EUR	All	24943344
61	hg19	Neurodegenerative disorder	Huntington's disease age onset	EUR	All	26232222
62	hg19	Behavioral phenotype	Educational attainment	EUR	All	27225129
63	hg19	Behavioral phenotype	Educational attainment (Males)	EUR	All	27225129

64	hg19	Behavioral phenotype	Educational attainment (Females)	EUR	All	27225129
65	hg18	Reproductive aging	Age at menarche	EUR	All	25231870
66	hg18	Reproductive aging	Age at menopause	EUR	All	26414677
67	hg18	Lipid panel outcomes	HDL		All	24097068
68	hg18	Lipid panel outcomes	LDL		All	24097068
69	hg18	Lipid panel outcomes	Total cholesterol		All	24097068
70	hg18	Lipid panel outcomes	Triglyceride		All	24097068
71	hg18	Reproductive aging	Leukocyte telomere length	EUR	All	23535734
72	hg19	DNAm biomarkers	AgeAccelGrim EUR	EUR	All	
73	hg19	DNAm biomarkers	DNAmGranAdjustedAge EUR	EUR	All	
74	hg19	DNAm biomarkers	AgeAccelHannum EUR	EUR	All	
75	hg19	DNAm biomarkers	DNAmPAI1AdjAge EUR	EUR	All	
76	hg19	DNAm biomarkers	IEAA EUR	EUR	All	
77	hg19	DNAm biomarkers	AgeaccelPhenoAge EUR	EUR	All	
78	hg19	DNAm biomarkers	AgeAccelGrim AFR	AFR	All	
79	hg19	DNAm biomarkers	DNAmGranAdjustedAge AFR	AFR	All	
80	hg19	DNAm biomarkers	AgeAccelHannum AFR	AFR	All	
81	hg19	DNAm biomarkers	DNAmPAI1AdjAge AFR	AFR	All	
82	hg19	DNAm biomarkers	IEAA AFR	AFR	All	
83	hg19	DNAm biomarkers	AgeaccelPhenoAge AFR	AFR	All	
84	hg19	DNAm biomarkers	AgeAccelGrim All	EUR+AFR	All	
85	hg19	DNAm biomarkers	DNAmGranAdjustedAge All	EUR+AFR	All	
86	hg19	DNAm biomarkers	AgeAccelHannum All	EUR+AFR	All	
87	hg19	DNAm biomarkers	DNAmPAI1AdjAge All	EUR+AFR	All	
88	hg19	DNAm biomarkers	IEAA All	EUR+AFR	All	
89	hg19	DNAm biomarkers	AgeaccelPhenoAge All	EUR+AFR	All	
90	hg19	Longevity	Father's attained age	EUR	All	29227965
91	hg19	Longevity	Mother's attained age	EUR	All	29227965
92	hg19	Longevity	Parental attained age	EUR	All	29227965
93	hg19	Age related phenotype	Atrial fibrillation	EUR	All	30061737
94	hg19	Neurodegenerative disorder	Alzheimer's disease	EUR	All	30617256
95	hg19	Cognitive related	Intelligence	EUR	All	29942086
96	hg19	Reproductive aging	AgeAtMenarche	EUR	All	28436984

97	hg19	Neurodegenerative disorder	Huntington's disease motor progression	EUR	All	28642124

EUR: Europeans; AFR: Africans; ASN: Asians.
GWAS resulted in index 72-89 are published in https://www.biorxiv.org/content/10.1101/2020.06.29.133702v1.

References

1. Arneson, A. *et al.* A mammalian methylation array for profiling methylation levels at conserved sequences. *bioRxiv*, 2021.01.07.425637 (2021).
2. Bors, E.K. *et al.* An epigenetic clock to estimate the age of living beluga whales. *bioRxiv*, 2020.09.28.317610 (2020).
3. Horvath, S. *et al.* DNA methylation study of age and sex in baboons and four other primates. *bioRxiv*, 2020.11.29.402891 (2020).
4. Horvath, S. *et al.* Reversing age: dual species measurement of epigenetic age with a single clock. *bioRxiv*, 2020.05.07.082917 (2020).
5. Horvath, S. *et al.* Epigenetic clock and methylation studies in the rhesus macaque. *bioRxiv*, 2020.09.21.307108 (2020).
6. Horvath, S. *et al.* DNA methylation age analysis of rapamycin in common marmosets. *bioRxiv*, 2020.11.21.392779 (2020).
7. Jasinska, A.J. *et al.* Epigenetic clock and methylation studies in vervet monkeys. *bioRxiv*, 2020.09.09.289801 (2020).
8. Kordowitzki, P. *et al.* Epigenetic clock and methylation study of oocytes from a bovine model of reproductive aging. *bioRxiv*, 2020.09.10.290056 (2020).
9. Lemaître, J.-F. *et al.* Epigenetic clock and DNA methylation studies of roe deer in the wild. *bioRxiv*, 2020.09.21.306613 (2020).
10. Prado, N.A. *et al.* Epigenetic clock and methylation studies in elephants. *bioRxiv*, 2020.09.22.308882 (2020).
11. Raj, K. *et al.* Epigenetic clock and methylation studies in cats. *bioRxiv*, 2020.09.06.284877 (2020).
12. Sailer, L.L. *et al.* Pair bonding slows epigenetic aging and alters methylation in brains of prairie voles. *bioRxiv*, 2020.09.25.313775 (2020).
13. Schachtschneider, K.M. *et al.* Epigenetic clock and DNA methylation analysis of porcine models of aging and obesity. *bioRxiv*, 2020.09.29.319509 (2020).
14. Sugrue, V. *et al.* Castration delays epigenetic aging and feminises DNA methylation at androgen-regulated loci. *bioRxiv*, 2020.11.16.385369 (2020).
15. Wilkinson, G.S. *et al.* Genome Methylation Predicts Age and Longevity of Bats. *bioRxiv*, 2020.09.04.283655 (2021).