bioRxiv preprint doi: https://doi.org/10.1101/2020.11.29.402891; this version posted November 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

DNA methylation study of age and sex in baboons and four other primates

Steve Horvath [1,2 #]*, Amin Haghani[1#], Joseph A. Zoller [2], Jason Ernst[3], Matteo Pellegrini [4], Anna J. Jasinska [5], Julie A. Mattison [6], Adam B. Salmon[7], Ken Raj[8], Susan Jenkins[9,10] Cun Li[9,10] , Peter W. Nathanielsz[9,10] *

Affiliations

1 Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, California, USA;
2 Department of Biostatistics, Fielding School of Public Health, University of California, Los Angeles, Los Angeles, California, USA;
3 Department of Biological Chemistry, University of California, Los Angeles, Los Angeles, CA, 90095, USA;
4 Department of Molecular, Cell and Developmental Biology, University of California Los Angeles, Los Angeles, CA 90095, USA;
5 Center for Neurobehavioral Genetics, Semel Institute for Neuroscience and Human Behavior, Department of Psychiatry and Biobehavioral Sciences, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, California, USA;
6 Translational Gerontology Branch, National Institute on Aging Intramural Research Program, National Institutes of Health, USA;
7 The Sam and Ann Barshop Institute for Longevity and Aging Studies, and Department of Molecular Medicine, UT Health San Antonio, and the Geriatric Research Education and Clinical Center, South Texas Veterans Healthcare System, San Antonio TX, USA;
8 Radiation Effects Department, Centre for Radiation, Chemical and Environmental Hazards, Public Health England, Chilton, Didcot, UK;
9 Texas Pregnancy & Life-course Health Center, Southwest National Primate Research Center, P.O Box 760549, San Antonio, Texas 78246, USA;
10 Department of Animal Science, College of Agriculture and Natural Resources Department 3684, 1000 East University Avenue, Laramie, Wyoming 82071-2000, USA

 # co-first authorship

** corresponding authors*
1) Steve Horvath, PhD, ScD E-mail: shorvath@mednet.ucla.edu Address: Gonda Building, 695 Charles Young Drive South, Los Angeles, CA 90095
Steve Horvath (shorvath@mednet.ucla.edu)

2) Peter W. Nathanielsz, MD, PhD, Sc.D. E-mail:peter.nathanielsz@uwyo.edu. Address: Department of Animal Science, College of Agriculture and Natural Resources, Department 3684, 1000 East University Avenue, Laramie, Wyoming 82071-2000

Key words: baboon, aging, development, epigenetic clock, DNA methylation

Emails:
shorvath@mednet.ucla.edu
ahaghani@g.ucla.edu
jaz18@g.ucla.edu
jason.ernst@ucla.edu
matteop@mcdb.ucla.edu
ankajasinska@gmail.com
Julie.mattison@nih.gov
ken.raj@phe.gov.uk

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.29.402891; this version posted November 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

salmona@uthscsa.edu
sjenki13@uwyo.edu
cli5@uwyo.edu
peter.nathanielsz@uwyo.edu

ABSTRACT
DNA methylation data have been successfully used to develop highly accurate estimators of age ("epigenetic clocks") in several mammalian species. With a view of extending epigenetic clocks to primates, we analyzed DNA methylation profiles from five primate species; *Papio hamadryas* (baboons), *Callithrix jacchus* (common marmoset), *Chlorocebus sabaeus* (vervet monkey), *Macaca mulatta* (rhesus macaque), and *Homo sapiens* (human). From these we present here, a highly accurate primate epigenetic clock. This clock is based on methylation profiles of CpGs that are highly conserved and are located on a custom methylation array (HorvathMammalMethylChip40). Furthermore, we carried out in-depth analysis of the baboon, as it is evolutionarily the closest primate to humans that can be employed in biomedical research. We present five epigenetic clocks for baboons (Olive-yellow baboon hybrid), one of which, the pan tissue epigenetic clock, was trained on seven tissue types (fetal cerebral cortex, adult cerebral cortex, cerebellum, adipose, heart, liver, and skeletal muscle) with ages ranging from late fetal life to 22.8 years of age. To facilitate translational capability, we constructed two dual-species, human-baboon clocks, whereby one measures ages of both species in units of years, while the other reports ages relative to the maximum lifespan of the species. Although the primate clock applies to all five primate species, the baboon-specific clocks exhibit only moderate age correlations with other primates. We also provide detailed gene and pathway analyses of individual CpGs that relate to age and sex across different primate species. Ten out of 739 sex related CpGs in primate species are located near 9 autosomal genes (including *FAM217A, CDYL, POU3F2*, and *UHRF2*). Overall, this study sheds light on epigenetic aging mechanisms in primates, and the potential influence of sex.

Keywords: Epigenetic clock, baboon, primate, age, sex, DNA methylation

INTRODUCTION

During development, germline DNA methylation is erased, and then re-established in tissue-specific patterns as the developmental program unfolds after implantation[1]. The primary role of DNA methylation is to regulate the expression of genes, ensuring the production of appropriate ones in different tissues and organs. It was recognized for a long while that the average level of DNA methylation decreases with age, but the consequence of this remained speculative as it was unclear whether expression of specific genes were altered or if this change was entropic in nature, and as such with no deterministic or predictive consequence. Our understanding has since progressed rapidly as a result of the technical advancement of methylation array platforms [2-11], which simultaneously quantify thousands of individual CpGs at known locations on the human genome. It was observed that age-based methylation changes accompany the functional decline of adult stem cells [12-14] and that even small methylation changes can lead to loss of regulatory control of transcription, either directly or via additive effects [15]. Crucially, the correlation between methylation changes (increased or decreased) of some CpGs with chronological age over the course of an entire lifespan was found to be very strong indeed [16-19]. This allowed methylation levels of several CpGs to be combined and developed into an accurate age estimator (reviewed in [19,20]). An example of this is the human pan tissue epigenetic age estimator which combines the weighted methylation average of 353 CpGs, into an age estimate that is referred to as DNAm age or epigenetic age. The quintessential observation that emerged from the use of these human epigenetic clocks is that the discrepancy between epigenetic age and chronological age, termed epigenetic age acceleration, is associated with mortality risk and a wide range of age-related conditions [21-25]. This implies that epigenetic clocks successfully capture biological age, if not in its entirety, then at least to some measurable level. This encouraged the development of similar clocks for animals, as the human pan tissue clock, whilst

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.29.402891; this version posted November 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

directly applicable to chimpanzee [26], could not be used for other species due to evolutionary sequence divergence.

Here, we describe the development of DNA methylation-based biomarkers known as epigenetic clocks, for primates. For ethical and practical reasons, these studies are not possible with the great apes. Instead, we focused on the olive-yellow hybrid baboons living in the Southwest National Primate Research Center (SNPRC). These Old World monkeys can be employed to study biological changes including age-related processes [27], which are expected to be very much closer to those of humans than those exhibited by mice. Although baboons are an attractive primate model for studying aging, their lifespan, while considerably shorter than that of humans (max. 122.5 years), is still substantial (max. 37.5 years)[28], and leads to high maintenance cost, especially for studies on anti-aging interventions that use longevity as the primary measure of effect. Here we present biomarkers of aging that promise to greatly shorten the duration of baboon studies and hence lower the burden and costs associated with evaluating anti-aging interventions in this important investigative primate species. In addition, these biomarkers can also be used to evaluate the effect of age-related conditions and various stress factors on the aging trajectory [19,29 26 21-25].
In previous publications, we have presented epigenetic clocks for vervet monkey [30], rhesus macaque [31], common marmosets [32]. Here we present epigenetic clocks for baboons and all five primate species. This article addressed four aims. First, to develop a DNA methylation-based estimator of chronological age across the entire lifespan based on new DNA methylation data from baboons (olive-yellow baboon hybrids). Second, to evaluate whether these biomarkers apply to humans, other Old World monkeys (vervet monkey, rhesus macaque) and New World monkeys (common marmoset). Third, to develop an epigenetic clock that applies to all five primate species at the same time. Fourth, to evaluate the effect of age and sex on individual cytosine methylation levels in different tissues and primate species.

Results

We generated high-quality DNA methylation data from 326 samples derived from seven different baboon tissues (fetal cerebral cortex, adult heart, adipose, cerebellum, cerebral cortex, liver, and skeletal muscle). Details of tissue numbers can be found in Table 1. It is a strength of our study that the adult baboon tissues come from animals with comparable age distributions (**Table 1**), which facilitates the comparison of age correlations. Unsupervised hierarchical clustering reveals that the samples cluster largely by tissue type (**Supplementary Figure S1**). The tissue samples from other primate species are described in companion papers (Methods).

Epigenetic clocks for baboons

To arrive at unbiased estimates of the epigenetic clocks, we carried out cross-validation analyses of the training datasets. This generates unbiased estimates of the age correlation R (defined as Pearson correlation between the estimated age (DNAm age) and chronological age) as well as the median absolute error, which indicates concordance of the estimated age with chronological age. The different baboon epigenetic clocks that we constructed differed from each other with regards to tissue type, species, and measure of age. Some clocks apply to all tissues (pan-tissue clocks, **Figure 1A**) while others are tailor-made for specific tissues/organs and named accordingly. We developed brain-specific epigenetic clocks, which were trained using baboon brain regions: cerebellum and frontal cortex (R=0.95, **Figure 1B**, **Supplementary Figure S2**) and the cerebral cortex clock (R=0.97, **Figure 1C**). The baboon pan-tissue clock, which was trained on all available tissues, is highly accurate in age estimation of all the different baboon tissue samples (R=0.96 and median error 1.1 years, **Figure 1A, Supplementary Figure S3**), including cerebellum (R=0.95, **Supplementary Figure S3C**) and frontal cortex samples (R=0.99, **Supplementary Figure S3D**). Although all of these different clocks exhibit high age correlations, pan tissue clocks may have different biological properties from tissue specific clocks, e.g. pan tissue clocks tend to be less correlated with changes in cellular composition [19].

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.29.402891; this version posted November 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

To generate hybrid human-baboon clock, human DNA methylation profiles that were generated using HorvathMammalMethylChip40 were added to the baboon DNA methylation profiles in the training dataset. From these, two human-baboon pan-tissue clocks were developed. One reports age estimates in unit of years while the other reports its estimates as relative age, which is the ratio of chronological age to maximum lifespan of the species, with values between 0 and 1. This ratio allows alignment and biologically meaningful comparison between species with different average and maximal lifespans (baboon, 37.5 years and human, 122.5 years), which is not afforded by mere measurement of absolute age. The human-baboon clock for relative age is highly accurate when applied to both species together (R=0.97, **Figure 1D**) and only marginally less so when the analysis is restricted to baboon tissues (R=0.96, **Figure 1E**).

Application to other primates

To determine the cross-tissue performance and the cross-species applicability of the baboon clocks, we applied them to an array of tissues from key human organs (n=852 from 16 tissues) and two Old World primate species, rhesus macaque (n=283 samples from eight tissues) and vervet monkeys (n=140 from 3 tissues). We also analyzed n=95 blood samples from a New World primate: common marmoset.

The application of pan-tissue baboon clock to *human* tissues produced epigenetic age estimates that are poorly calibrated; meaning that they exhibited poor concordance with chronological age (high median absolute error), but they nevertheless showed low to moderate correlations with ages with several human tissues (e.g. human adipose, human blood, human bone marrow, heart, kidney, lung, skeletal muscle, skin, and spleen, **Supplementary Figure S4**).

The estimated age of vervet monkeys by the baboon clocks also exhibited high correlations (R>=0.86, **Supplementary Figure S5**), but weak concordance with chronological age. The human-baboon clock for absolute age works particularly well, albeit with increasing off-set of DNAm age with chronological age (R=0.91, **Supplementary Figure S5D**). When pure baboon clocks were applied to DNA methylation profiles of rhesus macaque, the estimated ages correlated moderately with chronological age (r<=0.59, **Supplementary Figure S6A-C**) and performed rather poorly with regards to concordance with chronological age. Once again, the human-baboon clocks for chronological and relative age generated much better age correlations (R>=0.84, **Supplementary Figure S6D-E**). Overall, the human-baboon clocks generated age estimates with the highest correlations with age of both the non-human primates.

We also applied the baboon clocks to blood samples from common marmosets (**Supplementary Figure S7**). We observe moderately high age correlations (e.g. the baboon pan tissue clock leads to r=0.76, **Supplementary Figure S7A**) but the age estimates are highly biased and accompanied by large median errors.

Age related CpGs

Epigenome-wide association studies (EWAS) of chronological age revealed a largely tissue-specific pattern of DNAm change (**Figure 3A**). Age-associated CpGs methylation changes in one tissue are rarely conserved in another tissue (**Supplementary Figure S8**) which is similar to what has been observed in humans [33]. However, the low conservation and differences in p value ranges in our analyzed tissue types may reflect a limited sample size of non-cortex tissues (see Table 1). With this caveat in mind, it is nevertheless notable that aging of some tissues such as adipose, cerebellum and heart, is accompanied largely by increased methylation of age associated CpG, while that of muscle is primarily the opposite. To capture the top affected loci in all tissues, a nominal p value $< 10^{-4}$ was employed. This highlighted the top CpG methylation changes and their proximal genes in each tissue as follows: adipose, *KLF14* promoter (z=6.2); cerebellum, *ESRRB* intron (z=7.5); cerebral cortex, *HOXC4* promoter (z=19.9); heart, *KIF15* promoter (z=5.6); liver, *TNRC6A* exon (z=-9.4); and muscle, *ASCC2* exon (z=-6.1). To identify age associated CpGs that are shared across the tissue, we carried out meta-analyses on these six tissue samples, and the top DNAm changes are: hypermethylation in *KLF14* (z=16.7), *KIF15* (z=16.3), and *HOXC4* (z=15) promoters. In addition, two CpGs in the exon of *FOXD3* (z=13.3) were also hypermethylated with age (**Figure 3A**).

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.29.402891; this version posted November 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

In addition to identifying individual CpGs with methylation levels that alter with age, it is often more informative to also ascertain whether these age-associated CpGs of the various tissues are clustered in certain intracellular pathways, disease-related pathways and other biological effects. To ascertain this, enrichment analyses of age associated CpGs were carried out and the detailed results are presented in **Supplementary Figure S9**. In general, a few major characteristics are apparent. The first is that many enriched features are specific to the respective tissue, e.g. the near absence of any enrichment of the age related CpGs in heart tissue contrasts with the many enrichments observed for liver tissue. The age associated CpGs shared across different tissues (meta-analysis) mostly enriched in biological features related to development, Polycomb repressive complex 2 targets, and histone mark H3K27ME3 which plays a role in determining whether a stem cell will remain unspecified or will eventually differentiate.

We mapped the positions of CpGs to the genome and observed that age related CpGs map to both genic as well as intergenic regions, which are defined relative to transcriptional start and end sites (**Figure 3B**). Although the HorvathMammalMethylChip40 contained more CpGs in exons than other genomic regions, the number of age-related CpGs in adipose tissue is most common in the promoter region, with all of them becoming increasingly methylated with age. To a lesser degree, this is also observed with cerebellum, heart and liver. Interestingly, in all tissues except muscle, most age associated CpGs in promoter region gained methylation with age. In muscle however, the trend was in the negative direction. Interestingly, the distribution of age associated CpGs in muscle genome mirrors CpG distribution of the HorvathMammalMethylChip40, highlighting yet again that methylation patterns in muscle are distinct from those of other tissues.
We are not sure whether the distinct results observed for muscle arise from true biological differences or whether they represent a technical issue.

To obtain a perspective of over-lapping age associated CpGs between different tissues, an *upset* plot, which is akin to the alignment of multiple Venn diagram intersections was generated. From this, we identified CpGs that showed consistent age associated DNAm change in different combinations of multiple tissues. The *KLF14* promoter was implicated by the EWAS analysis of age in all tissues (**Figure 3C**). The KLF14 protein is a transcriptional factor that regulates TGFBII receptor expression and has been implicated in diabetes onset in humans [34]. Our EWAS of age also implicated another gene *KIF15* in all but one tissue.

We examined the transcriptional factor (TF)-binding motif enrichment of the top age-associated CpGs located within the promoter or 5'UTR in each tissue (**Figure 3D**). A very clear tissue-specific pattern is observed, with mutually exclusive motifs for each tissue type. This echoes the general tissue specific age-related CpGs seen in Figure 3A. The top TF motifs in the meta-analysis were ZIC1, ZIC2, and ZIC3, which had > 5 CpGs with age-associated hypomethylation. The other prominent motif that contained more than five CpGs that become increasingly methylated with age is TFAP2C, which regulates the expression of genes involved in development of multiple organs. We caution the reader that some of these motifs may reflect a g-c bias.

Conserved DNA methylation differences by sex in primates

We excluded the common marmoset from our DNA methylation study of sex differences in primates, because marmosets appear to have a unique characteristic when it comes to DNA methylation studies in mammals: it is not possible to predict sex on the basis of methylation data in common marmosets [32].
We used the data from the remaining four primates to identify sex related CpGs in multiple tissues. In total, the analysis involved 29 regression models that examined sex differences in 13 tissues from human, baboon, vervet monkey, and macaque from prenatal time to late adulthood. The median number of sex-associated methylation positions (SMP) in all these models was 1125 CpGs (5% False Discovery Rate), most of which were located on the X chromosome (**Figure 4A-B**). The sex difference was more prominent in older individuals than neonates, particularly in macaque and vervet monkey blood samples, which had >10000 SMPs compared to other tissues (**Figure 4A**). In humans, the heart and muscle samples had a higher number of SMPs (>2000) compared to other tissues. In specific tissues (heart and muscle in humans, blood in macaque

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.29.402891; this version posted November 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

and vervet monkey), we find that that a surprisingly large number of SMPs are located on autosomes (**Figure 4B**). But in most tissues, the number of *autosomal* SMPs is very limited (fewer than 200 in most tissues, Figure 4B). In total, 739 SMPs were highly conserved in all tissues of these four species and 729 of these are located on sex chromosomes. The 10 CpGs (1.3%) that are located on autosomal chromosomes are proximal to 9 genes (depending on the species) (**Figure 4A**). Four of these genes (*FAM217A*, *CDYL*, *POU3F2*, and *UHRF2*) were located in human chromosome 6. The remaining autosomal sex related CpGs are located near *MYCL*, *ZXDC*, *RRAS*, *RBM39* (Figure 4A).

Only 3 CpGs relate both to sex and age: these CpGs are proximal to *HUWE1*, *BCOR*, and *MECP2*.

As expected, enrichment analysis of the sex-associated CpGs highlighted gonosomal and X-linked inheritance for CpGs located on both sex or autosomal chromosomes in all primates (**Supplementary Figure S10**). The genes on sex chromosomes were also related to testicular development and several cognitive and neuronal pathways including intellectual disability, cognitive impairment, autism spectrum disorder, mental health, agitation, and thermal nociceptive threshold. Strikingly, the genes on the autosomal chromosome were also enriched in several neuronal pathways such as GABA receptor activation, intellectual disability, and synaptic function. Thus, our result suggests a possible difference in neuronal development, activity or function between sexes across these four primate species that are possibly regulated by differential methylation between them. These differences are apparent from the fetal stage to the end of lifespan. Detailed enrichment analysis of sex-differences in baboons can be found **Supplementary Figures S12.**

Large human cohort studies allowed us to test whether sex affected the methylation patterns of human genes. We analyzed human Illumina EPIC array data from three large studies: blood samples from individuals of European ancestry (Framingham Heart study), blood samples from individuals of African ancestry (Jackson Heart study), and postmortem prefrontal cortex samples from individuals from the Religious Order study (ROSMAP).

Although the EPIC array does not have good coverage for conserved CpGs between primates, it nevertheless contains almost all human gene-regions represented in mammalian array (**Supplementary Figure S13**). Thus, following EWAS of sex in EPIC array data, we only studied the gene-regions proximal to identified SMPs in primate analysis. We found highly significant sex related CpGs both in blood and brain samples (**Supplementary Figure S14**). Several gene-regions proximal to these SMPs were also differentially methylated between human males and females in both blood and brain samples (**Figure 5**). Sex-related genes found both in blood and brain samples include the *EFNB1* promoter and *BCOR* intron on X chromosome, and TLE1 5'UTR on human chromosome 9.

Discussion

Apart from a primate clock that applies to the five considered primate species, we present five epigenetic clocks that are tailor-made for baboons and apply to the entire life course (from birth to old age). The application of baboon clocks on DNA methylation profiles derived from the other two members of the Old World monkeys yielded an unexpected finding. The baboon and rhesus macaque lines separated approximately 9 million years ago, while the baboon and vervet monkeys separated a million years before that, i.e. 10 million years ago [35]. As such, the macaque is phylogenetically closer to baboon than the vervet monkey is. Despite this, the baboon epigenetic clock performed better with vervet monkeys, generating age estimates that are highly correlated with chronological age albeit with poor concordance. Rhesus macaque epigenetic age, as measured by the baboon clock, correlated considerably less well with chronological age and exhibited even lower age concordance. This curious observation is compounded by the fact that epigenetic ages of rhesus macaque that were derived using both the human-baboon epigenetic clocks (reporting in units of years or relative age) correlated and concurred better with their chronological ages. This improvement is puzzling because humans separated from the old-world monkey line approximately 23-25 million years ago. The effect of this divergence is also seen in this study where epigenetic ages of human tissues as estimated by the baboon pan tissue clock were poorly correlated with age and exhibited poor concordance. Yet, the inclusion of DNA methylation profiles from human tissues, improved the clocks' ability to estimate ages of rhesus macaque, which shared a line with baboons 9 million years ago. This improvement may reflect that the

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.29.402891; this version posted November 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

hybrid human-baboon clock is based on CpGs that are better preserved across primate species. The successful generation of the primate clock demonstrates the feasibility of building epigenetic clocks for five species based on a single mathematical formula. This effectively consolidates the notion that certain epigenetic aspects of aging are highly conserved between species. However, there are also profound differences in age related methylation changes across species and even tissues from the same species. When age associated CpGs were analyzed, their tissue specificity quickly became apparent. This can appear at first light to be inconsistent with the successful development of a pan tissue clock. It is however important to bear in mind that age associated CpGs that constitute the clock are the most predictive of age across all tissues. These are not necessarily the highest scoring ones for individual tissues. As such, it appears that as tissues age, two sets of CpGs undergo methylation change; the first is tissue-specific and the second is general across all tissues, which is consistent with what has been observed in humans [5,33,36].

The cross-tissue analysis of age-associated CpGs in baboons identified the locus of *KLF14*, which is a transcription regulator that can activate and also repress the expression of genes. Although its direct targets remain to be fully characterized, its effect on the pathology is known to include prevention of cardiovascular disease [37] and suppression of cancers [38]. Computational predictions of its targets include genes that are involved in regulation of cholesterol, body-mass index, insulin and blood glucose level [39]. Although blood was not one of the tissues analyzed in baboons, it is nevertheless noteworthy that methylation of the *KLF14* locus in blood of mice is predictive of chronic inflammation in adipose tissue [40], suggesting that this cross-tissue age-associated loci might extend to other tissues and may bear even greater predictive scope. Indeed, methylation of this loci in blood, saliva and buccal swabs is predictive of age in humans [41], and along with *ELOVL2*, the *KLF14* is one of the most consistently identified loci with age-associated methylation change [42]. Our meta analysis of age across baboon tissues also implicated CpGs near *HOXC4*, *HOXA11* and *SHOX2* genes that gain methylation with age (Figure 3C). These are members of the Hox family, which is one of the oldest gene sets that are conserved across not only in mammals but even arthropods. Hox proteins specify development of body segments of organisms and are pivotal for normal healthy development. Their expression is regulated by polycomb repressive complex [43], whose DNA targets are consistently identified, as well in this study, as being increasingly methylated with age. Hence the involvement of HOX loci further consolidates the importance of the process of development in aging. This is consistent with enrichment analyses of age associated CpGs that identified processes of development as being implicated in aging across the different tissues. This is further substantiated by the identification of TFAP2C binding sites as one of the major transcription factor motifs that are implicated in age-associated methylation. TFAP2C is a transcription factor that regulates expression of developmental genes [44,45]. These multiple lines of evidence point repeatedly to the importance of developmental processes in aging, which echoes the findings from human DNA methylation profiles and those of other mammals.

Sex is a potential biological variable that affects both the lifespan and healthspan of an organism. In humans, the rate of epigenetic aging in women is lower than that of men [46]. The underlying reason for this is still unknown and therefore the availability of DNA methylation profiles of both sexes in four different primate species presents an opportunity to identify conserved epigenetic features that may provide some clues. This analysis did indeed yield interesting results. First, it is very clear that except for three CpGs (Figure 4C), age-associated CpGs do not overlap with sex-associated ones, underlying the distinct nature of these two features. While the majority of the sex associated CpGs expectedly reside in sex chromosomes, some were found in autosomes. This is particularly pronounced in human muscle and heart. One of the autosomal genes with sex associated CpG methylation is *CDYL*, which encodes a positive regulator of polycomb repressive complex 2 (PRC2) [47,48]. This is particularly relevant as PRC2 target sites become increasingly methylated with age, across different mammalian species, including humans. The collective evidence from many different species analyzed thus far repeatedly points to PRC2 as a pivotal factor underlying epigenetic aging. Therefore, the difference in methylation of *CDYL* locus could potentially result in differences in PRC2 activity, which in turn could cause a difference in epigenetic aging rates between the sexes.

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.29.402891; this version posted November 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Sex affects the blood methylome in a species and life stage dependent manner. In neonatal blood from vervets, we observed few autosomal CpGs that relate to sex but this number increases dramatically in older animals. While blood samples from older macaques also reveal a large number of autosomal CpGs that relate to sex, the same could not be observed in human blood samples.

Beyond their utility, the primate and baboon epigenetic clocks reveal several salient features with regard to the biology of aging. First, the fact that these are pan-tissue clocks re-affirm the notion that aging is a coordinated biological process that is harmonized throughout the body. Second, the process of development is again highlighted as being implicated in the process of epigenetic aging. Third, although age-associated CpGs and sex-associated CpGs are largely mutually exclusive, they may nevertheless interact further downstream to affect different rates of aging between the sexes. Fourth, the ability to develop human-baboon clocks and even primate clocks attests to the high conservation of the aging process across evolutionary distinct primate species. This implies, albeit does not guarantee, that treatments that alter the epigenetic age of baboons, as measured using the human-baboon clock are very likely to exert similar effects in humans.

We expect that the development of the baboon and primate clocks will provide a significant boost to the potential use of baboons as models for human aging. Lifespan and healthspan studies in primates are expensive and time-consuming. These drawbacks are mitigated by these epigenetic clocks, which can be easily measured within the life-time of the animal, as opposed to the measure of longevity, which is a distinct parameter that is not necessarily indicative of biological aging.

Materials and Methods

Baboon care and maintenance

All animals were given a full veterinary examination prior to recruitment to the study and no obvious cause of ill health or pathology was observed. The animals were housed in group cages at the Southwest National Primate Research Center, at Texas Biomedical Research Institute (TBRI), in San Antonio, Texas in groups of up to 16. The remaining 4 females were housed in individual cages at the UT Health Sciences Center San Antonio (UTHSCSA).
All procedures were approved by the TBRI or UTHSCSA Animal Care and Use Committee and conducted in facilities approved by the Association for Assessment and Accreditation of Laboratory Animal Care.
Twenty-eight females and the ten males were fed ad libitum Purina Monkey Diet 5038 (12% energy from fat, 0.29% from glucose and 0.32% from fructose and metabolizable energy content of 3.07 kcal/g protein; Purina LabDiets, St Louis, MO, USA) (CTR). Water was continuously available to all animals. Animal health status was recorded daily.

Necropsy
None of the animals were euthanized for this project. Rather, we used left-over frozen tissue samples that had previously been collected as part of other projects. Necropsies were performed by either a qualified, experienced veterinarian or M.D investigator. At TRBI, baboons were pre-medicated with ketamine hydrochloride (10 mg/kg IM) and anesthetized using isoflurane (2%) general anesthesia as previously described [49]. Baboons were exsanguinated while still under general anesthesia as approved by the American Veterinary Medical Association. At UTHSCSA four animals were euthanized using Fatal Plus (Vortech, Dearborn, MI, USA) Pentobarbital 390 mg/ml. Following cardiac asystole and failure of reflex responses to skin pinch and eye touch stimulation, tissues (adipose, cerebellum, cerebral cortex, muscle, heart, liver) were rapidly dissected and immediately frozen in liquid nitrogen.

For the studies in which fetal tissue was obtained, all animals were housed in 20 foot × 20 foot × 15 foot metal and concrete group cages at the Texas Biomedical Research Institute. Experimental animals were obtained

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.29.402891; this version posted November 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

from appropriate groups of 16 healthy female baboons (Papio sp.) of similar pre-study body weights (10–15 kg) and morphometric features (13). Pregnancy was dated initially by timing of ovulation and changes in sex skin color and confirmed at 30 days of gestation (dG) by ultrasonography. Details of housing, feeding, and environmental enrichment have been published elsewhere (13). All procedures were approved by the University of Texas Health Science Center and Texas Biomedical Research Institute internal animal care and use committees and performed in Association for Assessment and Accreditation of Laboratory Animal Care–approved facilities.
Prior to Cesarean section, baboons were premedicated with ketamine hydrochloride (10 mg/kg, IM). Following tracheal intubation, isoflurane (2%, 2L/min, by inhalation) was used to maintain a surgical plane of anesthesia throughout surgery. Cesarean section was performed at gestational day 165 (0.9 of gestation) using standard sterile techniques as previously described [30,50].
Following hysterotomy, the umbilical cord was identified and used for fetal exsanguination with both maternal and fetal baboon under general anesthesia as approved by the American Veterinary Medical Association Panel on Euthanasia [51]. Postoperatively, mothers were placed in individual cages and watched until they were upright under their own power. They were returned to their group cage two weeks postoperatively. Maternal analgesia was administered for 3 days (buprenorphine hydrochloride injection; Hospira, Inc., Lake Forest, IL, USA; 0.015 mg/kg/day).

Animals were individually fed to enable precise regulation of intake either between 7:00 am and 9:00 am or 11:00 am and 1:00 pm as described in detail elsewhere [52]. At feeding time, the weight of each baboon was obtained via an electronic weighing system (GSE 665; GSE Scale Systems, Allen Park, Michigan) in a holding cage as the animal waited in line. Water was continuously available in each feeding cage (Lixit, Napa, California), and the animals were fed Purina Monkey Diet 5038 (Purina, St Louis, Missouri).

For this study, we selected samples representing the entire primate lifespan, from neonate to old age.

Data from other monkeys

The data from other monkeys are described in the companion papers for rhesus macaque (n=283 samples from skin, blood, adipose, cerebral cortex, liver, lung, muscle [31]), vervet monkey (n=240 samples from whole blood, prefrontal cortex and liver [30]), and common marmosets (n=95 blood samples [32]).

Human tissue samples profiled on the mammalian array

To build the human-baboon clock, we analyzed previously generated methylation data from n=850 human tissue samples (adipose, blood, bone marrow, dermis, epidermis, heart, keratinocytes, fibroblasts, kidney, liver, lung, lymph node, muscle, pituitary, skin, spleen) from individuals whose ages ranged from 0 to 93 years. The tissue samples came from three sources: tissue and organ samples from the National NeuroAIDS Tissue Consortium [53], Blood samples from the Cape Town Adolescent Antiretroviral Cohort study [54], skin and other primary cells provided by Ken Raj [55]. Ethics approval (IRB#15-001454, IRB#16-000471, IRB#18-000315, IRB#16-002028).

DNA extraction

DNA was extracted on an automated nucleic acid extraction platform Anaprep (Biochain) using a magnetic bead based extraction method and Tissue DNA Extraction Kit (AnaPrep).

DNA methylation data

In total, 36,523 probes from the HorvathMammalMethylChip40 were aligned to specific loci that are proximal to 6152 genes in the Olive baboon (Papio anubis.Panu_3.0.100) genome. The CpG-containing probes in this array were selected based on their conservation in mammalian genomes. Therefore, findings derived from these data would have high a high level of translatability to humans and other mammals.

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.29.402891; this version posted November 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

All methylation data were generated using a custom Illumina methylation array (HorvathMammalMethylChip40) based on 37492 CpG sites. Out of these 37492 sites, 1951 were selected based on their utility for human biomarker studies; these CpGs, which were previously implemented in human Illumina Infinium arrays (EPIC, 450K, 27K), were selected due to their relevance for estimating human age, human blood cell counts or the proportion of neurons in human brain tissue. The remaining 35541 probes were chosen due to their location in stretches of DNA that are highly conserved across mammalian species (Arneson, Ernst, Horvath, in preparation).
The particular subset of species for each probe is provided in the chip manifest file at the NCBI Gene Expression Omnibus (GEO) platform (GPL28271). The SeSaMe normalization method was used to define beta values for each probe [56].

Penalized Regression models
Details on the clocks (CpGs, genome coordinates) and R software code are provided in the Supplement. Penalized regression models were created with glmnet [57]. We investigated models produced by both "elastic net" regression (alpha=0.5). The optimal penalty parameters in all cases were determined automatically by using a 10-fold internal cross-validation (cv.glmnet) on the training set. By definition, the alpha value for the elastic net regression was set to 0.5 (midpoint between Ridge and Lasso type regression) and was not optimized for model performance.
We performed a cross-validation scheme for arriving at unbiased (or at least less biased) estimates of the accuracy of the different DNAm based age estimators. One type consisted of leaving out a single sample (LOOCV) from the regression, predicting an age for that sample, and iterating over all samples. A critical step is the transformation of chronological age (the dependent variable). While no transformation was used for the pan-tissue clock for baboons, we used a log linear transformation for the dual species clock of absolute age. For the primate clock, we used the following transformation: sqrt(Age+1), i.e. after adding an offset of 1 year, we formed the square root transformation.

Relative age estimation
To introduce biological meaning into age estimates of baboons and humans that have very different lifespan; as well as to overcome the inevitable skewing due to unequal distribution of data points from baboons and humans across age range, relative age estimation was made using the formula: Relative age= Age/maxLifespan where the maximum lifespan for the different species was chosen from the anAge data base [58], e.g. the Human maximum lifespan was determined to be 122.5.

Epigenome wide association studies of age
EWAS was performed in each tissue separately using the R function "standardScreeningNumericTrait" from the "WGCNA" R package[59]. Next the results were combined across tissues using Stouffer's meta-analysis method.

Transcription factor enrichment and chromatin states
The FIMO (Find Individual Motif Occurrences) program scans a set of sequences for matches of known motifs, treating each motif independently [60]. We ran TF motif (FIMO) scans of all probes on the HorvathMammalMethyl40 chip using motif models from TRANSFAC, UniPROBE, Taipale, Taipaledimer and JASPAR databases. A FIMO scan p-value of 1E-4 was chosen as cutoff (lower FIMO p-values reflect a higher probability for the local DNA sequence matching a given TF motif model). This cutoff implies that we find almost all TF motif matches that could possibly be associated with each site, resulting in an abundance of TF motif matches. We caution the reader that our hypergeometric test enrichment analysis did not adjust for GC content and may therefore be biased.

Human studies on the EPIC array

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.29.402891; this version posted November 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Framingham Heart Study. We used blood methylation data from 2,356 individuals composed of 888 pedigrees from the Framingham Heart cohort [61], a large-scale longitudinal study started in 1948, initially investigating risk factors for cardiovascular disease (CVD). The FHS cohort contains blood DNA methylation profiling at exam 8. A human epigenetic clock analysis of these data is presented in many articles including [24,46].

Jackson Heart Study (JHS, N=1747). The JHS is a large, population-based observational study evaluating the etiology of cardiovascular, renal, and respiratory diseases among African Americans residing in the three counties (Hinds, Madison, and Rankin) that make up the Jackson, Mississippi metropolitan area. The age at enrollment for the unrelated cohort was 35-84 years; the family cohort included related individuals >21 years old. JHS ancillary study ASN0104, available with both phenotype and DNA methylation array data.

ROSMAP. We analyzed previously generated DNA methylation data from Caucasian subjects from the Religious Order Study (ROS) and the Rush Memory and Aging Project (MAP) [62,63] Both are longitudinal community based cohort studies of aging and dementia. The majority of participants in both studies are 75–80 years old at baseline with no known dementia. All participants agree to organ donation at death. Participants sign and informed consent, repository consent, and Anatomical Gift Act. A human epigenetic clock analysis of these methylation data is presented in [64].

Acknowledgements
This work was supported by the Paul G. Allen Frontiers Group (SH). We would like to acknowledge support through a 1U19AG057758-01A1 Womb to Tomb: Developmental programming and aging interactions in primates to PWN.
This investigation used resources that were supported by the Southwest National Primate Research Center grant P51 OD011133 from the Office of Research Infrastructure Programs, National Institutes of Health.
We would like to thank the veterinary and technical staff at the Primate Center. For their consistent help and advice on animal management. We would specifically like to express our gratitude to Ms. Susan Jenkins for maintain the databases on our colony.
The rhesus macaque data were funded in part by the Intramural Research Program, National Institute on Aging, NIH.
The Jackson Heart Study (JHS) is supported by contracts HHSN268201300046C, HHSN268201300047C, HHSN268201300048C, HHSN268201300049C, HHSN268201300050C from the National Heart, Lung, and Blood Institute and the National Institute on Minority Health and Health Disparities.
The Framingham Heart Study is funded by National Institutes of Health contract N01-HC-25195 and HHSN268201500001I. The laboratory work for this investigation was funded by the Division of Intramural Research, National Heart, Lung, and Blood Institute, National Institutes of Health. The analytical component of this project was funded by the Division of Intramural Research, National Heart, Lung, and Blood Institute, and the Center for Information Technology, National Institutes of Health, Bethesda, MD. JMM and KLL were supported by R01AG029451.
The Religious Order study and Rush Memory and Aging Project were funded by P30AG10161, R01AG17917, RF1AG15819, R01AG34374, R01AG36042, U01AG46152 (Bennett).

The funding bodies played no role in the design, the collection, analysis, or interpretation of the data.

Conflict of Interest Statement
SH is a founder of the non-profit Epigenetic Clock Development Foundation which plans to license several patents from his employer UC Regents. All of these patents list SH as inventor. One of the patents also lists JE as inventor. The other authors declare no conflicts of interest.

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.29.402891; this version posted November 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Figure legends

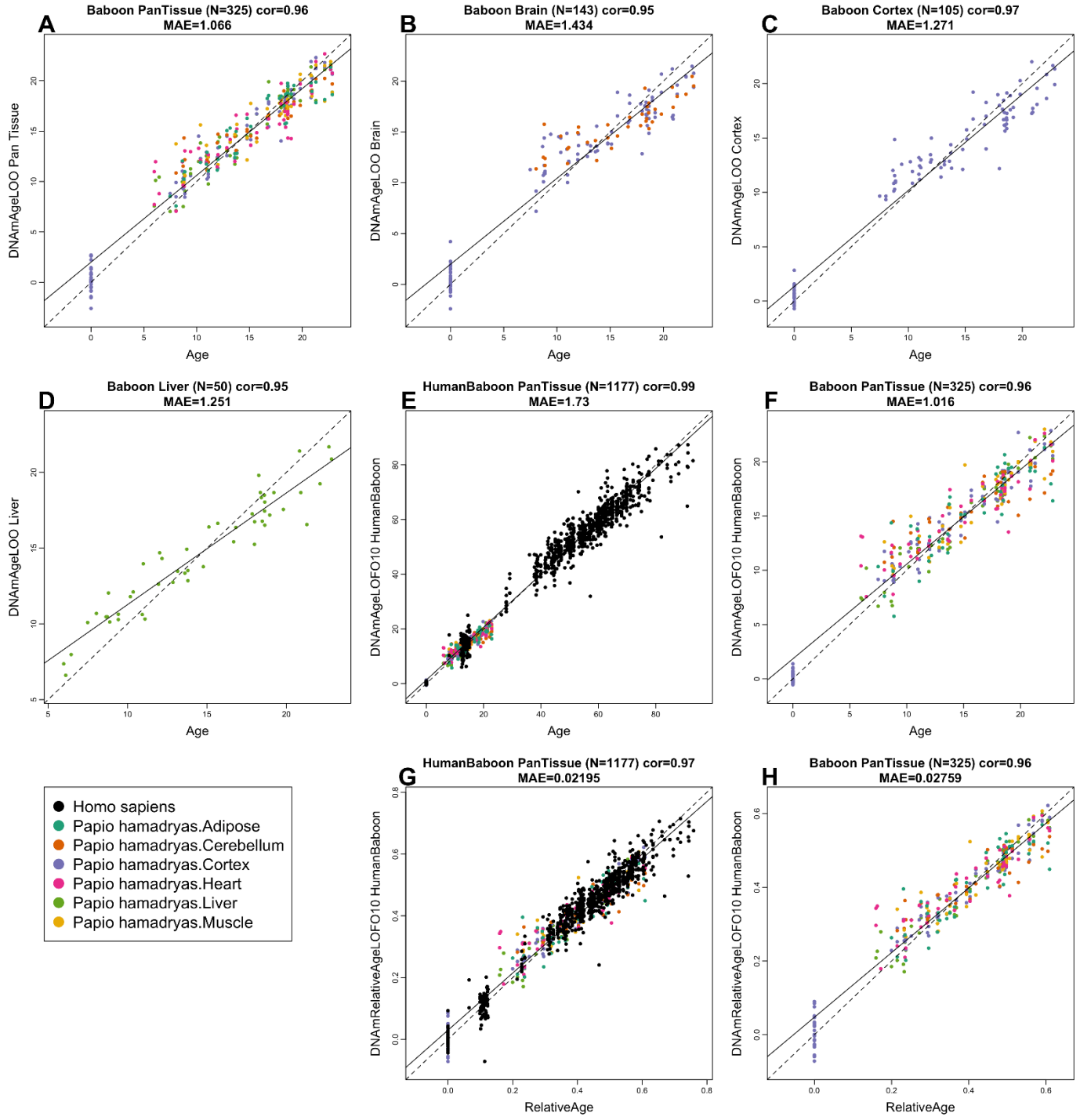
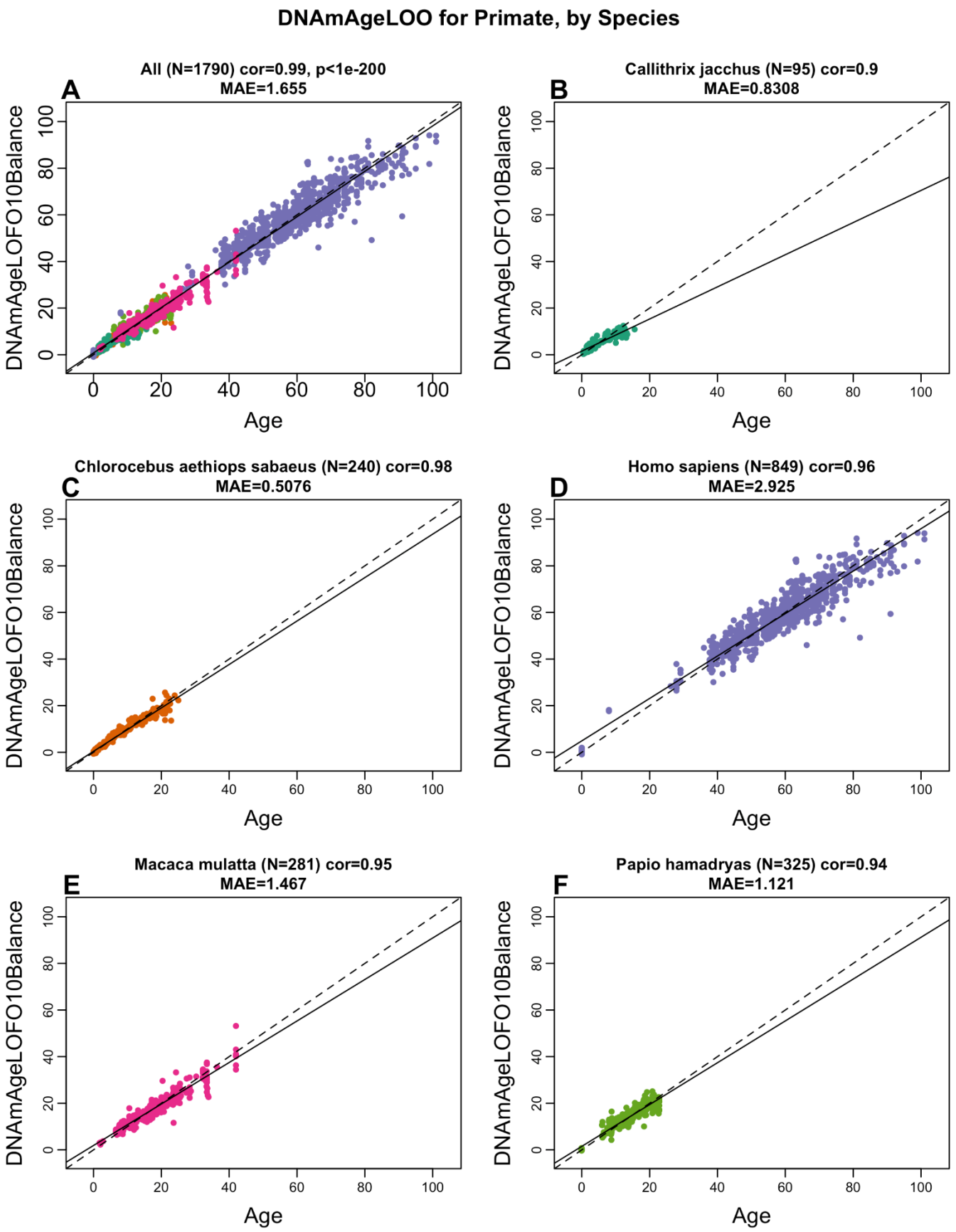


Figure 1: Cross-validation study of epigenetic clocks for baboons and humans. A-C) Three epigenetic clocks that were trained on baboon tissues: A) all tissues, B) all brain tissues, C) cerebral. D,E) Human-baboon clock for relative age applied to D) both species and E) baboons only. Relative age was defined as ratio of chronological age to maximum lifespan. D) Human samples are colored in (black) and baboon samples (colored by tissue type). Each panel reports the sample size, correlation coefficient, median absolute error (MAE). Dots are colored by tissue type or species.
A-C) "LOO" denotes the leave-one-out cross validation estimates of DNA methylation age (y-axis, in units of years). D,E) "LOFO10" denotes the ten-fold cross-validation estimates of age (y-axis, in years).

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.29.402891; this version posted November 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

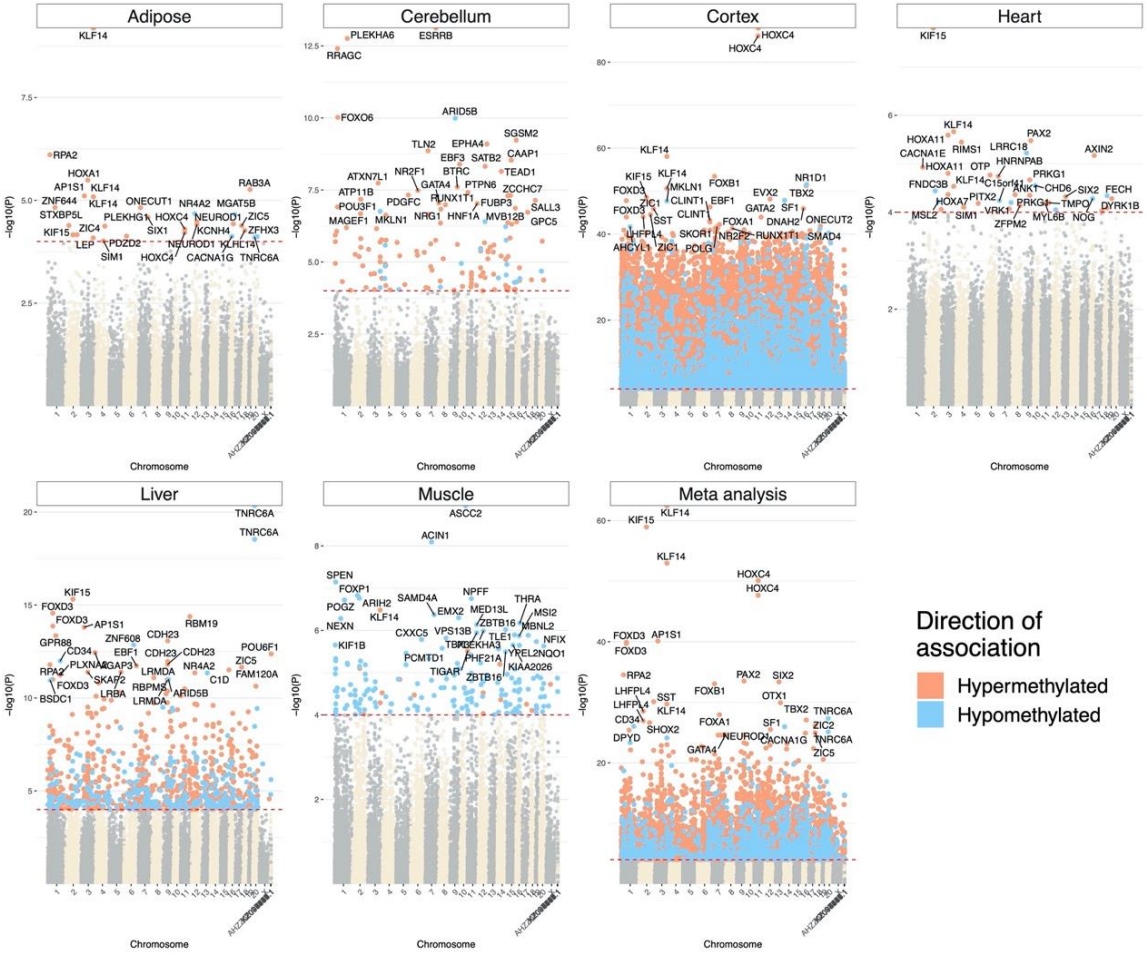


bioRxiv preprint doi: https://doi.org/10.1101/2020.11.29.402891; this version posted November 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

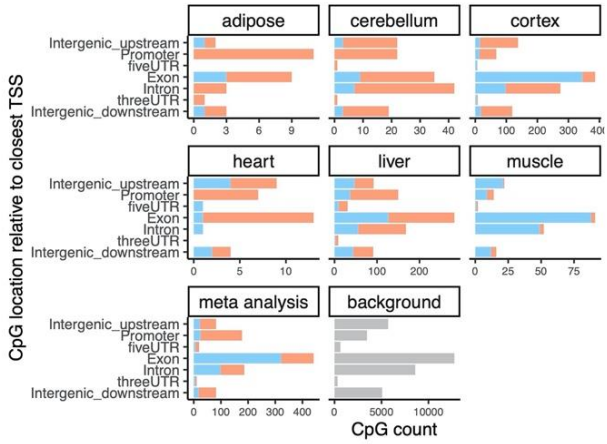
Figure 2: Cross-validation study of the primate clock. A) Epigenetic clock based on 5 different primate species (colored by species as indicated in the other panels). B-F) are excerpts from panel A but restricted to B) Callithrix jacchus (common marmoset), C) Chlorocebus sabaeus (vervet monkey), D) homo sapiens, E) macaca mulatta (rhesus macaque), F) papio hamadryas (baboons). Chronological age (x-axis) versus the ten-fold cross validation (balanced by species) of the species (y-axis). Each panel reports the sample size, correlation coefficient, median absolute error (MAE). Dots are colored by species.

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.29.402891; this version posted November 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

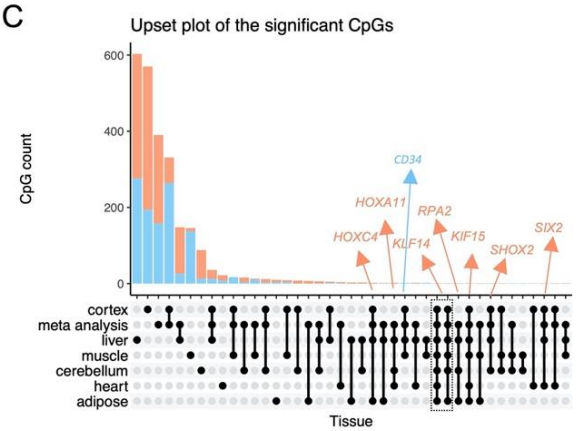
A



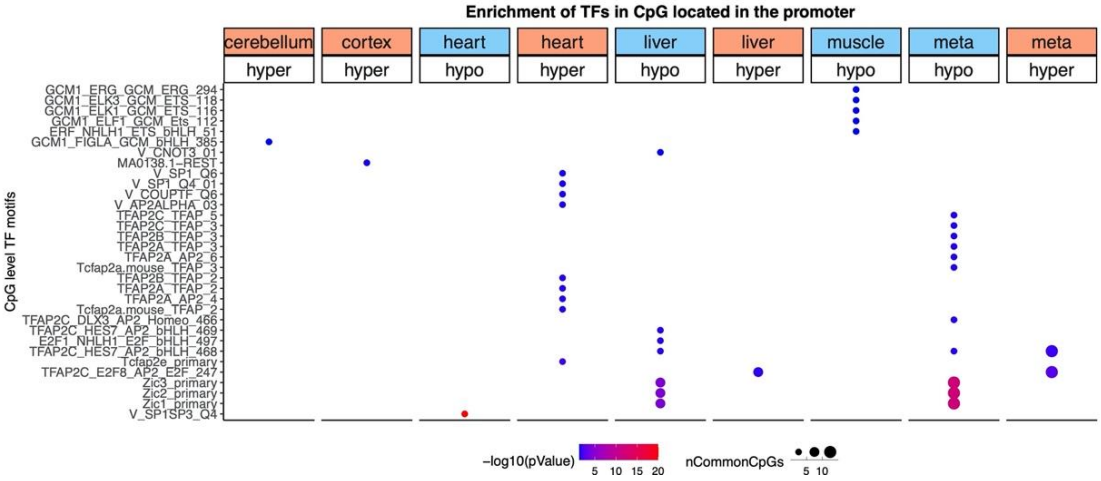
B



C



D



bioRxiv preprint doi: https://doi.org/10.1101/2020.11.29.402891; this version posted November 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Figure 3. Epigenome-wide association (EWAS) of chronological age in baboons. Meta-analysis of correlation tests in adipose, cerebellum, cerebral cortex, heart, liver, muscle of olive baboons. A) Manhattan plots of the EWAS of chronological age. Note that the cerebral cortex leads to more significant p values (-log base 10 transformation of the p value, y-values) which reflects the broader age range of this tissue. The coordinates are estimated based on the alignment of Mammalian array probes to Panu_3.0.100 genome assembly. The direction of associations with $p < 10^{-4}$ (red dotted line) is highlighted by red (hypermethylated) and blue (hypomethylated) colors. The top 30 CpGs were labeled by neighboring genes. B) Location of top CpGs in each tissue relative to the closest transcriptional start site. Top CpGs were selected at $p < 10^{-4}$ and further filtering based on z score of association with chronological age for up to 500 in a positive or negative direction. The number of selected CpGs: adipose, 29; cerebellum, 142; cerebral cortex, 1000; heart, 35; liver, 823; muscle, 197; and tissue meta-analysis, 1000. The grey color in the last panel represents the location of 36523 mammalian BeadChip array probes mapped to Panu_3.0.100 genome. C) Upset plot representing the overlap of aging-associated CpGs based on meta-analysis or individual tissues. Neighboring genes of the overlapping CpGs were labeled in the figure. D) Transcriptional motif enrichment for the top CpGs in the promoter and 5`UTR of the neighboring genes.

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.29.402891; this version posted November 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

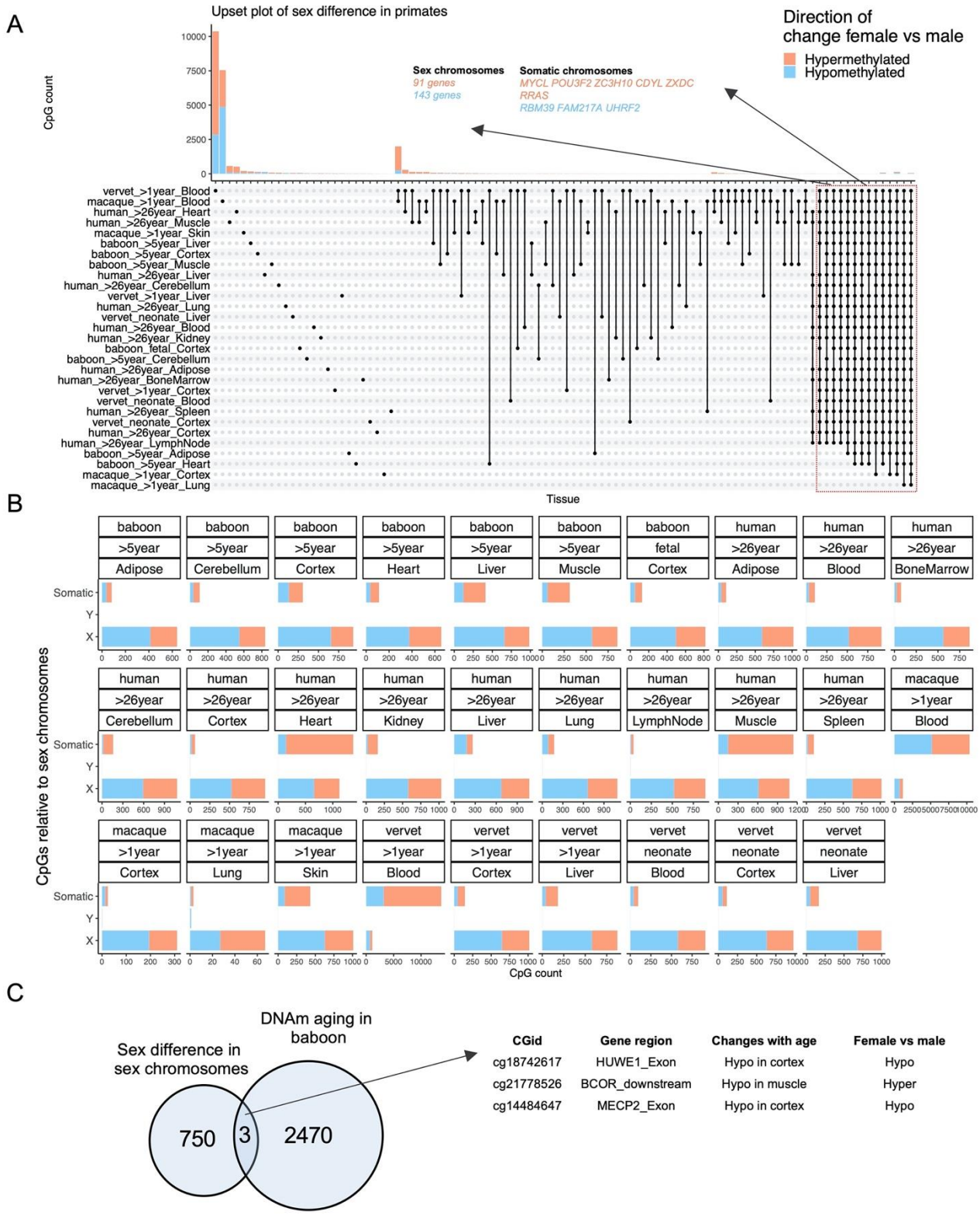


Figure 4. Sex effects on DNA methylation in primates. A) The upset plot of the sex-mediated differential methylated CpGs in 13 tissues from four mammalian species (human, baboon, vervet monkey, and macaque) at different age ranges. The analysis included 29 multivariate regression models (stratified by species, tissues, and age categories), including chronological age as the covariate. The y-axis reports the number of CpGs that are significant at a 5% FDR rate. Neonates in vervet monkeys were aged <1 year. B) Distribution of differentially methylated CpGs relative to sex chromosomes. Gene annotations are based on the alignment of

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.29.402891; this version posted November 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

the probes in mammalian array to the target species: human, HG38; vervet monkey, Chlorocebus_sabaeus.ChlSab1.1.100; macaque, Macaca_mulatta.Mmul_10.100; and baboon, Papio_anubis.Panu_3.0.100. C) The overlap of conserved sex-dependent CpGs and DNAm aging in the baboons. Manhattan plots of the sex differences in specific baboon tissues can be found in Supplementary Figure S11.

Figure 5. Sex related methylation changes in primates implicated by two array platforms.
The Manhattan plots only involves EPIC array probes that map to the same human gene regions as the corresponding probe on the mammalian array. The y-axis reports sex effects (Z statistics) based on the human EPIC array. The x-axis reports genomic locations (gene regions) that were significant (5% FDR) in at least one of the primates blood samples (human, macaque, vervet monkey). A) The overlap of primates EWAS of sex with the meta-analysis of Jackson Heart (JHS) and Framingham Heart (FHS) Studies. The meta-analysis z score is calculated by Stouffer's method from the EWAS of sex (adjusted for age) in JHS and FHS. The red lines are z scores that corresponds to $p<10^{-100}$ in EPIC array meta-analysis. The top two gene regions from each chromosome were labeled. B) The overlap of primates EWAS of sex (adjusted for age) prefrontal cortex samples from ROSMAP. The probes were also further filtered for the gene regions that were significant (5%

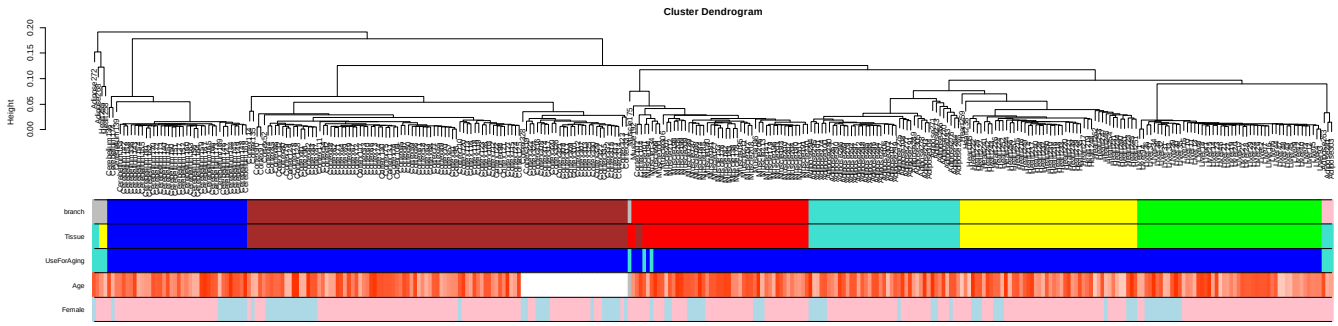
bioRxiv preprint doi: https://doi.org/10.1101/2020.11.29.402891; this version posted November 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

FDR) in at least one of the primates' cerebral cortex samples (baboon, vervet monkey). The red lines are T scores that corresponds to $p<10^{-8}$ in EPIC array EWAS analysis. The human EPIC array covers most of the genes that are covered by the mammalian methylation array (Supplementary Figure 10). Results for all probes on the EPIC array are presented in Supplementary Figure S14.

Tissue	N	No. Female	Mean Age	Min. Age	Max. Age
Adipose	41	32	15.6	7.49	22.8
Cerebellum	38	29	15.8	8.04	22.8
Cortex	76	60	15.3	7.49	22.8
Fetal Cortex	29	16	0	0	0
Heart	48	38	14.5	5.98	22.8
Liver	50	40	14.6	5.98	22.8
Muscle	44	34	16.4	8.04	22.8

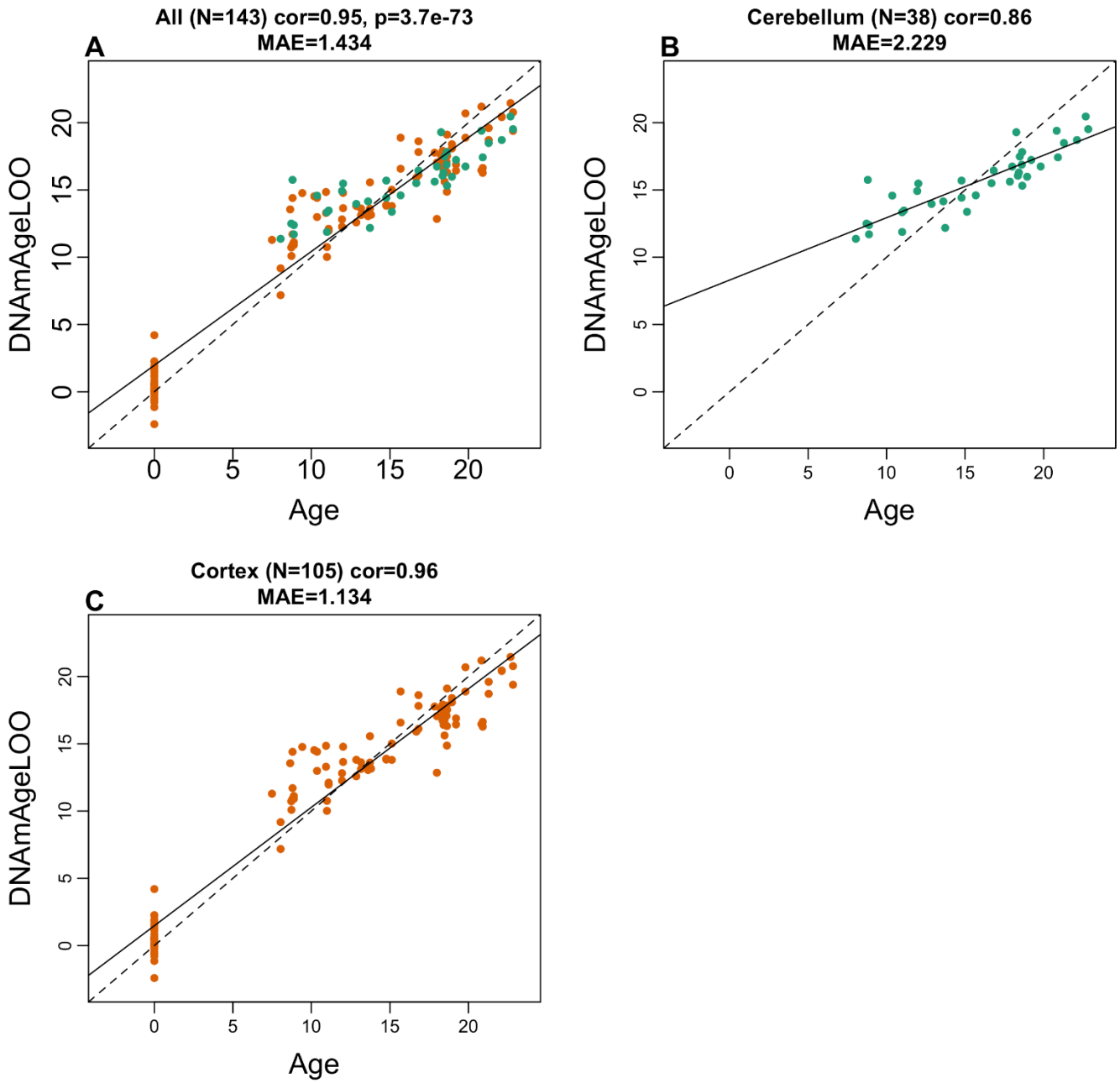
Table 1. Description of the DNA methylation data from baboons.
Tissue type. N=Total number of samples per tissue. Number of females. Age: mean, minimum and maximum. The fetal brain cortex samples were collected at gestational age 165 days.

SUPPLEMENTARY MATERIAL



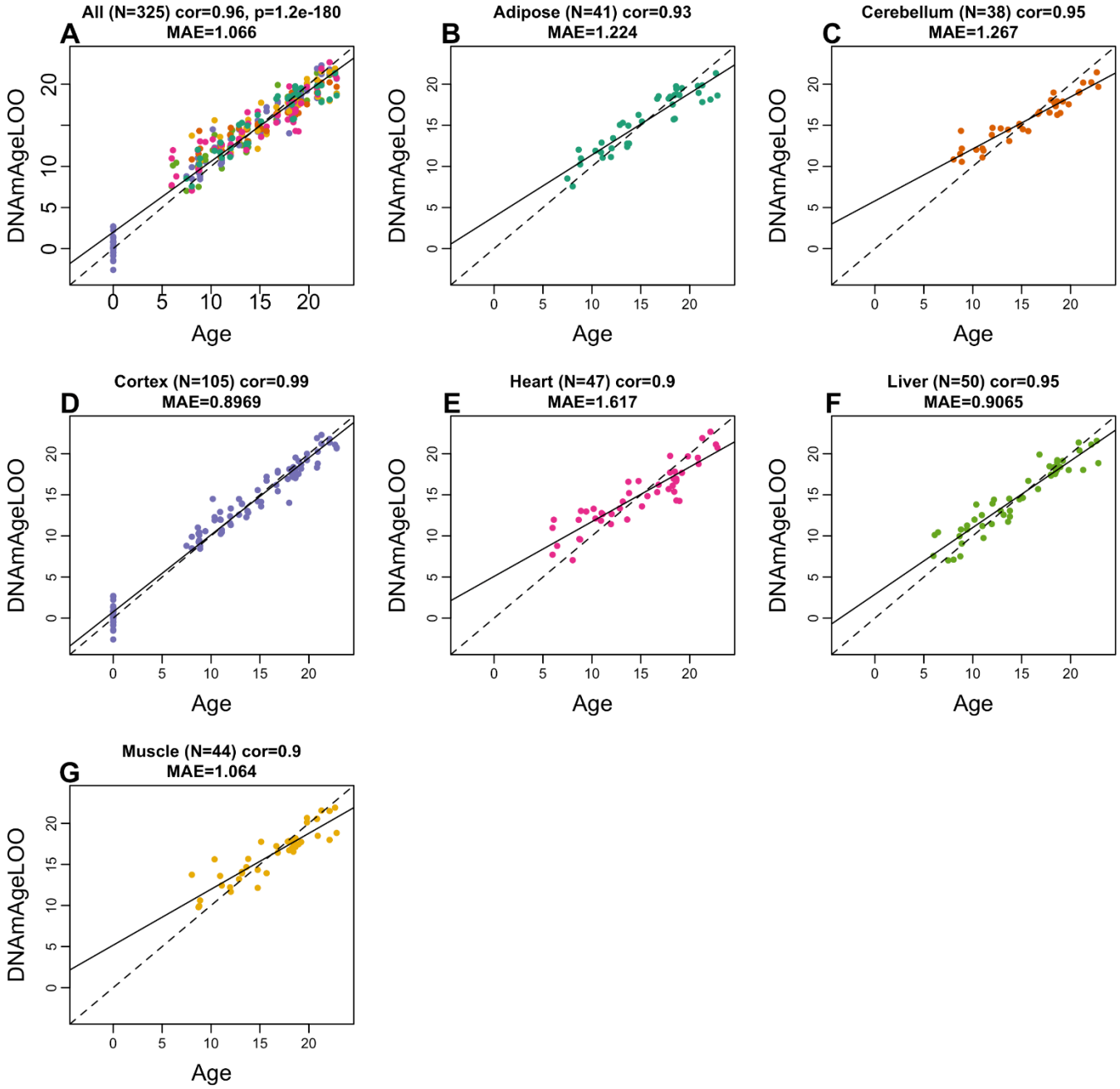
Supplementary Figure S1. Unsupervised hierarchical clustering of tissue samples. Average linkage hierarchical clustering based on the interarray correlation coefficient (Pearson correlation). A height cut-off of 0.07 led to branch colors that largely correspond to Tissue type (second panel): cerebral cortex=brown, muscle=red, heart=yellow, liver=green, adipose=turquoise, cerebellum=blue. A handful of outlying arrays were removed from the analysis (turquoise color in the third color band). Branches largely correspond to tissue type as one can see by comparing the first two color bands.

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.29.402891; this version posted November 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



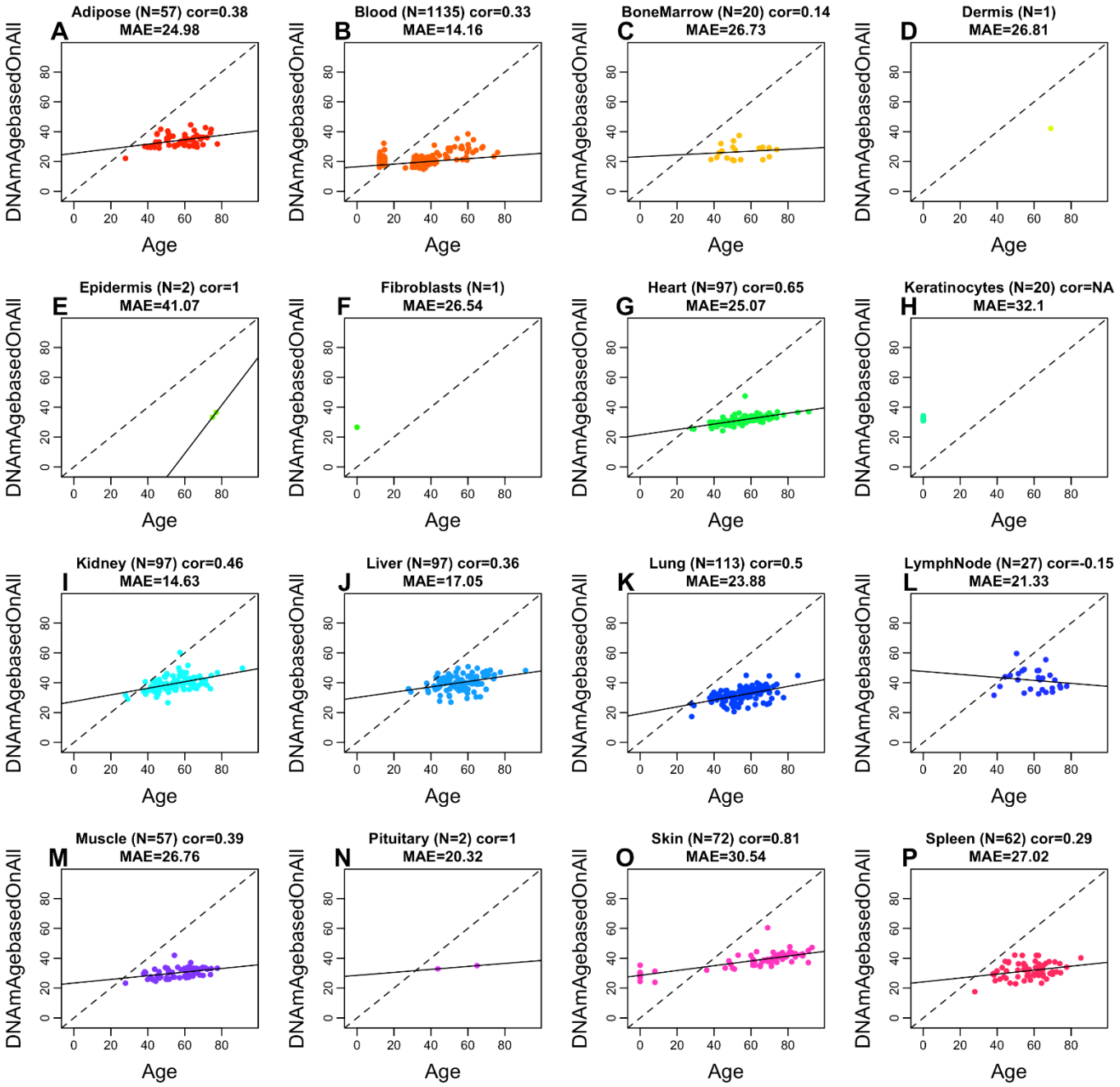
Supplementary Figure S2. Baboon brain clock Leave-one-sample-out estimate of age based on DNA methylation data (y-axis) versus chronological age (in units of years). A) The brain clock was developed using cerebellum, frontal cortex and temporal cortex samples. Results restricted to B) cerebellar samples and C) cortical samples Each title reports the sample size, Pearson correlation coefficient and median absolute deviation (median error).

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.29.402891; this version posted November 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



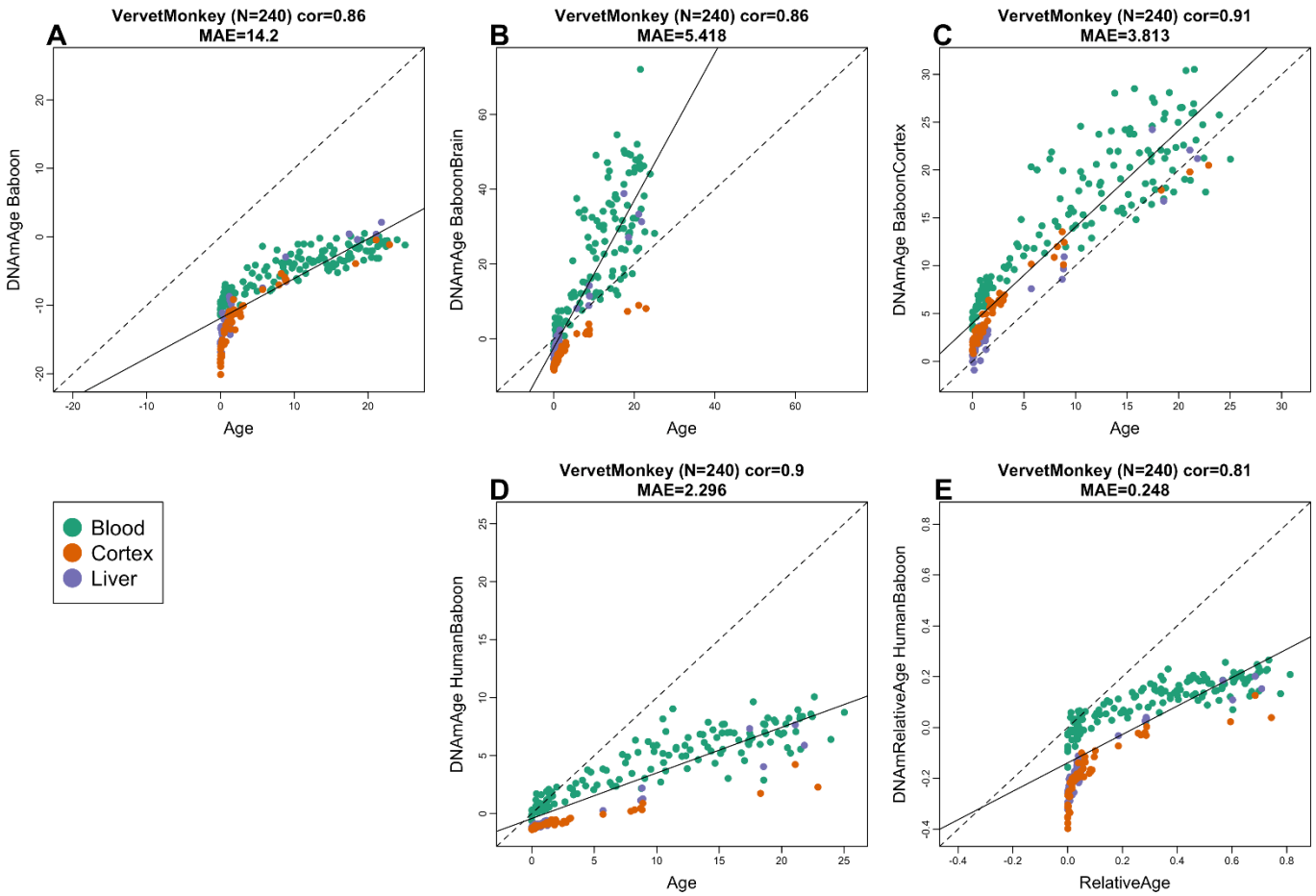
Supplementary Figure S3. Pan-tissue baboon epigenetic clock. Leave-one-sample-out estimate of age based on DNA methylation data (y-axis) versus chronological age (in units of years). Results for A) all tissues, B) adipose, C) brain cerebellum, D) cerebellar cortex, E) heart, F) liver, G) skeletal muscle. Each panel reports the sample size, Pearson correlation coefficient and median absolute deviation (median error).

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.29.402891; this version posted November 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



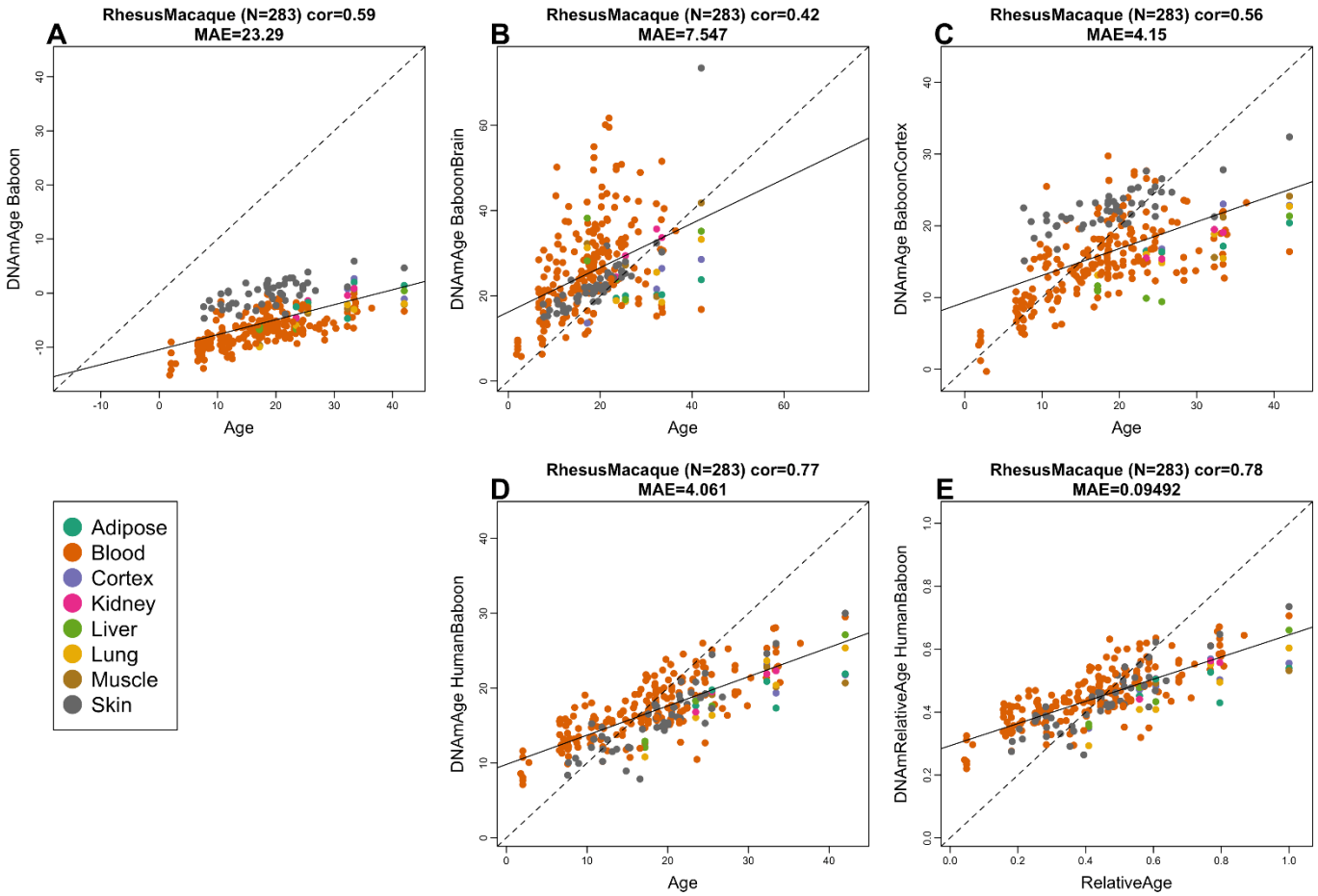
Supplementary Figure S4. Pan-tissue clock from baboons applied to human tissues. Baboon pan-tissue clock (trained on baboon tissues) applied to human A) adipose, B) blood, C) bone marrow, D) dermis, E) epidermis, F) fibroblasts G) heart, H) keratinocytes, I) kidney, J) liver, K) lung, L) lymph node, M) muscle, N) pituitary gland, O) skin, P) spleen. Estimate age based on the baboon pan tissue clock (y-axis) versus chronological age in humans (in units of years). Title: sample size, Pearson correlation coefficient and median absolute deviation (median error).

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.29.402891; this version posted November 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



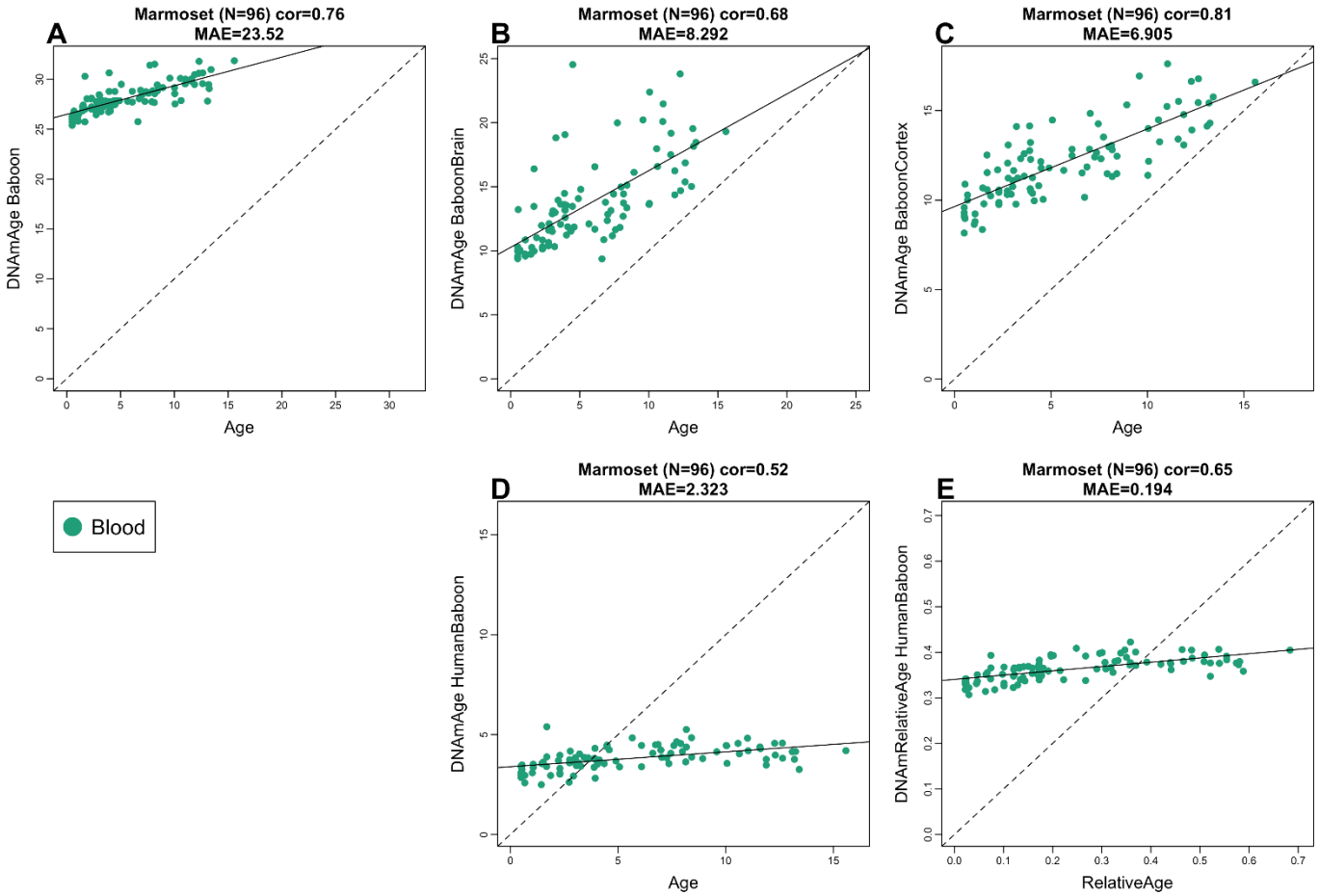
Supplementary Figure S5. Baboon clocks applied to tissues from vervet monkey (*Chlorocebus sabaeus*). Each dot corresponds to a tissue sample from vervet monkey. Chronological age of the vervet specimens (x-axis) versus DNAm age estimate based on the respective baboon clock. A) Baboon pan tissue clock, B) brain clock, C) cortex clock, D) human-baboon clock for chronological age, E) human-baboon clock for relative age. The number of samples is shown in parentheses; cor – Pearson's correlation, MAE – median absolute error.

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.29.402891; this version posted November 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



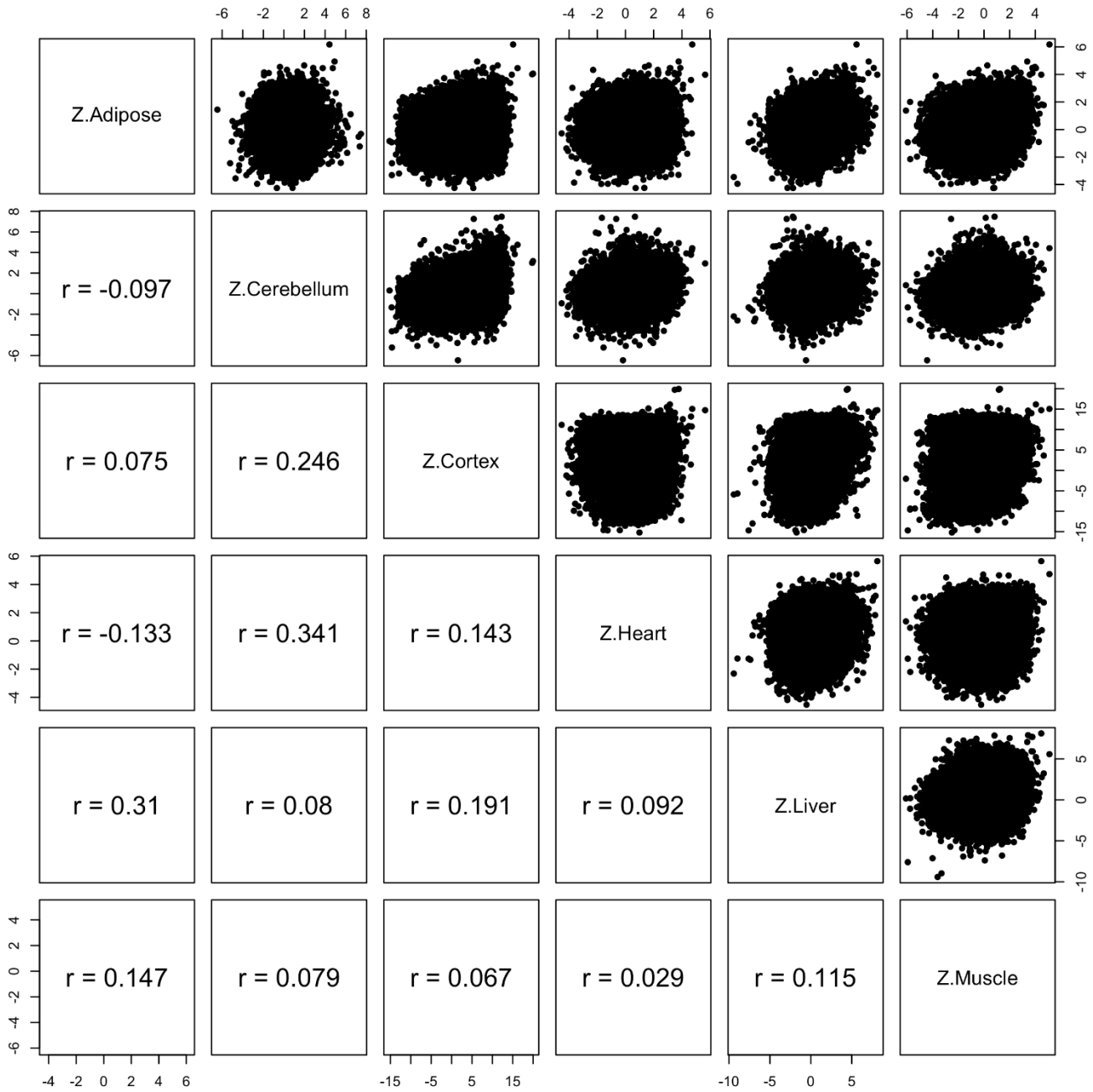
Supplementary Figure S6. Baboon clocks applied to tissues from rhesus macaques. Each dot corresponds to a tissue sample from rhesus macaques: Chronological age of the vervet specimens (x-axis) versus DNAm age estimate of the baboon A) pan tissue clock, B) brain clock, C) cortex clock, D) human-baboon clock for chronological age, E) human-baboon clock for relative age. The number of samples is shown in parentheses; cor – Pearson's correlation, MAE – median absolute error.

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.29.402891; this version posted November 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



Supplementary Figure S7. **Baboon clocks applied to tissues from marmosets**. Each dot corresponds to a blood sample from marmosets. Chronological age of the marmoset (x-axis) versus DNAm age estimate of the respective baboon clock. A) Baboon pan tissue clock, B) brain clock, C) cortex clock, D) human-baboon clock for chronological age, E) human-baboon clock for relative age. The number of samples is shown in parentheses; cor – Pearson's correlation, MAE – median absolute error.

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.29.402891; this version posted November 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

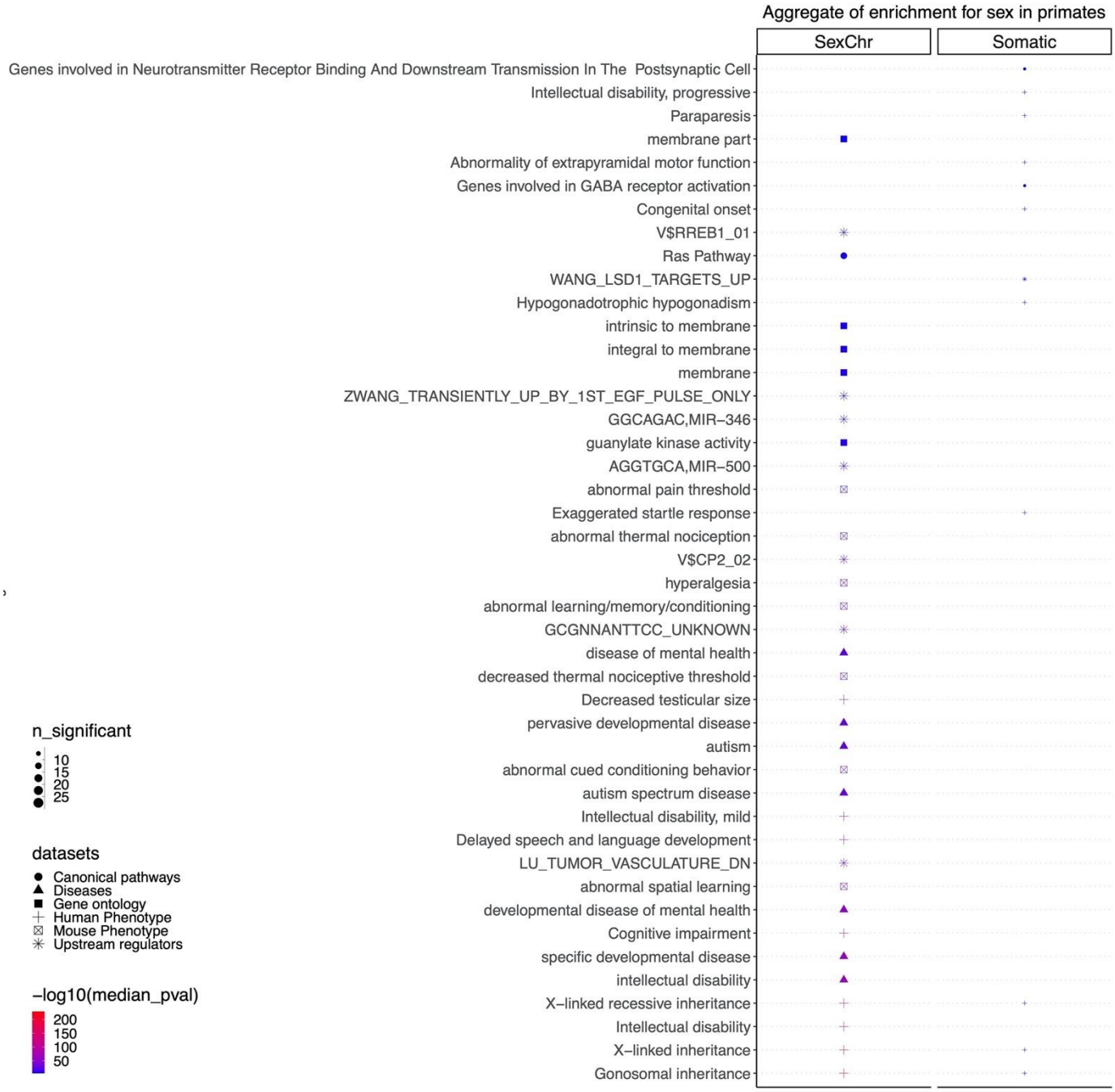


Supplementary Figure S8. Epigenome wide association study of correlation in different baboon tissues. Each dot corresponds to a CpG. Z statistics for a correlation test of age in adipose, cerebellum, brain cortex, heart, liver, muscle, temporal cortex.

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.29.402891; this version posted November 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

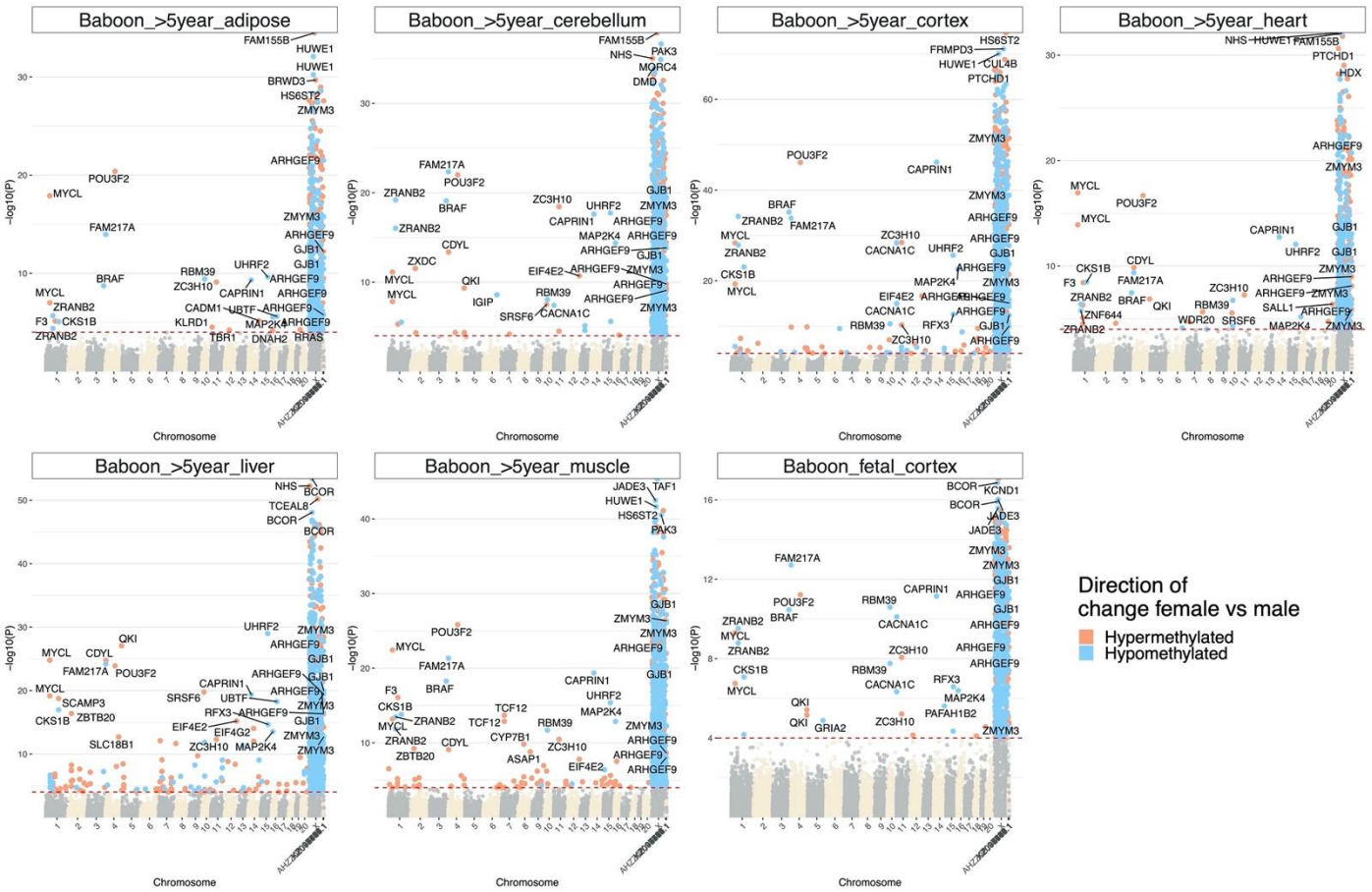
Supplementary Figure S9. Enrichment analysis of the top CpGs associated with DNAm age in the baboon. The analysis was done using the genomic region of enrichment annotation tool [65]. The gene level enrichment was done using GREAT analysis [65] and human Hg19 background. The background probes were limited to 25,677 probes that were mapped to the same gene in the olive baboon genome. The top three enriched datasets from each category (Canonical pathways, diseases, gene ontology, human and mouse phenotypes, and upstream regulators) were selected and further filtered for significance at $p < 10^{-6}$.

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.29.402891; this version posted November 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



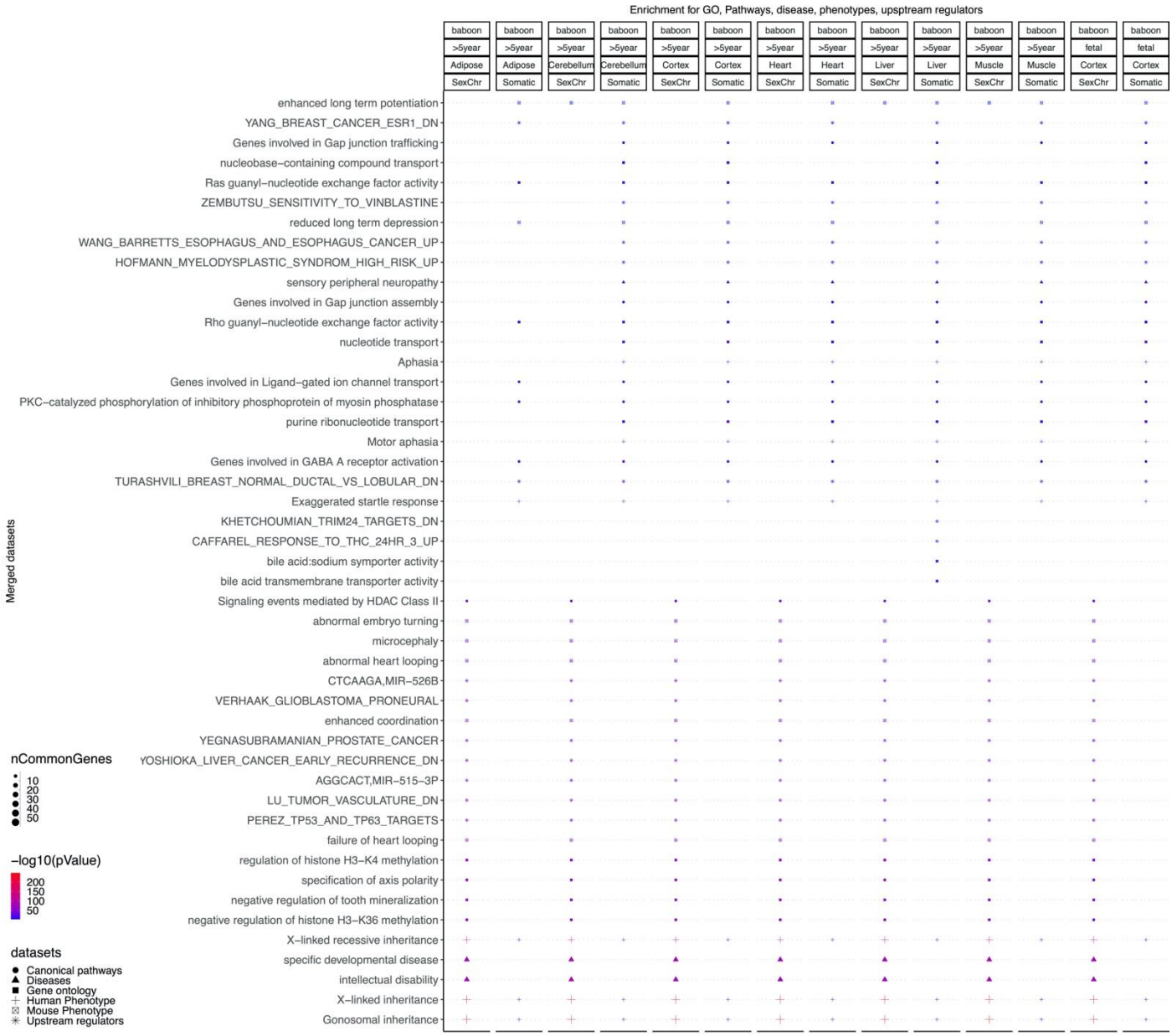
Supplementary Figure S10. Summary of enrichment analysis of sex-related CpGs in primate tissues by sex chromosomes. The figure summarizes the results of 27 EWAS of sex corresponding to 27 different strata comprised by primate species and tissue type. For each term (y-axis) we obtained 27 enrichment p values (hypergeometric test) corresponding to the 27 strata. These 27 p values were summarized by their median value. This unusual meta analysis approach leads to a *descriptive* measure of significance as opposed to an inferential measure. The size of the symbol (n_significant) corresponds to the number of times that a term was significant ($p<0.001$) across the 27 analyzed datasets. The top three enriched datasets from each category (Canonical pathways, diseases, gene ontology, human and mouse phenotypes, and upstream regulators) were selected and further filtered for significance at median pvalue $< 10^{-5}$. We excluded marmosets from the analysis for reasons mentioned in the text.

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.29.402891; this version posted November 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



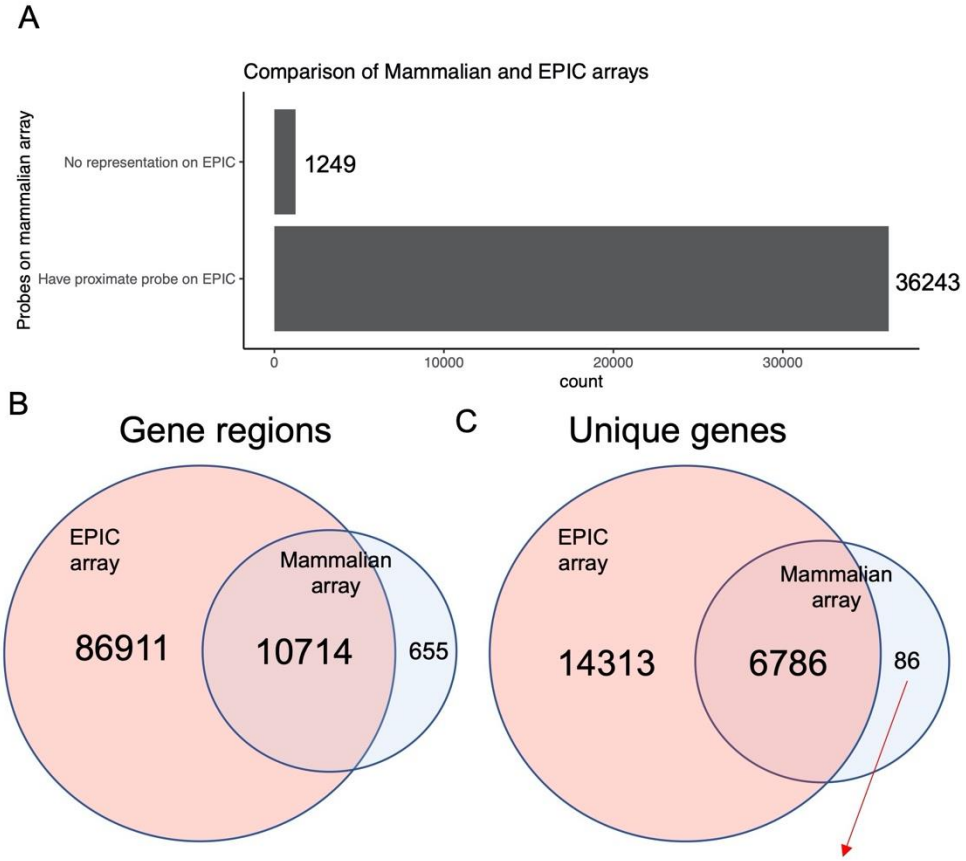
Supplementary Figure S11. Manhattan plots of sex difference in baboon samples. The coordinates are estimated based on the alignment of Mammalian array probes to Panu_3.0.100 genome assembly. The direction of change for female vs male with $p < 10^{-4}$ (red dotted line) is highlighted by red (hypermethylated) and blue (hypomethylated) colors. The top 5 X chromosome CpGs, and top 30 non-X chromosome CpGs were labeled by neighboring genes.

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.29.402891; this version posted November 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



Supplementary Figure S12. Enrichment analysis of the top CpGs associated with sex in baboons. The gene level enrichment was done using GREAT analysis and human Hg19 background. The background probes were limited to probes that were mapped to the same gene in the olive baboon genome. The top two enriched datasets from each category (Canonical pathways, diseases, gene ontology, human and mouse phenotypes, and upstream regulators) were selected and further filtered for significance at $p < 10^{-10}$.

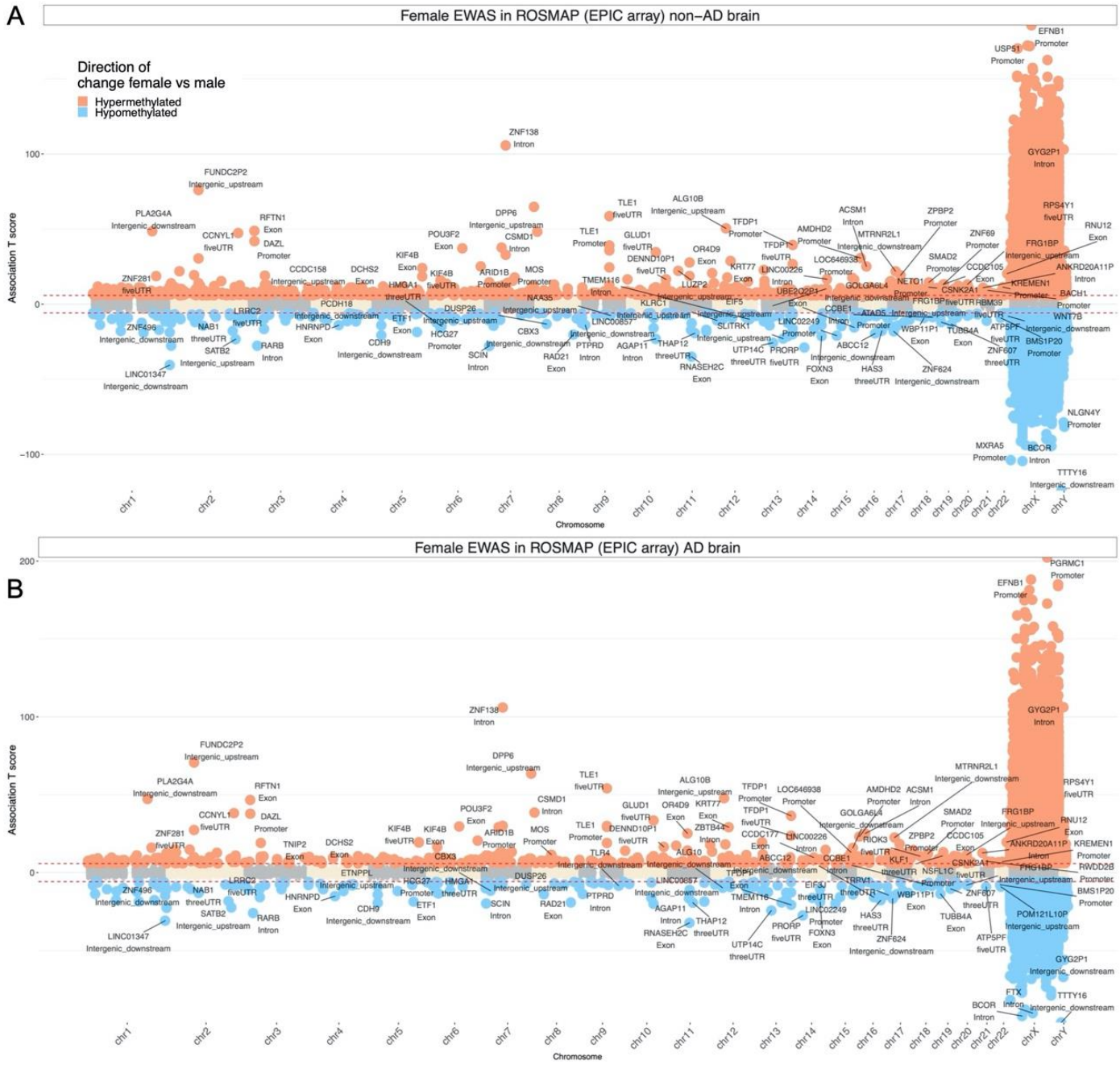
bioRxiv preprint doi: https://doi.org/10.1101/2020.11.29.402891; this version posted November 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



PELO ; ZBTB12 ; MAB21L2 ; MIR181A1 ; FMC1-LUC7L2 ; EWSAT1 ; RBM12 ; HOXC6 ; MIRLET7G ; CHIC2 ; DKFZP434H168 ; ZNF551 ; MMP24 ; PAX6-AS1 ; ASPA ; MIR30E ; HOXB5 ; MIR618 ; MIR3184 ; MAZ ; CORO1A ; SNORA16A ; LOC100130872 ; HOXA6 ; ALDOA ; SENCR ; HOXB4 ; ZBED6CL ; DUSP8 ; DLX6 ; MIR181A2HG ; ITGB1BP2 ; MIR218-1 ; MIR1306 ; PDCD1LG2 ; OR51B5 ; STARD10 ; NONO ; GABPB1-IT1 ; CDSN ; KCNQ1OT1 ; MIR330 ; MIR628 ; MIR22 ; MIR301A ; MIR26A1 ; TBC1D10C ; MRPL38 ; CNKSR3 ; PRR34-AS1 ; PITPNM2-AS1 ; MIR217 ; MIR5047 ; LOC100126784 ; DNAL4 ; MIR142 ; MIR136 ; MIR133A1 ; MIR99A ; SDCBP2-AS1 ; PRRT2 ; UBE2F ; POC1B-GALNT4 ; LRRTM2 ; MIR423 ; MIR3064 ; LRTM2 ; MIR127 ; GSTM4 ; APOBEC2 ; MIR383 ; RFX4 ; MIR3652 ; ADORA2A ; HOXA5 ; SMG1P1 ; CHERP ; MIR499A ; MIR548I4 ; SNORA50A ; MIR3618 ; MIR10A ; MIR3666 ; MIR5004 ; KCTD13 ; ASF1A

Supplementary Figure S13. Mammalian array compared to the human Illumina EPIC array. The EPIC array covers most of the genomic regions represented on mammalian array. A) Most of the mammalian array probes are located on gene regions that have at least one probe represented on the EPIC array. B) Venn Diagram visualizing the overlap between gene regions covered by the EPIC array and the mammalian array. Most of the gene regions on the mammalian array are also covered in EPIC array. C) Venn Diagram visualizing the overlap between genes covered by the EPIC array and the mammalian array. There are 86 genes that are specifically presented on mammalian but not the EPIC array.

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.29.402891; this version posted November 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



Supplementary Figure S14. Sex differences in DNAm pattern in human postmortem prefrontal cortex samples. Results for individuals with (A) and without (B) Alzheimer's disease (AD) neuropathology. Sex differences were adjusted for age in the model. The red lines represent T scores at $p < 10^{-8}$.

References

1 Cedar, H. & Bergman, Y. Programming of DNA methylation patterns. *Annual review of biochemistry* **81**, 97-117, doi:10.1146/annurev-biochem-052610-091920 (2012).

2 Christensen, B. *et al.* Aging and Environmental Exposures Alter Tissue-Specific DNA Methylation Dependent upon CpG Island Context. *PLoS Genet* **5**, e1000602 (2009).

3 Bollati, V. *et al.* Decline in genomic DNA methylation through aging in a cohort of elderly subjects. *Mech Aging Dev* **130**, 234-239, doi:http://dx.doi.org/10.1016/j.mad.2008.12.003 (2009).

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.29.402891; this version posted November 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

4 Rakyan, V. K. *et al.* Human aging-associated DNA hypermethylation occurs preferentially at bivalent chromatin domains. *Genome research* **20**, 434-439, doi:10.1101/gr.103101.109 (2010).

5 Teschendorff, A. E. *et al.* Age-dependent DNA methylation of genes that are suppressed in stem cells is a hallmark of cancer. *Genome research* **20**, 440-446, doi:10.1101/gr.103606.109 (2010).

6 Numata, S. *et al.* DNA Methylation Signatures in Development and Aging of the Human Prefrontal Cortex. *Am J Hum Genet.* **90**, 260-272, doi:10.1016/j.ajhg.2011.12.020 (2012).

7 Alisch, R. S. *et al.* Age-associated DNA methylation in pediatric populations. *Genome Res.* **22**, 623-632, doi:10.1101/gr.125187.111 (2012).

8 Johansson, A., Enroth, S. & Gyllensten, U. Continuous Aging of the Human DNA Methylome Throughout the Human Lifespan. *PLoS One* **8**, e67378, doi:doi: 10.1371/journal.pone.0067378 (2013).

9 Day, K. *et al.* Differential DNA methylation with age displays both common and dynamic features across human tissues that are influenced by CpG landscape. *Genome Biol.* **14**, R102 (2013).

10 Garagnani, P. *et al.* Methylation of ELOVL2 gene as a new epigenetic marker of age. *Aging Cell* **11**, 1132-1134, doi:10.1111/acel.12005 (2012).

11 Lin, Q. *et al.* DNA methylation levels at individual age-associated CpG sites can be indicative for life expectancy. *Aging (Albany NY)* **8**, 394-401 (2016).

12 Przybilla, J., Galle, J. & Rohlf, T. Is adult stem cell aging driven by conflicting modes of chromatin remodeling? *Bioessays* **34**, 841-848, doi:10.1002/bies.201100190 (2012).

13 Przybilla, J., Rohlf, T., Loeffler, M. & Galle, J. Understanding epigenetic changes in aging stem cells--a computational model approach. *Aging Cell* **13**, 320-328, doi:10.1111/acel.12177 (2014).

14 Beerman, I. *et al.* Proliferation-dependent alterations of the DNA methylation landscape underlie hematopoietic stem cell aging. *Cell Stem Cell* **12**, 413-425, doi:10.1016/j.stem.2013.01.017 (2013).

15 Beerman, I. & Rossi, D. J. Epigenetic regulation of hematopoietic stem cell aging. *Exp Cell Res* **329**, 192-199, doi:10.1016/j.yexcr.2014.09.013 (2014).

16 Gibbs, W. Biomarkers and aging: The clock-watcher. *Nature* **508**, 168–170, doi:doi:10.1038/508168a (2014).

17 Jung, M. & Pfeifer, G. P. Aging and DNA methylation. *BMC biology* **13**, 1-8, doi:10.1186/s12915-015-0118-4 (2015).

18 Benayoun, B. A., Pollina, E. A. & Brunet, A. Epigenetic regulation of aging: linking environmental inputs to genomic stability. *Nat Rev Mol Cell Biol* **16**, 593-610, doi:10.1038/nrm4048 (2015).

19 Horvath, S. & Raj, K. DNA methylation-based biomarkers and the epigenetic clock theory of aging. *Nat Rev Genet*, doi:10.1038/s41576-018-0004-3 (2018).

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

21 Marioni, R. *et al.* DNA methylation age of blood predicts all-cause mortality in later life. *Genome Biol.* **16**, 25 (2015).

22 Christiansen, L. *et al.* DNA methylation age is associated with mortality in a longitudinal Danish twin study. *Aging Cell* **15**, 149-154, doi:10.1111/acel.12421 (2016).

23 Perna, L. *et al.* Epigenetic age acceleration predicts cancer, cardiovascular, and all-cause mortality in a German case cohort. *Clin Epigenetics* **8**, 64, doi:10.1186/s13148-016-0228-z (2016).

24 Chen, B. H. *et al.* DNA methylation-based measures of biological age: meta-analysis predicting time to death. *Aging (Albany NY)* **8**, 1844-1865, doi:10.18632/aging.101020 (2016).

25 Horvath, S. *et al.* Decreased epigenetic age of PBMCs from Italian semi-supercentenarians and their offspring. *Aging (Albany NY)* **7**, 1159-1170, doi:10.18632/aging.100861 (2015).

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

27 Cox, L. A. *et al.* Baboons as a model to study genetics and epigenetics of human disease. *ILAR journal* **54**, 106-121, doi:10.1093/ilar/ilt038 (2013).

28 Bronikowski, A. M. *et al.* The aging baboon: Comparative demography in a non-human primate. *Proceedings of the National Academy of Sciences* **99**, 9591, doi:10.1073/pnas.142675599 (2002).

29 Field, A. E. *et al.* DNA Methylation Clocks in Aging: Categories, Causes, and Consequences. *Mol Cell* **71**, 882-895, doi:10.1016/j.molcel.2018.08.008 (2018).

30 Jasinska, A. J. *et al.* Epigenetic clock and methylation studies in vervet monkeys. *bioRxiv*, 2020.2009.2009.289801, doi:10.1101/2020.09.09.289801 (2020).

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.29.402891; this version posted November 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

31 Horvath, S. *et al.* Epigenetic clock and methylation studies in the rhesus macaque. *bioRxiv*, 2020.2009.2021.307108, doi:10.1101/2020.09.21.307108 (2020).

32 Horvath, S. *et al.* DNA methylation age analysis of rapamycin in common marmosets. *bioRxiv*, 2020.2011.2021.392779, doi:10.1101/2020.11.21.392779 (2020).

33 Zhu, T., Zheng, S. C., Paul, D. S., Horvath, S. & Teschendorff, A. E. Cell and tissue type independent age-associated DNA methylation changes are not rare but common. *Aging (Albany NY)* **10**, 3541 (2018).

34 Small, K. S. *et al.* Regulatory variants at KLF14 influence type 2 diabetes risk via a female-specific effect on adipocyte size and body composition. *Nature Genet.* **50**, 572-580, doi:10.1038/s41588-018-0088-x (2018).

35 Kumar, S., Stecher, G., Suleski, M. & Hedges, S. B. TimeTree: A Resource for Timelines, Timetrees, and Divergence Times. *Molecular Biology and Evolution* **34**, 1812-1819, doi:10.1093/molbev/msx116 (2017).

36 Teschendorff, A. E., West, J. & Beck, S. Age-associated epigenetic drift: implications, and a case of epigenetic thrift? *Human Molecular Genetics* **22**, R7-R15, doi:10.1093/hmg/ddt375 (2013).

37 Guo, Y. *et al.* Perhexiline activates KLF14 and reduces atherosclerosis by modulating ApoA-I production. *J Clin Invest* **125**, 3819-3830, doi:10.1172/jci79048 (2015).

38 Fan, G. *et al.* Loss of KLF14 triggers centrosome amplification and tumorigenesis. *Nat Commun* **6**, 8450, doi:10.1038/ncomms9450 (2015).

39 Yang, Q. & Civelek, M. Transcription Factor KLF14 and Metabolic Syndrome. *Front Cardiovasc Med* **7**, 91, doi:10.3389/fcvm.2020.00091 (2020).

40 Iwaya, C. *et al.* DNA methylation of the Klf14 gene region in whole blood cells provides prediction for the chronic inflammation in the adipose tissue. *Biochemical and biophysical research communications* **497**, 908-915, doi:10.1016/j.bbrc.2017.12.104 (2018).

41 Jung, S. E. *et al.* DNA methylation of the ELOVL2, FHL2, KLF14, C1orf132/MIR29B2C, and TRIM59 genes for age prediction from blood, saliva, and buccal swab samples. *Forensic Sci Int Genet* **38**, 1-8, doi:10.1016/j.fsigen.2018.09.010 (2019).

42 Kananen, L. *et al.* Aging-associated DNA methylation changes in middle-aged individuals: the Young Finns study. *BMC genomics* **17**, 103, doi:10.1186/s12864-016-2421-z (2016).

43 Gentile, C. & Kmita, M. Polycomb Repressive Complexes in Hox Gene Regulation: Silencing and Beyond: The Functional Dynamics of Polycomb Repressive Complexes in Hox Gene Regulation. *Bioessays* **42**, e1900249, doi:10.1002/bies.201900249 (2020).

44 Kuckenberg, P., Kubaczka, C. & Schorle, H. The role of transcription factor Tcfap2c/TFAP2C in trophectoderm development. *Reprod Biomed Online* **25**, 12-20, doi:10.1016/j.rbmo.2012.02.015 (2012).

45 Auman, H. J. *et al.* Transcription factor AP-2gamma is essential in the extra-embryonic lineages for early postimplantation development. *Development* **129**, 2733-2747 (2002).

46 Horvath, S. *et al.* An epigenetic clock analysis of race/ethnicity, sex, and coronary heart disease. *Genome Biol* **17**, 171, doi:10.1186/s13059-016-1030-0 (2016).

47 Zhang, Y. *et al.* Corepressor protein CDYL functions as a molecular bridge between polycomb repressor complex 2 and repressive chromatin mark trimethylated histone lysine 27. *The Journal of biological chemistry* **286**, 42414-42425, doi:10.1074/jbc.M111.271064 (2011).

48 Liu, Y. *et al.* Chromodomain protein CDYL is required for transmission/restoration of repressive histone marks. *J Mol Cell Biol* **9**, 178-194, doi:10.1093/jmcb/mjx013 (2017).

49 Schlabritz-Loutsevitch, N. E. *et al.* Metabolic adjustments to moderate maternal nutrient restriction. *British journal of nutrition* **98**, 276-284 (2007).

50 Kavitha, J. V. *et al.* Down-regulation of placental mTOR, insulin/IGF-I signaling, and nutrient transporters in response to maternal nutrient restriction in the baboon. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* **28**, 1294-1305, doi:10.1096/fj.13-242271 (2014).

51 Li, C. *et al.* Effects of maternal global nutrient restriction on fetal baboon hepatic insulin-like growth factor system genes and gene products. *Endocrinology* **150**, 4634-4642, doi:10.1210/en.2008-1648 (2009).

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.29.402891; this version posted November 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

52 Schlabritz-Loutsevitch, N. E. *et al.* Development of a system for individual feeding of baboons maintained in an outdoor group social environment. *Journal of Medical Primatology* **33**, 117-126, doi:10.1111/j.1600-0684.2004.00067.x (2004).
53 Morgello, S. *et al.* The National NeuroAIDS Tissue Consortium: a new paradigm in brain banking with an emphasis on infectious disease. *Neuropathol Appl Neurobiol* **27**, 326-335. (2001).
54 Horvath, S. *et al.* Perinatally acquired HIV infection accelerates epigenetic aging in South African adolescents. *AIDS (London, England)* **32**, 1465-1474, doi:10.1097/QAD.0000000000001854 (2018).
55 Kabacik, S., Horvath, S., Cohen, H. & Raj, K. Epigenetic aging is distinct from senescence-mediated aging and is not prevented by telomerase expression. *Aging (Albany NY)* **10**, 2800-2815, doi:10.18632/aging.101588 (2018).
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-e123, doi:10.1093/nar/gky691 (2018).
57 Friedman, J., Hastie, T. & Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software* **33**, 1-22 (2010).
58 de Magalhaes, J. P., Costa, J. & Church, G. M. An analysis of the relationship between metabolism, developmental schedules, and longevity using phylogenetic independent contrasts. *J Gerontol A Biol Sci Med Sci* **62**, 149-160 (2007).
59 Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).
60 Bailey, T. L. *et al.* MEME Suite: tools for motif discovery and searching. *Nucleic Acids Research* **37**, W202-W208, doi:10.1093/nar/gkp335 (2009).
61 Dawber, T. R., Meadors, G. F. & Moore, F. E., Jr. Epidemiological approaches to heart disease: the Framingham Study. *Am J Public Health Nations Health* **41**, 279-281 (1951).
62 Bennett, D. A., Schneider, J. A., Arvanitakis, Z. & Wilson, R. S. Overview and findings from the religious orders study. *Curr Alzheimer Res* **9**, 628-645 (2012).
63 Bennett, D. A. *et al.* Overview and findings from the rush Memory and Aging Project. *Curr Alzheimer Res* **9**, 646-663 (2012).
64 Levine, A. J. *et al.* Accelerated epigenetic aging in brain is associated with pre-mortem HIV-associated neurocognitive disorders. *J Neurovirol*, doi:10.1007/s13365-015-0406-3 (2015).
65 McLean, C. Y. *et al.* GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol* **28**, doi:10.1038/nbt.1630 (2010).