The impact of sex, age, and genetic ancestry on DNA methylation across tissues

Winona Oliveros[1,2], Jose Miguel Ramirez[1,2], Marta Melé[1,3*]

Affiliations

[1]Department of Life Sciences, Barcelona Supercomputing Center (BCN-CNS), Barcelona, Catalonia, 08034, Spain

[2]Universitat de Barcelona, Barcelona, Catalonia, 08034, Spain

[3] Lead contact

*Correspondence: marta.mele@bsc.es

ABSTRACT (226 words, 1629 characters)

Background: Understanding the consequences of individual DNA methylation variation is crucial for the study of human biology and disease. However, the collective impact of demographic traits on DNA methylation and subsequent gene expression across human tissues remains unexplored. Here, we quantify the contributions of sex, age, genetic ancestry, and BMI on DNA methylation variation across 9 human tissues and 424 individuals from the Genotype-Tissue Expression project.

Results: We observe that inter-individual DNA methylation variation is higher than inter-individual gene expression variation. We show that genetic ancestry and age have a greater impact on DNA methylation than does sex, with aging effects being more widespread but less pronounced. On average, less than 10% of the gene expression variation in sex, age, and ancestry is mediated by DNA methylation differences, with ancestry showing the largest percentage of mediation. We show that DNA methylation differences between populations accumulate at regions with outlier DNA methylation profiles (high/lowly methylated) and are largely under genetic control. The female genome is predominantly hypermethylated across tissues, likely driven by the polycomb repressive complex. Ultimately, we demonstrate how aging increases DNA methylation levels at targets of the polycomb repressive complex in most tissues except the gonads.

Conclusions: Overall, our multi-individual and multi-tissue analysis provides a comprehensive characterization of the main drivers of human DNA methylation variation under healthy conditions, laying the foundations for clinical studies considering demographic perspectives.

Keywords: DNA methylation, age, sex, genetic ancestry, GTEx, polycomb

Background

The deposition and maintenance of DNA methylation is essential for normal mammalian development. DNA methylation occurs when a methyl group (-CH3) is transferred into, almost exclusively, the C5 position of the cytosine in CpG dinucleotides[1–4]. DNA methylation is fundamental for the regulation of X-inactivation[4], genomic imprinting[5,6], and the repression of repetitive elements[6]. Most (60–80%) of the 28 million CpGs in the human genome are methylated, which has mainly been associated with gene silencing[5]. The remaining 20–40% of unmethylated CpGs are mostly located in CpG islands, which are genomic regions with very high CpG content that are resistant to DNA methylation. However, the role of DNA methylation on transcriptional regulation depends on the genomic context[7]. DNA methylation at promoters, which are usually located in genomic regions containing CpG islands, is associated with long-term gene silencing[5,8], whereas DNA methylation at the gene body can, in certain contexts, stimulate transcription elongation[5,9]. Importantly, aberrations in DNA methylation are associated with many diseases[6], such as various cancer types[10–13], autoimmune and inflammatory diseases[14–18], and cardiovascular diseases[19–21]. Thus, DNA methylation plays a fundamental role in genome regulation.

DNA methylation can vary between individuals, influenced by both intrinsic (e.g., age, genetics, and sex), and environmental factors. Age-related changes in DNA methylation are among the most extensively studied, including many individual CpGs for which methylation status has been reported to correlate with age[22]. In particular, it has been shown that with age, DNA becomes more hypermethylated at bivalent chromatin regions and CpG islands[23,24], whereas it becomes more hypomethylated at open-sea regions (i.e., isolated CpGs)[25,26]. In fact, the accumulation of DNA methylation changes can accurately predict chronological age and mortality[27–29]. Sex-based differences in DNA methylation further contribute to the complexity of inter-individual variation. Methylation profiles exhibit pronounced differences between males and females across various tissues, including both sex chromosomes and autosomes[30–32]. For example, female genomes display higher levels of autosomal methylation than male genomes do in whole blood[33]. Finally, genetic variation significantly influences DNA methylation through mechanisms such as methylation quantitative trait loci (meQTLs), which link specific genetic variants to methylation levels at particular CpG sites. This genetic influence is evident in the differential methylation patterns observed across human populations in whole blood[34–37], which reflect both the ancestral genetic backgrounds and the impact of environmental exposures.

Despite the valuable insights provided by these studies, the study of DNA methylation variation has predominantly been confined to a limited number of tissues and single traits[24,34,38], with very limited efforts trying to disentangle its regulatory effect[39,40]. A recent study in whole blood dissected the collective contribution of different environmental and intrinsic factors to DNA methylation variation[41]. This study revealed that many factors, including age and sex, have a direct effect on DNA methylation, even when controlling for cell type composition effects. However, beyond whole blood, the comparative impact of demographic traits, such as age, genetic ancestry, sex, and body mass index (BMI), across tissues on DNA methylation and its effect on transcriptional regulation, remains largely unexplored.

In this study, we leveraged Genotype-Tissue Expression (GTEx) data to systematically investigate the relationships among sex, age, ancestry, and BMI, with DNA methylation variation across 9 human tissues encompassing 856 samples from a total of 424 individuals. We identified differentially methylated positions (DMPs) and differentially variably methylated positions (DVPs) across tissues with different demographic traits and their associations with gene expression. We identified common, trait- and tissue-specific patterns, which we further characterized via publicly available data, such as chromatin states, to determine their functional consequences. Overall, our multi-trait and multi-tissue analysis provides an exhaustive characterization of the global and tissue-specific impact of age, sex, ancestry, and BMI on DNA methylation variation.

RESULTS

Individually variable methylated CpGs are preferentially located at promoters

DNA methylation is known to vary both between individuals and between tissues[24,30,34]. However, the extent to which DNA methylation varies more across tissues than across individuals has only been estimated with few donors[38,42]. In addition, whether DNA methylation variability is comparable to gene expression variability remains unexplored. To first quantify how DNA methylation varies across individuals and tissues, we used data from 9 GTEx tissues from 424 individuals across 754,054 CpG sites from the Illumina EPIC methylation array (**Fig. 1A, Fig. S1A**)[43]. We used linear mixed models to quantify the contribution of individual and tissue to DNA methylation variation (see Methods). Variation in DNA methylation is substantially greater among tissues (mean ~53% of the total variance in

DNA methylation) than among individuals (mean ~3.7% of the total variance) (**Fig. 1B**), as previously reported[38,42]. We then calculated the proportion of gene expression variation explained by individual and tissue using the same tissues. Compared with methylation, gene expression varies less between tissues (mean ~45% of the total variance, Mann–Whitney–Wilcoxon test p value < 2.2e–16, **Fig. 1C**) and individuals (mean ~2.6% of the total variance, Mann–Whitney–Wilcoxon test p value = 5.815e–05), which is consistent with previous studies with smaller sample sizes[44]. Thus, tissue and individual factors explain a larger proportion of DNA methylation variability than gene expression variability does, probably reflecting the higher stability of DNA methylation.

We then evaluated whether DNA methylation variation impacts gene expression variation. We defined a set of highly variable CpGs and genes (>50% variance explained either by tissue or individual for DNA methylation or gene expression, respectively). Although there was no overlap between individually variable genes at the DNA methylation and gene expression levels, we found significantly higher gene expression variance between individuals for genes that we detected to be highly individually variable in DNA methylation (Mann–Whitney–Wilcoxon test p value < 0.05) (**Fig. S1B–C**). Accordingly, we found the same pattern for tissue variable genes (**Fig. S1D**). This finding indicates that DNA methylation, although significant, might have a modest contribution to explaining inter-individual and inter-tissue variation in gene expression.

Next, we wanted to explore whether the proportion of DNA methylation variation explained by tissue or individual variation changes depending on the genomic context of the CpGs. We divided the CpGs into enhancers, promoters, gene bodies, or intergenic regions according to the EPIC array annotation and found that enhancer-associated CpGs have higher tissue variability than the rest (OR= 3.81, FDR= 0, two-tailed Fisher's exact test), with promoter-associated CpGs being highly stable across tissues (**Fig. S1E-F)**. This observation is consistent with previous studies showing enrichment of cell type-specific methylation patterns at enhancer regions[38,45]. Although the influence of the genomic context on the proportion of variation explained by individual is modest (**Fig. S1E**), we detected a significantly larger impact at promoters (OR= 1.55, FDR= 2.3e-29, two-tailed Fisher's exact test; **Fig. 1D**). We then evaluated whether the contribution of DNA methylation variation to gene expression variation changes on the basis of the genomic context[5]. When stratifying the inter-individual gene expression variance between genes highly variable in DNA methylation for each genomic location, we found that only DNA methylation variation in promoters significantly impacts inter-individual gene expression variation (**Fig. 1E, Fig. S1G-H**).

Finally, we wanted to explore whether highly tissue or highly individually variable CpGs were associated with genes of particular functions. Highly tissue-variable CpGs are more often located near genes involved in developmental and metabolic processes (functional enrichment, FDR < 0.05) (**Table S1**), which likely reflects the developmental fine-tuning of tissue-specific gene expression and is consistent with previous observations where cell types cluster on the basis of developmental origin[38]. Highly individually variable CpGs (**4810** CpGs) are enriched in genes belonging to antigen processing and presentation pathways (functional enrichment, FDR < 0.05; **Table S1; Fig. 1F**). This enrichment is driven mostly by the presence of highly inter-individual variable CpGs located in the HLA genes located in the MHC locus (**Table S1**), which is one of the most polymorphic loci in vertebrates[46] and has been suggested to be transcriptionally regulated by DNA methylation[47]. Consistent with this observation, genes belonging to antigen processing and presentation pathways have higher inter-individual gene expression variation than other genes (Mann-Whitney-Wilcoxon Test p.value = 0.0012) (**Fig. 1G**).

Overall, we find higher tissue and individual variability of DNA methylation compared to gene expression. Interestingly, highly individually variable CpGs are located in the promoters of HLA genes, moderately impacting their gene expression variability and potentially reflecting differences in immune responses between individuals.

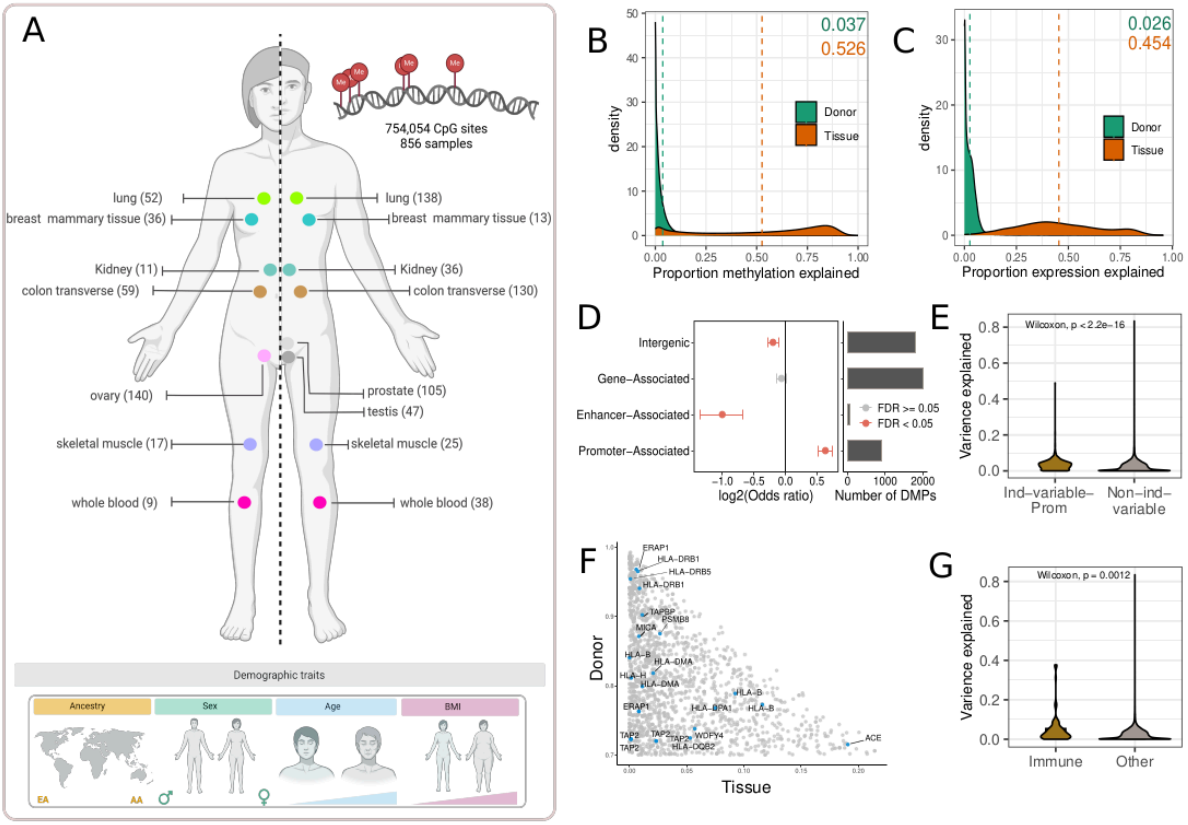


Figure 1. DNA methylation variation across tissues and individuals. A. Overview of the DNA methylation data, with the number of female samples (left) and the number of male samples (right) for each tissue. The number of analyzed CpGs and the total number of samples are shown. **B**. Proportion of DNA methylation variation explained by individual (green) and tissue (orange). The mean numbers are shown. **C.** Proportion of gene expression variation explained by individual (green) and tissue (orange). The mean numbers are shown. **D. Left.** Enrichment of each genomic location on individually variable CpGs (>50% variance explained either by individual for DNA methylation or gene expression). Dots are colored according to the significance of the enrichment (two-tailed Fisher's exact test). The confidence interval is shown. **Right.** Number of individually variable CpGs within each genomic location. **E.** Proportion of gene expression variation explained for genes that are individually-variable (brown) or not individually-variable (gray) at the DNA methylation level of their promoters. **F.** Individually variable CpGs. The X-axis represents the proportion of DNA methylation variation explained by tissue. The Y-axis represents the proportion of DNA methylation variation explained by individual. The blue dots represent CpGs associated with genes belonging to antigen processing and presentation pathways. **G.** Proportion of gene expression variation explained for genes belonging to antigen processing and presentation pathways (brown) versus other genes (gray).

Age and ancestry have a larger impact on DNA methylation than sex across tissues

Genetic effects, age, and sex simultaneously contribute to DNA methylation variation in whole blood, with genetic effects having the strongest impact[41]. To further study the simultaneous association between demographic traits and DNA methylation variation beyond whole blood, we quantified DNA methylation changes with genetic ancestry, sex, age, and BMI simultaneously across 9 human tissues. We identified differentially methylated

positions (DMPs) for each demographic trait while controlling for known sources of technical variation and other confounders, such as cell-type composition (see Methods). In most tissues, age has the largest number of DMPs, followed by ancestry and sex (**Fig. 2A**). However, tissues with smaller sample sizes have more ancestry-DMPs than age-DMPs (**Fig. 2A**). This is likely because ancestry-DMPs have larger effect sizes (effect sizes of ancestry-DMPs vs. age-DMPs, Mann–Whitney–Wilcoxon test p value < 2.2e–16 in all tissues) (**Fig. S2A**) and thus can be detected in tissues with low sample sizes. BMI shows no association with DNA methylation changes in the studied tissues, consistent with previous observations in whole blood[48,49]. The largest number of age-DMPs and ancestry-DMPs are located in the colon and lung respectively, which have the largest sample sizes. Conversely, skeletal muscle, despite not having many samples, has the largest number of sex-DMPs, consistent with previous reports of widespread sex DNA methylation and physiological differences[50–52]. Downsampling tissues to the same number of samples shows that skeletal muscle remains the tissue with the highest number of sex-DMPs and that ancestry has the largest number of DMPs across tissues (**Fig. S2B**).

Next, we tested the specific contribution of each demographic trait to DNA methylation variation per tissue (see Methods). We found that in every tissue, the demographic trait with the largest number of DMPs also explains most of the variation in DNA methylation variation (**Fig. 2A**). In general, however, the percentage of total variation explained by demographic traits per CpG is low (mean DNA methylation variation explained: 10.7% ancestry, 6% age, 3.8% sex, and 0.9% BMI). From the 23,120 DMPs for which a single trait explained more than 10% of their inter-individual DNA methylation variation in at least one tissue, age had the largest number of DMPs, followed by ancestry and sex (age 43%, ancestry 42%, sex 15%, **Fig. 2B**). Among those, some are known to be related to trait-associated phenotypes, such as cg01878807 and cg05171021, two CpGs located on the promoter of the *DHRS4* gene (ADME gene, highly related to the absorption, distribution, metabolism, and excretion of drugs), where ancestry explains more than 10% of the variability in DNA methylation. Higher DNA methylation of both CpGs in African-American individuals is correlated with lower gene expression[53] and could impact ethnic differences in treatment response and drug ADME[54].

Finally, we evaluated whether the contribution of each demographic trait to DNA methylation variation in each tissue could be correlated with the contribution to gene expression variation. To compare both data modalities, we studied the genes associated with DNA methylation changes in the same tissues. Compared to methylation, the percentage of total gene expression variation explained by each demographic trait per gene is lower (mean

gene expression variation explained: 1.3% ancestry, 2.7% age, 1.7% sex, and 1.9% BMI). Importantly, the contribution of each demographic trait to explaining gene expression variation mirrors previous results[55] but is different from that of DNA methylation, with an overall higher contribution from BMI and a lower contribution from ancestry (**Fig. S2C**).

Overall, age and ancestry have larger contributions to inter-individual variation in DNA methylation variation than sex, with aging effects being more abundant but weaker. Interestingly, we report a higher contribution of ancestry, age, and sex to DNA methylation variation than to gene expression.

Few age-, sex-, or ancestry-related methylation differences explain expression differences

Although we identified a modest contribution of DNA methylation variation to gene expression variation, we asked if previously observed differences in gene expression with age, sex, and ancestry[55] could be explained by changes in DNA methylation. To address this, we first selected the DMPs that had an associated differentially expressed gene (DEG) for the same trait and tissue combination in a previous study[55] (see Methods). The proportion of DMP-DEG pairs was very low (0.4%-13.7%), suggesting that most expression differences with respect to age, sex, or ancestry might not be explained by methylation differences. In fact, fewer than 10% of the gene expression differences correlated with DNA methylation changes in the DMP-DEG pairs (**Fig. 2C, Fig. S2D-E**). Ancestry was the demographic trait with the highest percentage of correlated genes (mean ~8%, maximum 20% of correlated DEGs). Contrarily, age showed overall few significant correlations (mean 4% correlated DEGs), consistent with previous results[24]. The exception to this trend is ovary, which yielded 432 correlated DMP-DEGs (corresponding to 239 DEGs, 10%) located near genes involved in blood circulation pathways (**Table S2**). As expected, most correlations involving CpGs located at promoters are negative compared with a lower proportion of negative correlations in enhancers and gene bodies (**Fig. 2D, Table S3**). We explored whether the low correlation between gene expression and DNA methylation could be explained by the limited statistical power of our differential DNA methylation analysis, which includes fewer samples than the previously published differential gene expression analysis[55]. To address this, we correlated the significant trait-DMPs (p value nominal < 0.05) with the corresponding DEGs. The proportion of gene expression differences that could be explained by DNA methylation substantially increased (up to 44% correlated DEGs in ovaries with genetic ancestry) (**Fig**

S2F), suggesting that other genes transcriptionally regulated by DNA methylation could be missing from existing correlation analyses.

Finally, we further explored the causal relationship between DNA methylation and gene expression by testing whether the effect of each demographic trait on gene expression could be mediated by changes in DNA methylation. To address this, we applied a regularized mediation analysis[56] to estimate the average causal mediation effect of DNA methylation on every DEG per trait and tissue (see Methods). On average, ~4% of the DEGs are mediated by DNA methylation (**Fig. 2E, Table S4**). In this mediation analysis, we included all the DMPs associated with a DEG in the same model as multiple mediators. Therefore, the similar proportion of mediated effects (mean ~4%) compared with our previous correlation analysis (mean ~10%) probably indicates that when there is an effect of DNA methylation on gene expression, a single CpG is sufficient to detect it (**Fig. S2G, Table S4**).

Together, these results indicate that DNA methylation changes in single CpGs do not normally modulate gene expression under healthy conditions.

DNA methylation variability is higher in individuals of African descent

DNA methylation has been shown to increase in variability with age across tissues[41,57]. We thus wanted to test whether sex and ancestry could also be associated with increased DNA methylation variability and compare their effects to age. We identified differential variable positions (DVPs) for age, sex, and ancestry while controlling for the same confounders as in our previous differential DNA methylation analysis (see Methods). In all tissues, ancestry has the largest number of DVPs (**Fig. 2F**). Interestingly, ~99% of the ancestry-DVPs show larger variability in individuals of African descent (binomial test; FDR < 0.05 in all tissues, **Fig. 2G**), consistent with previous findings that African individuals are more genetically diverse[58]. We identified a limited number of age- and sex-related DVPs, suggesting that ancestry-related DVPs might have larger effect sizes (**Fig. S2H**). We only observe the bias toward higher methylation variability with age in colon transverse, which has the largest number of detected age-related DVPs. Despite the limited number of sex-DVPs, they are biased toward greater variability in females in most tissues (binomial test; FDR < 0.05 in the kidney cortex and whole blood). Finally, the overlap between DMPs and DVPs is significant in all tissues only for ancestry (~25% of DMPs are also DVPs, two-tailed Fisher's exact test, FDR < 0.05), likely because both capture genetic effects and have larger effect sizes (**Fig. S2A**) compared to age and sex.

We then examined whether changes in methylation variability with sex, age, or ancestry could have an impact on gene expression. To address this, we first applied the same methodology to detect differentially variable genes (DVGs) using the gene expression data available for the same tissues from the GTEx dataset. On average, we detect ~290 age-DVGs, ~1800 sex-DVGs and ~1020 ancestry-DVGs across tissues (**Fig. S2I)**. In 99% of cases, ancestry-DVGs increase in expression variability in African individuals (binomial test; FDR < 0.05 in all tissues except testes, **Fig. 2H**). However, we find almost no overlap between trait-DVP-DVGs (mean overlap: 0 age-DVP-DVG, 1 sex-DVP-DVG, and 11 ancestry-DVP-DVG, **Fig. S2J**), indicating that variable CpGs have little to no impact on the regulation of gene expression.

Overall, we report widespread differences in DNA methylation between populations, which might reflect the higher genetic variability found in African populations and/or environmental effects.

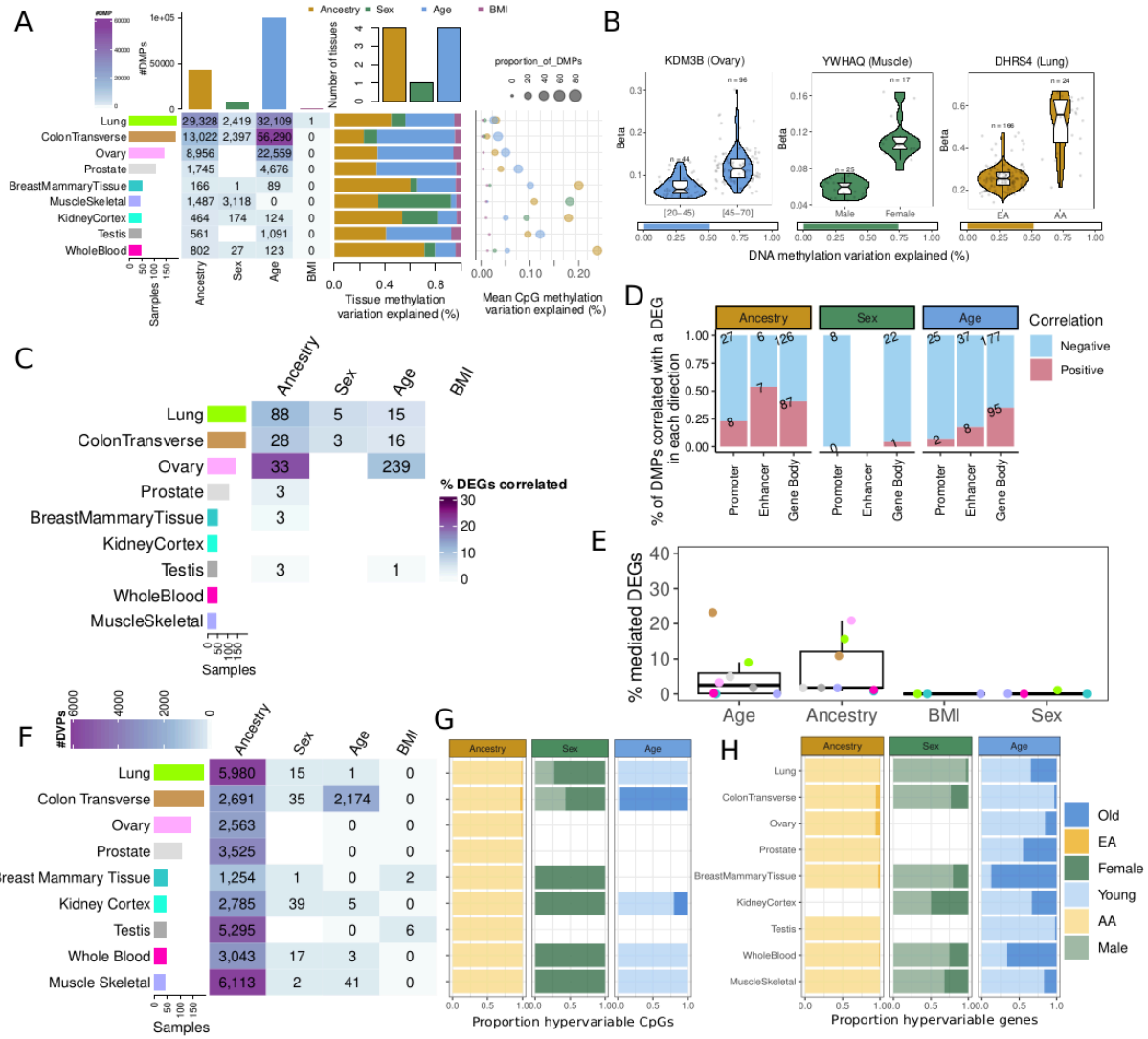


Figure 2. DNA methylation variation across tissues and demographic traits. A. (Left) Number of DMPs per tissue and demographic trait. Top bars show the number of CpGs differentially methylated by each demographic trait. Left bars represent the sample size. **(Center)** Proportion of total tissue DNA methylation variation explained by each demographic trait. Top bars show the number of tissues for which each demographic trait explains the largest proportion of variation. **(Right)** Mean DNA methylation variation explained by each demographic trait in each tissue. **B.** Examples of CpGs with a large proportion of DNA methylation variation explained by a demographic trait. The aging variable was categorized for visualization purposes. Top violin plots represent the beta values of each CpG across individuals, and the bottom bars represent the proportion of DNA methylation variation explained by each demographic trait. **C.** Number of DEGs per tissue and demographic trait correlated with at least 1 DMP. Cells are colored according to the proportion of DEGs correlated with DMPs out of the total number of DEGs with a CpG. Tissues are sorted by sample size. **D.** Proportion and number of DMPs correlated with DEGs per genomic location and demographic trait, colored according to the direction of correlation. **E.** Percentage of DEGs per tissue and demographic trait that are mediated by DNA methylation. Dots are colored according to tissue; see barplot in panel A. **F.** Number of DVPs per tissue and demographic trait. Left bars show sample size. Cells are colored according to the number of DVPs. **G.** Proportion of hyper-variably methylated CpGs per tissue and demographic trait. **H.** Proportion of hyper-variably expressed genes per tissue and demographic trait.

Africans and Europeans display divergent methylation at TSSs and quiescent regions

To better understand the relationship between genetic ancestry and epigenetics, we tested whether ancestry-DMPs are located in regions of the genome with specific chromatin states across tissues. We find that hypermethylated ancestry-DMPs in European-descent individuals are more often located in quiescent regions whereas ancestry-DMPs hypermethylated in African-descent individuals are consistently enriched in TSSs (two-tailed Fisher's exact test, FDR < 0.05) (**Fig. S3A**) across tissues. When selecting the shared ancestry-DMPs in the same direction (hypermethylated in 3 or more tissues), we observe the same enrichment (two-tailed Fisher's exact test, FDR < 0.05). (**Fig. 3A**). Interestingly, quiescent regions are, on average, the chromatin states with the highest methylation levels, whereas promoters have the lowest methylation levels (**Fig. S3B**). Hence, we evaluated whether our previous observation was driven by Africans being more diverse; and thus, regions with extreme DNA methylation profiles, such as promoters and quiescent regions, fluctuate in only one direction. To test this, we selected CpGs with extreme methylation values (i.e., CpGs with average methylation levels > 0.9 and < 0.1) and tested whether ancestry-DMPs significantly impact these extremely methylated and unmethylated CpGs. Indeed, we found that, on average, CpGs with high DNA methylation levels were hypomethylated in Africans compared with Europeans (**Fig 3B**), and, on average, CpGs with low DNA methylation levels were hypermethylated in Africans (**Fig S3C**). This observation is consistent with the higher genetic variability observed in Africans[58] and the larger DNA methylation variability we observed in African-descent individuals (**Fig. 2G**).

We then hypothesized that the enrichment in TSSs of ancestry-DMPs hypermethylated in Africans (**Fig. 3A, S3A**) might have an impact on the expression of the associated genes, as higher methylation at promoter regions is usually associated with lower gene expression[7]. To test this hypothesis, we compared the mean expression levels of genes with hypermethylated ancestry-DMPs in Africans. As expected, we found that the expression of these genes is lower in Africans than in Europeans (Wilcoxon paired test; p value<0.05).

Overall, our results show that widespread differences in DNA methylation profiles between African and European populations localize to regions of the genome with extreme methylation profiles, consistent with the higher genetic diversity and DNA methylation variability observed in African individuals. Finally, although weak, differences in DNA methylation at promoter regions between these populations seem to play a regulatory role.

Genetic effects explain ~70% of the differences in DNA methylation between human populations across tissues

Ancestry-associated DNA methylation changes can reflect the impact of both genetic and environmental factors[37,59]. A previous study in whole blood comparing methylation levels between a balanced number of European-descent and African-descent individuals estimated that 70% of ancestry-DMPs were driven by cis genetic effects[34]. Here, we wanted to quantify the proportion of ancestry-DMPs that are driven by genetic effects across tissues. To do this, we selected the list of *cis*-methylation quantitative trait loci (*cis*-mQTLs) (i.e., local associations between DNA methylation variation at CpGs and SNPs located within a 500-kb window) identified in the same samples in a previous study[43] and overlapped them with the ancestry-DMPs in the same tissue. On average, we find 25 to 50% of ancestry-DMPs in CpGs with associated *cis*-mQTLs (**Fig. 3C**). In the tissues with low sample sizes (whole blood, testis, kidney cortex, muscle, and breast), few of the ancestry-DMPs were previously associated with a *cis*-mQTL, suggesting that there might be genetic effects that have yet to be identified. However, we find that ancestry-related DMPs are significantly enriched in *cis*-mQTLs in 5 tissues (mCpGs) (two-tailed Fisher's exact test, FDR < 0.05) (**Fig. S3D**). To test whether the observed differences in DNA methylation between African and European populations are explained by genetic effects, we included the associated *cis*-mQTLs in our differential methylation models and tested whether those CpGs were still differentially methylated (see Methods). We find that, on average, ~70% of ancestry-DMPs that have an associated *cis*-mQTL can be explained by *cis*-genetic effects (here referred to as *cis*-driven DMPs) (**Fig. 3D**).

We then assessed whether the ancestry-related DVPs could also be driven by *cis*-genetic effects, as they are highly shared across tissues (**Fig. S3E**) and have a strong bias toward increased DNA methylation variability in African-descent individuals (**Fig. 2G**). Similar to DMPs, few DVPs are associated with *cis*-mQTLs (**Fig. 3E**). Nevertheless, when they are present, most (average of ~75%) are explained by genetic variants (here referred to as *cis*-driven DVPs) (**Fig. 3F**).

Among the four analyzed demographic traits, ancestry demonstrates the highest cross-tissue DMP sharing, with 25% (moderate sharing, 9419 CpGs) of its DMPs found in more than one tissue and 2% (high sharing, 903 CpGs) in 4 or more tissues (**Fig. 3G**). As expected, *cis*-driven DMPs were significantly more tissue-shared than DMPs without an

identified *cis*-mQTL and more significantly shared than non-*cis*-driven DMPs (Mann-Whitney-Wilcoxon Test p value < 0.05; **Fig. 3H**). Additionally, non-*cis*-driven DMPs were more significantly shared than DMPs without an identified mQTL (Mann–Whitney-Wilcoxon Test p value < 0.05, **Fig. 3H**), consistent with the observation that *cis*-mQTL effects are highly shared across tissues[43].

Cis-driven DMPs were enriched in enhancers (**Fig. S3F**) consistent with the fact that *cis*-mQTLs are usually located in non-genic regulatory regions[34,43]. As recent work has shown that chromatin states, specifically enhancers, might differ between African and European individuals[60], we explored whether ancestry-DMPs could be located in regions that had different chromatin states between these human populations. To address this, we used PBMC chromHMM states for an African-descent and a European-descent individual and assayed the preferential location of ancestry-DMPs in each individual (see Methods). Overall, 38% of whole-blood ancestry-related DMPs were located in differently annotated genomic regions between African and European populations (OR = 1.23, p value = 0.0044; **Fig. S3G**), with higher divergence at regulatory and repressed regions (**Fig. 3I**). Indeed, DNA hypermethylation in European-descent individuals was preferentially located in European-specific enhancers (OR=1.94, FDR=0.004; **Fig. S3H**), whereas DNA hypermethylation in African-descent individuals was preferentially located in African-specific bivalent enhancers (OR=2.23, FDR=0.04) (**Fig. S3I)**.

We next sought to characterize the functional pathways altered by ancestry-related DNA methylation changes. No pathway was significantly altered by overall ancestry-DMPs; however, enhancer-associated ancestry-DMPs were enriched in wound-healing and inflammation-related processes in the lung and colon (**Table S5**). Wound healing has been reported to play a role in inflammation[61–63] and inflammatory diseases[64]. Interestingly, 15 DM genes associated with wound healing are differentially expressed between populations during infection (8 DEGs in monocytes, 9 DEGs in CD8+ T cells, 8 DEGs in CD4+ T cells, 9 DEGs in NK cells and 11 DEGs in B cells), with 4 additional DM genes being differentially responsive during infection (*ABHD2*, *MAX*, *MYH9*, and *AHNAK* in monocytes and *AHNAK* in B cells)[65]. Additionally, one of the most tissue-shared ancestry-DMP, cg16999677 (**Fig. 3J**), is hypermethylated in European-descent individuals across 8 tissues, and is the CpG with its DNA methylation variation more strongly explained by genetic ancestry. This CpG is located in the gene body of the *ZDHHC11* gene, which plays a key role in the antiviral innate immune response[66], and has been recently identified as a positive modulator of *NF-κB* signaling, a critical transcription factor (TF) for inflammatory responses[67]. Additionally, this gene was reported to be significantly overexpressed in African-American individuals in four

tissues (two brain regions, the liver, and the thyroid)[55], consistent with higher methylation of this gene in European-American individuals resulting in lower expression.

Our results suggest that although *cis*-genetic effects mainly drive systemic DNA methylation differences between African and European human populations, many potential *cis*-mQTLs have yet to be identified. Additionally, differences in DNA methylation between these populations significantly affect divergent regulatory chromatin regions.

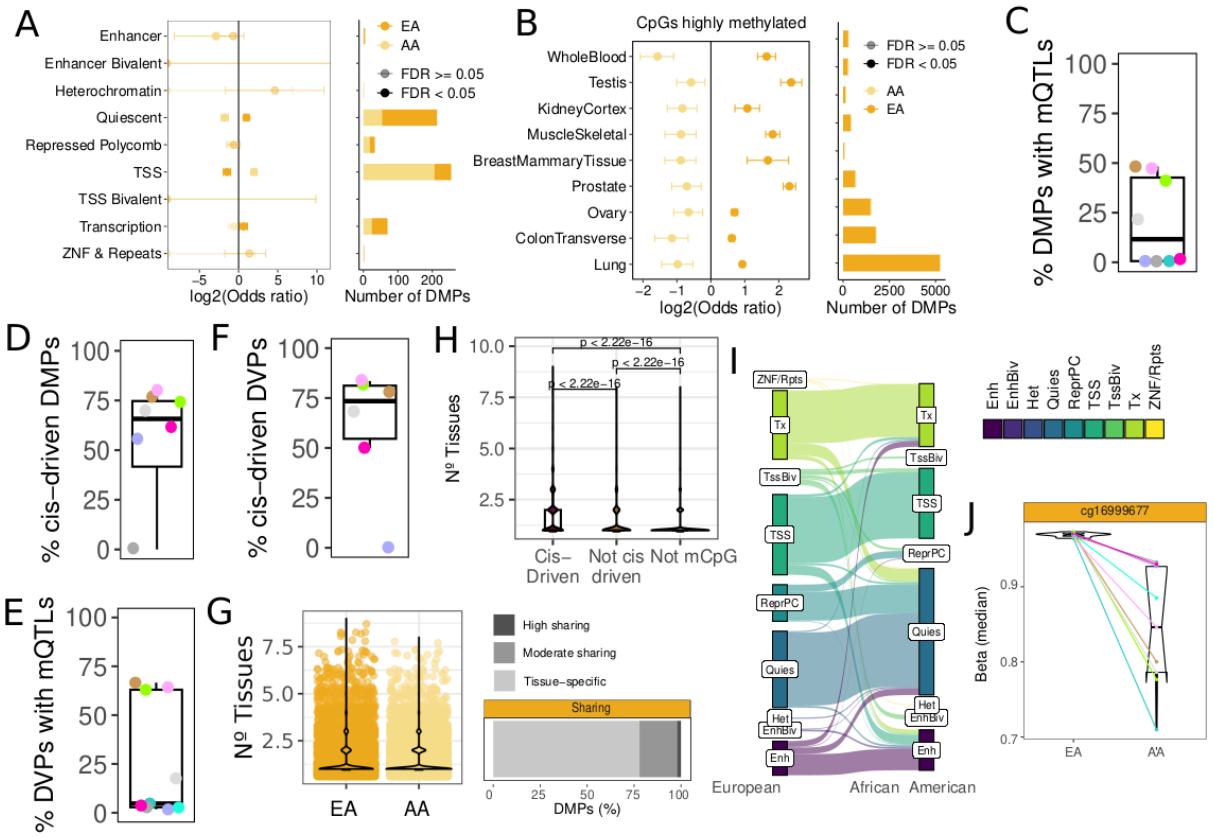


Figure 3. Ancestry-associated DNA methylation changes. A. Enrichment of shared ancestry-DMPs (CpGs differentially methylated with ancestry across 3 or more tissues) in chromatin states. **B.** Enrichment of ancestry-related DMPs in highly methylated CpGs in the genome (mean beta value >= 0.9). **C.** Percentage of DMPs with an associated mQTL in each tissue. **D.** Percentage of DMPs with an associated mQTL that is *cis*-driven. **E.** Percentage of DVPs with an associated mQTL in each tissue. **F.** Percentage of DVPs annotated as mCpGs in which the ancestry differences are driven by previously reported *cis*-driven effects. **G.** Number of tissues in which the ancestry-DMPs are differentially methylated (left) and the proportion of shared ancestry-DMPs (right). **H.** Number of tissues for which a CpG is differentially methylated, comparing Ancestry-DMPs that are *cis*-driven (genetically driven by an mQTL), Ancestry-DMPs that are not *cis*-driven and Ancestry-DMPs without a defined mQTL. **I.** Concordance of whole-blood ancestry-DMPs between chromatin states annotated in a European individual and an African-American individual. **J.** Example of highly shared ancestry-DMPs (CpGs differentially methylated across 8 tissues).

The female genome is hypermethylated at Polycomb binding sites in tissue-shared DMPs

Hypermethylation in females has been previously observed in whole blood[33]. We wanted to address whether autosomal DNA methylation was biased toward hypermethylation in females across tissues. In all tissues, we find significantly more hypermethylated sex-DMPs in females compared to males (two-tailed binomial test; 3.22e-18 < FDR < 8.77e-240 in all tissues except breast and whole blood (two-tailed binomial test; FDR > 0.05), **Fig. 4A**). Consistent with this finding, 75% of the shared sex-DMPs in more than one tissue (tissue-shared sex-DMPs, 15% of the total sex-DMPs), are significantly more methylated in females (two-tailed binomial test; p value = 9.26e-65; **Fig. 4B**).

We then wanted to determine whether hypermethylated sex-DMPs in females are associated with genes with specific functions. We only identified one pathway, male gamete formation, enriched in hypermethylated CpGs in the female transverse colon (**Table S6**). Finally, to assess replication, we compared our findings with those of an independent study in whole blood[33] (see Methods). Similarly, 74% of their sex-DMPs were hypermethylated in females, and when comparing their hypermethylated sex-DMPs with our results, we found a significant overlap (two-tailed Fisher's exact test, p value = 1.29e-9, OR = 331).

We then investigated whether hypermethylated positions in females were located in particular chromatin regions. Sex-associated changes in DNA methylation show tissue-specific enrichment in chromatin states (**Fig. 4C, Fig. S4A**), with the strongest signal found in Polycomb target regions (bivalent TSSs in the lung and Polycomb-repressed regions in the colon) and in enhancers in the muscle and kidney. Interestingly, hypermethylated sex-DMPs in females shared in more than one tissue are preferentially located at repressed Polycomb regions (**Fig. 4D**). Furthermore, when we evaluated the preferential location of tissue-shared sex-DMPs, including those in previous studies[33] (tissue-shared sex-DMPs shared in 2 or more tissues), we observed a higher enrichment and more significant p-value at repressed polycomb regions (OR 2.66 vs. OR 2.71; FDR 3.17e-8 vs. FDR 9.43e-9). To expand these results genome-wide, we evaluated the preferential location of sex-DMRs (differentially methylated regions) recovered using lung WGBS (whole-genome bisulfite sequencing) data from a previous study[68] (see Methods). Similar to our array results, hypermethylated sex-DMRs in females are preferentially located at repressed Polycomb regions and TSSs (**Fig. S4B**). Finally, to further assess the sex specificity of shared sex-DMPs, we compared the DNA methylation values of tissue-shared sex-DMPs between sex-specific tissues (ovary, prostate, and testis). We observed

consistent replication of the female-hypermethylation bias, with DNA methylation values being significantly higher in the ovary than in both the prostate and testis (**Fig. S4C**).

To further explore this pattern of hypermethylation in repressed polycomb regions, we examined whether the female hypermethylated positions were enriched at specific transcription factor-binding sites (TFBSs) based on TF ChIP-seq data from the same tissues[69]. We found enrichment of 30 TFBS in female hypermethylated sex-DMPs (**Table S7**). The tissue-shared female hypermethylated sex-DMPs are enriched in 177 TFBSs (**Fig. 4E, Table S7**), including TFBSs from the polycomb repressive complex and other TFs involved in the maintenance of X chromosome inactivation (i.e., *ASH1L, SETDB1, SAFB*)[70–72]. Additionally, the 177 TFBs significantly located in shared hypermethylated sex-DMPs are enriched in the polycomb repressive complex (**Table S8**). In line with these results, multiple Polycomb repressive complex genes were consistently overexpressed in females across tissues, such as *BMI-1*, which is consistently overexpressed in females across seven different tissues (pancreatic, breast, adipose, skin, thyroid, and muscle)[55]. Finally, the polycomb repressive complex is responsible for the deposition of the H3K27me3 chromatin mark. Therefore, we evaluated whether sex differences in the DNA methylation of Polycomb target regions could also be detected in H3K27me3 patterns across tissues, as demonstrated in human and mouse placentas[73]. We analyzed publicly available H3K27me3 ChIP-Seq datasets from the colon sigmoid (two males and one female, see Methods) and found a strong bias toward female-specific peaks compared with male-specific peaks (324 female-biased peaks vs. 170 male-biased peaks, two-tailed binomial p value 3.9e-17). However, we did not observe this sex bias when analyzing ChIP-Seq data from other tissues, such as the heart or colon transverse.

Together, these results suggest that the female genome is consistently hypermethylated across tissues, especially at binding sites of the polycomb repressive complex.

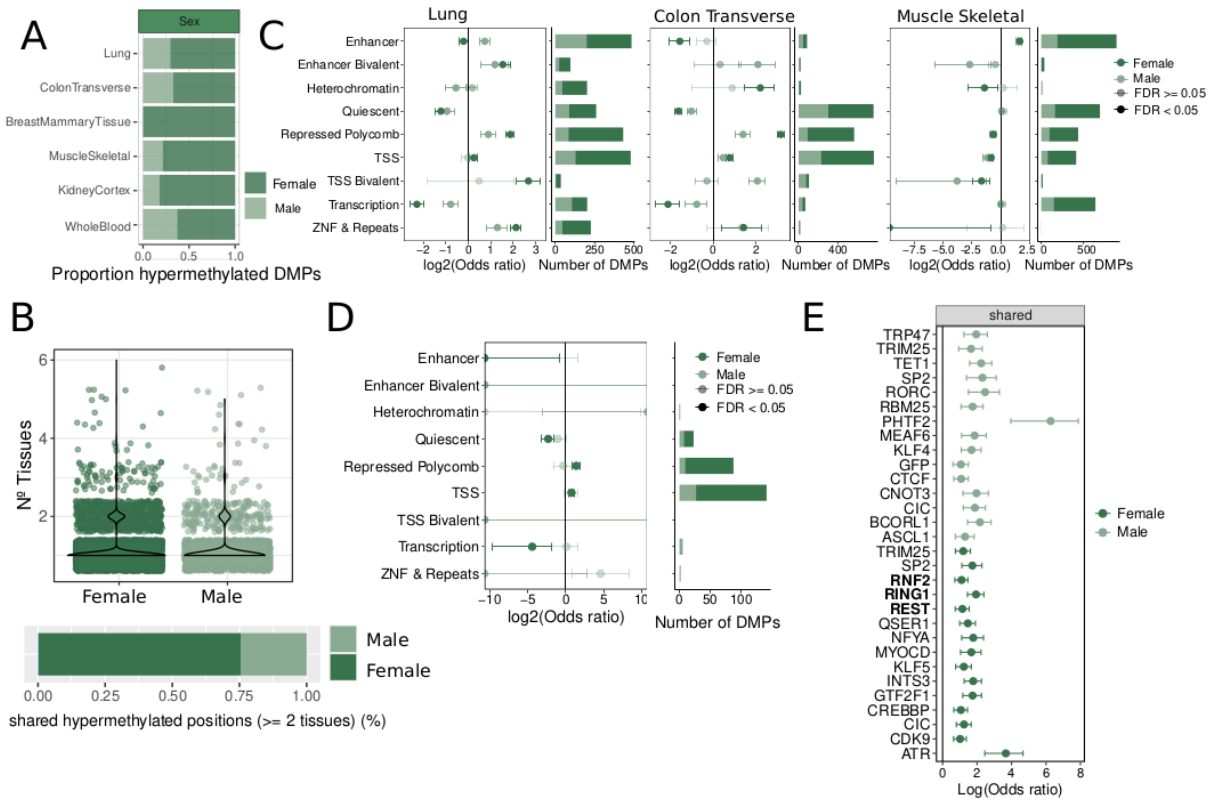


Figure 4. Sex-associated DNA methylation changes. A. Proportion of hypermethylated DMPs in males/females across tissues. **B.** Number of tissues in which the sex-DMPs are differentially methylated (top) and the proportion of shared sex-DMPs in more than 1 tissue by the direction of change (bottom). **C.** Sex-DMPs enrichment at different chromatin states across tissues with the higher number of DMPs. Left. Enrichment and confidence intervals of each tissue-DMPs. Dots are colored according to the direction of the change in DNA methylation and shaded according to the significance of the enrichment (two-tailed Fisher's exact test). Right. Number of DMPs within each chromatin state per tissue. **D.** Enrichment of shared sex-DMPs (CpGs differentially methylated with sex across 2 or more tissues) in chromatin states. **E.** Transcription factor binding sites (TFBS) enriched at shared female hypermethylated sex-DMPs. Polycomb TFs are shown in bold.

DNA methylation consistently increases with age near developmental genes, except in the gonads

DNA methylation changes with age have been extensively studied in different tissues[23,24,74,75]. However, differences in age-related methylation changes between males and females have not been explored before in other tissues beyond whole blood[76,77]. To address this, we identified age-DMPs for each sex separately (see Methods). In most tissues, age had a greater impact on males (**Fig. S5A**), with the largest number of male-age-DMPs and female-age-DMPs located in the colon and ovary, respectively. Downsampling tissues to the same number of samples per tissue and sex shows that colon and ovary remain the tissues

with the highest number of sex-age-DMPs, but we lose the ability to detect DMPs in the tissues with lower sample sizes. (**Fig. S5B**).

Next, we overlapped sex-age-DMPs with the different chromatin states defined by chromHMM in the corresponding tissues[78] to find common and sex-specific patterns of aging DNA methylation variation. Hypermethylated age-DMPs were located across most tissues and for both sexes in chromatin states defined by the H3K27me3 mark (Polycomb target regions, i.e., bivalent TSS, bivalent enhancer, and repressed Polycomb states) (two-tailed Fisher's exact test, FDR < 0.05) (**Fig. 5A-B**), as previously reported in other mammalian tissues[25]. However, the ovaries and testes displayed methylation patterns different from those of the other tissues, with hypermethylation located in the enhancer and quiescent regions, respectively, but not in regions marked by H3K27me3 (**Fig. S5C**).

To further explore the common pattern of aging in both sexes, we focused on the overall aging-related DMPs found in the non-sex stratified age–differential methylation analysis (**Fig. 2A**). We investigated whether the hypermethylated positions in older individuals were enriched at specific TFBSs. Consistent with our previous observations, we found that, among others, TFBS from the polycomb repressive complex (e.g., *CBX8*, *REST*, *RNF2, EZH2, and KDM2B*) were enriched among hypermethylated age-DMPs of the colon, lung, and prostate (**Fig. 5C, Fig. S5D**) but not in the ovary or testis (**Table S9**). Finally, as the Polycomb target regions were hypermethylated in both old (**Fig. 5A-B**) and female (**Fig. 4D**) individuals, we assessed whether age- and sex-related DMPs had independent (additive) contributions to the same CpGs. We find that the additive contribution of age and sex in colon and lung is higher than expected by chance (two-tailed Fisher's exact test, FDR <0.05). As expected, hypermethylated CpGs in females and older individuals were significantly enriched in target regions of the polycomb repressive complex (**Fig. 5D, Fig. S5E**). This finding suggests that a similar mechanism drives DNA methylation changes in females and old individuals.

We next sought to characterize the functional pathways affected by age-methylation changes. Aging influenced the DNA methylation of genes belonging to a wide range of biological pathways, mainly from hypermethylated DMPs with age (**Fig. 5E, Table S10**), and developmental pathways being altered across four tissues (**Fig. S5F**). In addition to aging affecting developmental pathways, it also influences angiogenesis in the ovary (**Table S10**), increasing its DNA methylation with age and probably altering the expression of the associated genes (**Table S2**). Finally, to explore the relationship between changes in biological pathways and genomic location, we stratified DNA methylation changes by

chromatin region. We observed that hypermethylated age-related DMPs in promoters and repressed polycomb and bivalent enhancers were located near genes involved in developmental and metabolic pathways in the lung and colon (**Table S10**).

Age-related DMPs shared across tissues may be significant contributors to phenotypic differences between individuals. Most age-related DMPs (85%) were differentially methylated in a single tissue (**Fig. 5F**). However, age-shared DMPs showed a clear bias toward hypermethylation in older individuals, with ~84% of the shared DMPs (363 out of 435 DMPs in 4 or more tissues) consistently increasing in methylation with age (two-sided binomial test; p value=2.795349e-30). Tissue-shared hypermethylated age-DMPs are enriched in developmental and differentiation pathways (**Table S11**) and are preferentially located in core promoter regions (**Fig. 5G**), possibly impacting the expression of the associated genes.

Overall, we find that aging leads to the hypermethylation of developmental genes, with most hypermethylation occurring at promoters and target regions of the polycomb repressive complex for both sexes, except for the ovary and testis. Finally, angiogenesis could be altered by DNA methylation changes in their associated genes in the ovary, probably indicating reproductive-specific aging effects on DNA methylation.

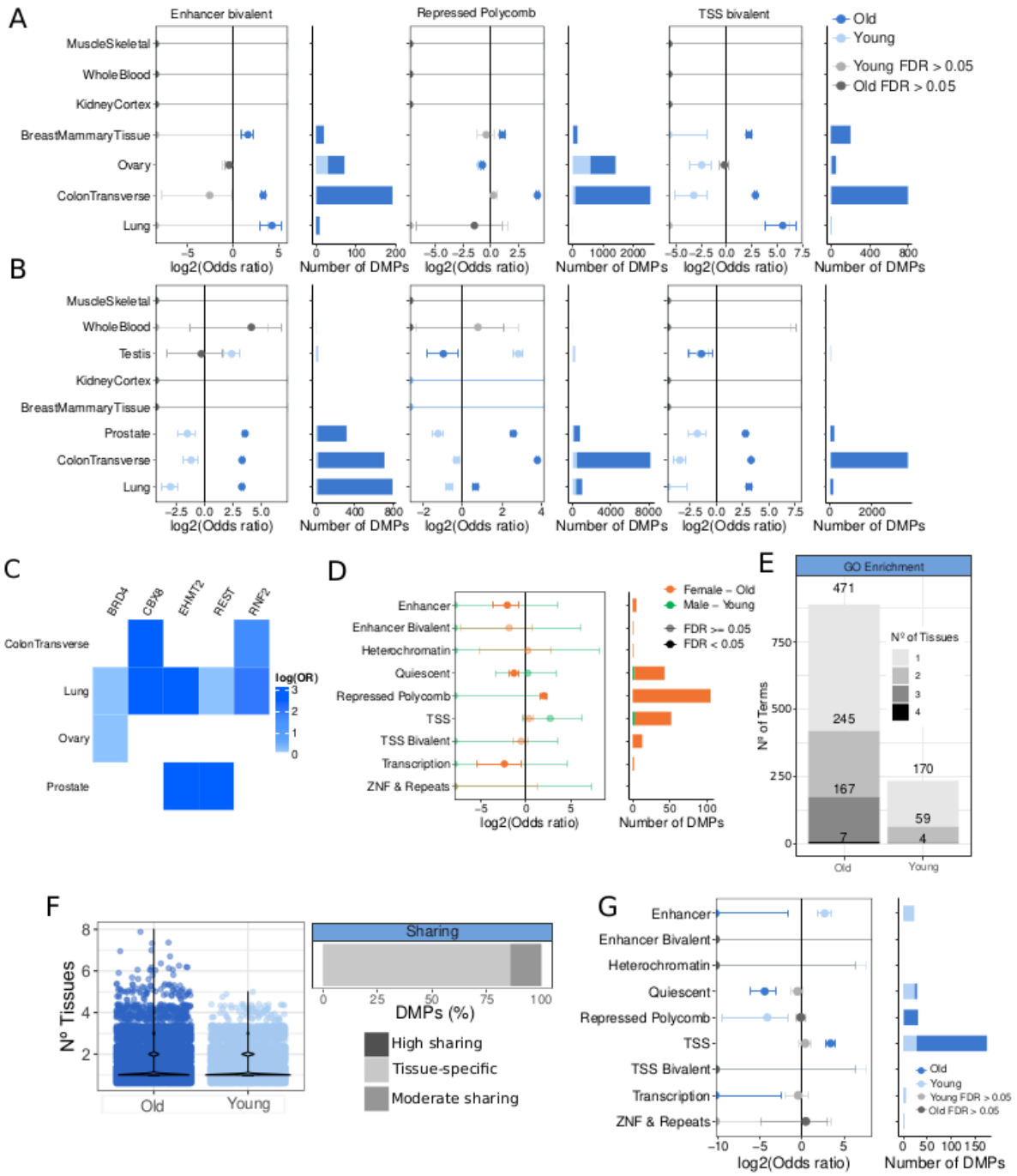


Figure 5. Age-associated changes in DNA methylation. A. Female age-related DMP enrichment at 3 chromatin states across tissues. **Left.** Enrichment of each tissue-DMPs. Dots are colored according to the direction of the change in DNA methylation. Non-significant enrichments are colored in gray tones (two-tailed Fisher's exact test). Confidence interval is shown. **Right.** Number of DMPs within each chromatin state per tissue. **B.** Male age-related DMP enrichment at 3 chromatin states across tissues. **Left.** Enrichment of each tissue-DMPs. Dots are colored according to the direction of the DNA methylation change and significance (two-tailed Fisher's exact test). Confidence interval is shown. **Right.** Number of DMPs within each chromatin state per tissue. **C.** TFBSs enriched in hypermethylated age-DMPs shared between 2 tissues. OR = odds ratio. **D.** Enrichment of sex–age additive effects across chromatin states in colon transverse. **Left.** Enrichment of DMPs. Dots are colored according to the direction of the DNA methylation change and significance (two-tailed Fisher's exact test). Confidence interval is shown. **Right.** Number of DMPs within each chromatin state per tissue. **E.** Number of tissue-specific and tissue-shared gene ontology enriched terms for age-hyper and age-hypo-methylated positions (FDR < 0.05, GO biological process). **F.** Number of tissues in which the age-DMPs are differentially methylated (top) and the proportion of shared age-DMPs (bottom). **G.** Enrichment of shared age-related DMPs (CpGs differentially methylated with age across 3 or more tissues) in chromatin states. Dots are colored according to the direction of the change in DNA methylation. Non-significant enrichments are colored in gray tones (two-tailed Fisher's exact test). Confidence interval is shown.

DISCUSSION

The study of interindividual DNA methylation variation is essential for advancing our understanding of human biology and disease. Here, we jointly quantify the associations of four demographic traits with DNA methylation across multiple tissues to compare their relative contributions to DNA methylation variation. We find that genetic ancestry and age are the main contributors to DNA methylation variation across tissues, with aging effects being less strong and, hence, more difficult to detect with lower sample sizes. We finally characterized DNA methylation differences across individuals and hypothesize about their functional consequences.

Differences in DNA methylation between human populations have been described in whole blood[34–37]. Here, we report a generalized difference in the DNA methylation levels of quiescent and promoter regions between African and European populations that is not influenced by differences in chromatin states. We hypothesize that the higher genetic variability in African populations[58] is coupled with higher DNA methylation variability in this population, consistent with the accumulation of ancestry DNA methylation changes in genomic regions with extreme DNA methylation levels (i.e., quiescent regions being highly methylated and promoters being not methylated), where subtle differences in DNA methylation are easier to detect. Importantly, we show that DNA methylation variation in promoter regions has a small, but consistent, influence on the gene expression of associated genes. This finding suggests potentially important functional consequences of the differences in DNA methylation between populations and highlights the importance of studying epigenetic inter-individual variation among diverse genetic ancestries. Finally, although we find a large portion of the ancestry DNA methylation variation to be genetically driven, only a very small proportion of the DMPs and DVPs were associated with *cis*-mQTLs, with especially low numbers in tissues with small sample sizes. This suggests that many potential *cis*-mQTLs still need to be identified. Therefore, future efforts are needed to elucidate the effects of genetics on DNA methylation variation in larger and more diverse cohorts and under multiple conditions and cell types.

Recent work has shown that chromatin states differ between African and European individuals[60], with previous studies demonstrating extensive differences in DNA methylation profiles, specifically impacting immune regulation[34,79]. Our findings expand this previous knowledge by revealing that differences in DNA methylation between these populations

significantly affect chromatin states that are differentially annotated between populations. In addition, the differences in DNA methylation between these populations in enhancers impacting inflammation-related pathways could be linked to the increased levels of inflammation in African American individuals previously reported[60,79,80]. Together, these findings suggest a direct role for DNA methylation in regulating immune responses, highlighting the need for further investigation into this complex relationship to advance into personalized medicine and inform clinical responses given an individual's genetic ancestry.

Hypermethylation in females has been previously observed in whole blood[33]. Here, we provide evidence that the female autosomal genome is hypermethylated across tissues, specifically at Polycomb-repressed regions. The polycomb repressive complex is known to play a role in repressing one of the female X chromosomes in the process of X chromosome inactivation[81,82]. One possibility that could explain our findings is that the continuous repression of the X chromosome, and hence, the increased activity of the polycomb repressive complex and the DNA methyltransferases, could drive methylation beyond CpGs located in the sex chromosomes. This possibility is concordant with female-biased genes significantly enriched in targets of Polycomb repressive complex 2 (*PRC2*) and H3K27me3 chromatin mark deposition[83]. Additionally, sex-specific deposition of the H3K27me3 chromatin mark has been previously reported in the placenta[73] and liver[84], consequently impacting gene expression. In line with this, multiple genes from the polycomb repressive complex (*BMI1, EZH2, SUZ12*) and genes known to interact with *XIST* (*ERH, RNMT, TRA2B*), a major effector of the X inactivation process[70], are upregulated in females in multiple tissues[55]. Additional studies are needed to elucidate the underlying mechanisms and functional implications of the hypermethylation of the female genome, specifically at Polycomb target regions.

Finally, the accumulation of DNA methylation changes with age has been previously identified[75]. In particular, hypermethylation in target regions of the polycomb repressive complex has been shown to be conserved across mammalian species and tissues[25]. However, the differences in age-related methylation changes between males and females have not yet been well characterized. Here, we provide evidence for systematic hypermethylation of polycomb-target regions and regulatory regions associated with developmental genes in both sexes. Hypermethylation in older individuals is specifically located in CpG-rich regions (CpG islands), and even though the mechanism remains unclear, one of the strongest hypotheses involves DNA damage. The DNA repair machinery relocalizes methylation enzymes and the polycomb repressive complex to sites of DNA damage. However, the proteins recruited to damaged sites preferentially target CpG-rich

regions[74,85]. These findings suggest that pervasive DNA damage with aging leads to the hypermethylation of CpG islands and most prominently affects the targets of the polycomb repressive complex. Despite the wide tissue sharing of age-related hypermethylation in Polycomb repressive complex target regions, we did not observe the same trend in the gonads (ovary and testis). This finding suggests an important, conserved, and distinct role of the polycomb repressive complex in the gonads[86–88] as well as distinct aging DNA methylation patterns[89–92]. Further research on the sex-specific effects of aging with larger and sex-balanced cohorts is needed to deepen our understanding of how aging affects males and females differently, specifically in the gonads.

Overall, our multi-individual and multi-tissue approach provides an extensive characterization of the main drivers of human DNA methylation variation across individuals in healthy conditions. This study provides a baseline for future studies aiming to elucidate DNA methylation changes related to disease phenotypes, creating a foundation for advancing personalized medicine.

Limitations of the study

Although our study employs a multi-tissue and multi-individual approach, it has certain inherent limitations that should be considered in the design of future studies. First, the GTEx dataset has a limited representation of human populations (restricted to African Americans and European Americans) and a skewed age distribution that favors older individuals. Our findings emphasize the pressing need to incorporate individuals with diverse ancestries, as well as developmental and pediatric samples, into multi-omic analyses, as these demographic traits significantly influence DNA methylation and gene expression variation. Second, our ability to detect DNA methylation changes is constrained by reduced sample sizes and the use of arrays instead of whole-genome bisulfite sequencing, which limits statistical power. Finally, the limited number of available tissues and the analysis of bulk samples hamper our ability to detect cell type-specific methylation changes. Taken together, future organ single-cell atlases including more tissues, donor diversity and multiple conditions will help deepen our understanding of how demographic traits influence DNA methylation.

Conclusions

In summary, we characterized the contributions of demographic traits to DNA methylation variation across multiple human tissues. Our findings reveal that genetic ancestry and age are the primary determinants of DNA methylation variation, with aging effects being subtler

and harder to detect in smaller cohorts. We identified significant DNA methylation differences between African and European populations, particularly in quiescent and promoter regions, which may have functional consequences on gene expression and chromatin states. Additionally, we defined sex-specific DNA methylation patterns, such as female hypermethylation at Polycomb-repressed regions, and systematic age-related hypermethylation in CpG-rich regions, emphasizing their potential roles in gene regulation and aging processes. This work provides a comprehensive characterization of the key contributors to DNA methylation variation and its functional implications. Future research incorporating larger, more diverse cohorts, single-cell analyses, and advanced sequencing technologies will be essential to fully understand the functional implications of these methylation variations, paving the way for personalized medicine and better insights into human health and disease.

ACKNOWLEDGEMENTS

We thank Aida Ripoll-Cladellas and Fairlie Reese for useful discussions at the end of the work. Fig. 1a was created with BioRender.com.

FUNDING

J.M.R. was supported by a predoctoral fellowship from the “la Caixa” Foundation (ID 100010434) with the code LCF/BQ/DR22/11950022. M.M. was supported by grant PID2019-107937GA-I00 funded by MCIN/AEI/10.13039/501100011033 and grant RYC-2017-22249 funded by MCIN/AEI/10.13039/501100011033 and by “ESF Investing in your future”.

AUTHORS’ CONTRIBUTIONS

M.M. conceived the study and designed and supervised all the analyses;

W.O. and J.M.R. analyzed the data;

W.O. led the data analysis related to differential methylation; J.M.R. led the data analysis related to differential variability.

M.M., W.O., and J.M.R. wrote the manuscript.

Corresponding author

Correspondence to Marta Mele (marta.mele@bsc.es).

AVAILABILITY OF DATA AND MATERIALS

All the GTEx protected data are available at dbGap under the accession number phs000424.v8. (https://gtexportal.org/home/protectedDataAccess). DNA methylation data is available in GEO under the accession number GSE213478. Public-access data, including QTL summary statistics and expression levels, are available on the GTEx Portal, as downloadable files and through multiple data visualizations and browsable tables (https://www.gtexportal.org), as well as in the UCSC and Ensembl browsers.
Analysis scripts are available at https://github.com/Mele-Lab/, and all the results tables derived from the analyses conducted in this paper are deposited at zenodo (doi: X).

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

All human donors were deceased, with informed consent obtained via next-of-kin consent for the collection and banking of deidentified tissue samples for scientific research. The research protocol was reviewed by Chesapeake Research Review Inc., Roswell Park Cancer Institute's Office of Research Subject Protection, and the institutional review board of the University of Pennsylvania.

CONSENT FOR PUBLICATION

Not applicable.

COMPETING INTERESTS

The authors declare that they have no competing interests.

METHODS

GTEx subjects

There were 838 donors (557 biological sex males and 281 biological sex females). The age of the donors ranged from 20-70 years, with most enrolled donors being older individuals. For more details on donor characteristics and sample collection, see the GTEx v8 main paper (DataS2)[93]. For details on the generation of the DNA methylation data and the QC pipeline, see Oliva et al.[43], and for more details on the gene expression pipeline, see the GTEx v8 main paper[93].

Ancestry annotation

We used GTEx v8 release genotype data[93] to infer ancestry for each donor. We filtered the phased GTEx v8 whole-genome sequencing variant call file (VCF) (dbGaP accession number phs000424.v8) from the reference genome Human Build 38 (hg38) with Plink 2.0 (ref) to (1) generate a pruned subset of SNPs based on Linkage disequilibrium with the

option --indep-pairwise 50 10 0.1, and (2) filter variants to keep only those on chromosome 21 as a reference with the option --chr 21 and –make-bed to create a final bed file with the filtered SNPs. The final bed file was filtered to keep only individuals with European–American and African–American ancestry, as well as admixed individuals, as identified by the GTEx consortium[93]. We then estimated the admixture proportion for each individual based on the pruned and filtered set of SNPs using the ADMIXTURE program (v.1.3.0)[94]. ADMIXTURE was run within a two/three/four-ancestry population model in 48 threads. To define the most accurate number of ancestry populations (K parameter), we selected the model with the lowest cross-validation error (ancestry population model with K = 2).

Differential methylation analysis

We downloaded normalized beta counts of the 754,054 CpGs from the Infinium Methylation EPIC array generated in Olivia et al.[43] stored in GEO (GSE213478). For CpG QC, Oliva et al.[43] excluded CpGs that had low detection P and low bead counts, cross-reactive CpGs and those overlapping a variant (or within a single base pair extension) and mapping to sex chromosomes. We used the *limma* pipeline implemented in the R package limma (v3.58.1)[95] to run linear fixed-effects models on M values[96] and corrected for the following set of covariates to be as similar as possible to previously published research[55,97]:

M values ~ HardyScale + IschemicTime + PEER1 + PEER2 + PEER3 + PEER4 + PEER5 + ancestry + sex + age + BMI

To account for hidden batch effects and other confounders of DNA methylation variance, we used the probability estimation of expression residuals (PEER) method. We included 5 PEERs in the models, as it was the maximum number of PEERs available in all tissues, previously computed in Oliva et al.[43]. These factors correct for known sources of technical variation and other confounders, such as cell type composition, as shown in the original publication[43]. We corrected all analyses for multiple testing using false discovery rate (FDR) through the Benjamini–Hochberg method and considered CpGs differentially methylated (DMPs) at an adjusted p value below 0.05. The aging variable was categorized for visualization purposes.

To investigate the sex-specific aging effect, we divided the donors for each tissue into males and females, and we used the same *limma* pipeline and ran linear fixed-effects models on M values for each sex separately. Owing to the low variability of some of the previously included covariates, we retained only the first two PEER factors:

M values ~ PEER1 + PEER2 + Ancestry + Age + BMI

We corrected all sex-specific analyses for multiple testing separately by using false discovery rate (FDR) through the Benjamini–Hochberg method and considered CpGs differentially methylated (sex-age-DMPs) at an adjusted p value less than 0.05.

To investigate additive effects, we conducted two-tailed Fisher's exact tests using the fisher.test function in the R package stats (v4.3.2) with default parameters for the common DMPs between each pair of demographic traits per tissue. The analysis was corrected using FDR through the Benjamini–Hochberg method, and significant effects with adjusted p values less than 0.05 were considered. To determine whether the direction of change was concordant across demographic traits, we conducted a chi-square test and corrected for multiple testing in the same way.

To validate the DMPs between ancestries, we compared our results to those obtained in an independent study using DNA methylation arrays[60] (Table S5 from the original publication). We performed two-tailed Fisher's exact tests to check whether the DMPs we identified overlapped with the positions observed in this previous study. The background was the set of studied positions from the EPIC array that were found in their study.

To validate the DMPs between sexes, we compared our results to those obtained with an independent dataset using DNA methylation arrays. Whole blood sex-DMPs were obtained from Additional file 1 from Grant et al.[33]. We performed two-tailed Fisher's exact tests to check whether the DMPs we identified overlapped with the positions observed in this previous study. The contingency table for Fisher's exact test includes 4 DMPs common between both datasets, 23 DMPs only in the GTEx dataset, and 391 DMPs only in the external dataset. The background was the set of studied positions from the EPIC array that were found in their study.

Downsampling for differential methylation analysis

To fairly compare the number of DMPs between tissues, we performed a downsampling analysis on 35 samples. We performed 50 random permutations on 35 samples from every tissue and reported the median number of DMPs per tissue and trait. We chose this approach, as downsampling to the same distribution of the tissue with the lowest sample size (skeletal muscle) would result in not enough young individuals to observe any signal in the other tissues. Using this approach also allowed us to downsample all the tissues (including skeletal muscle).

To fairly compare the number of DMPs between sexes in the sex-specific model of aging, we performed a downsampling analysis on the number of samples available for the sex with the lowest number of available donors per tissue.

Differential variability methylation analysis

We used the function varFit from the *missMethyl*[98] R package (v3.58.1) to test for differential variability running a linear fixed-effects model on M values. We corrected for the same covariates as those used in the differential methylation analysis:

M values ~ HardyScale + IschemicTime + PEER1 + PEER2 + PEER3 + PEER4 + PEER5 + ancestry + sex + age + BMI

We corrected all analyses for multiple testing using false discovery rate (FDR) via the Benjamini–Hochberg method and considered CpGs that were variably methylated at an adjusted p value below 0.05.

To investigate common differential variability in methylation and differential methylation effects, we conducted two-tailed Fisher's exact tests on the common DMP-DVPs associated with each demographic trait per tissue. The analysis was corrected using FDR through the Benjamini–Hochberg method, and significant effects with adjusted p values less than 0.05 were considered.

We then applied the same methodology to identify differentially variable genes at the transcriptional level. We ran a linear fixed-effects model on the gene expression counts, correcting for technical covariates to have the most similar model to DNA methylation as possible:

gene counts ~ HardyScale + IschemicTime + RIN + ExonicRate + PEER1 + PEER2 + ancestry + sex + age + BMI

To account for hidden batch effects and other confounders of gene expression variance, we used the probability estimation of expression residuals (PEER) method. We included 2 PEERs in the models as they were mostly correlated with cell type heterogeneity (see Figure S4A from [99]), and the second PEER factor was mostly correlated with the sequencing batch (see Figure S8A from [100]). We corrected all analyses for multiple testing using false discovery rate (FDR) via the Benjamini–Hochberg method and considered genes that were variably expressed at an adjusted p value below 0.05.

Annotation of DMPs

General classification. The DMPs were classified based on their genomic location at promoters, enhancers, gene bodies, or intergenic regions on the basis of the annotations provided in the EPIC v1.0 array manifest b5. Positions annotated as "Promoter_Associated" under "Regulatory_Feature_Group" or under "TSS200" or "TSS1500" under

“UCSC_RefGene_group” were assigned as promoters. The ones with any value in “Phantom5_Enhancers” were assigned as enhancers. From the rest, the positions including “Body”, “1stExon”, “ExonBnd”, “3UTR” or “5UTR” under “UCSC_RefGene_group”, were annotated as gene bodies. The rest were annotated as intergenic. To study the enrichment of demographic trait-DMPs in the different genomic locations, we performed a two-tailed Fisher’s exact test for each region separately for hypomethylated and hypermethylated DMPs. We used as background the set of studied positions from the Illumina EPIC array. The analysis was corrected using FDR through the Benjamini–Hochberg method, and significant effects with adjusted p values less than 0.05 were considered.

Tissue-specific chromatin state classification. To annotate the chromatin states around each CpG, we used the 18 chromatin states inferred with ChromHMM[101] generated by the ROADMAP Epigenomics consortium[102] and analyzed by EpiMap[78] for the lung (BSS01190), colon transverse (BSS01848), breast (BSS00145), testis (BSS01718), prostate (BSS01459), ovary (BSS01399), muscle skeletal (BSS01577), and kidney cortex (BSS01096) and PBMCs (BSS01419). TssFlnkD, TssFlnk, TssA and TssFlnkU are reported together under “TSS". EnhA1, EnhA2, EnhG1, EnhG2, and EnhWk are reported together under “Enh”. Tx and TxWk are reported together under “Tx”, and ReprPC and ReprPCWk are reported together under “ReprPC”. (**Supplementary Fig. 3A**). The data were downloaded from https://personal.broadinstitute.org/cboix/epimap/ChromHMM/observed_aux_18_hg19/CALLS/. Whole-blood chromatin states from African American and European individuals were downloaded from ENCODE (African individual ENCFF249LTX; European individual ENCFF158URQ). Colon chromatin states from 2 female and 2 male individuals were downloaded from ENCODE (Females BSS01848 and BSS01849; Males BSS01850 and BSS01851). Enhancers that were differentially annotated between populations were obtained from Table S5 from Aracena et al.[60].
To perform enrichment of DMPs in the different chromatin states, we performed a two-tailed Fisher’s exact test, similar to the analysis of the different genomic locations. We used as background the set of studied positions from the Illumina EPIC array. To perform enrichment of DMRs in the different chromatin states, we performed a two-tailed Fisher’s exact test, similar to the analysis of the different genomic locations. We used as background the whole set of discovered DMRs and, as a significant group, the DMRs with 3 or more CpGs in the region.

Transcription factor-Binding Sites (TFBS). To study the enrichment of TBFS around DMPs, we downloaded processed ChIP-seq data on transcription factors for the human v19 Lung, Kidney, Blood, Muscle, Breast, Digestive Tract, Prostate and a general catalog available in

ChipAtlas[69]. The transcription factors involved in the polycomb repressive complex available in the ChIP-seq data were *EZH1/2, SUZ12, YY1, KDM2B, REST, PCGF2, CBX2, CBX8, RNF2, and RYBP*.

Gene assignment. The probe–gene pairs were retrieved first from the EPIC v1.0 manifest b5. We considered a probe to be part of a pair for every gene annotated under "UCSC_RefGene_Name". For the probes annotated as promoters or enhancers for which no assigned genes were present, we annotated the closest gene via the function *matchGenes* from the R package *bumphunter* (v1.44.0)[103]. We downloaded genes belonging to the MHC complex assembly and antigen processing pathways from the GO database (GO:0002396 and GO:0019882, respectively) for Figure 1F.

Enrichment of Transcription Factors

We performed Fisher's exact test for each transcription factor (TF) separately for hypomethylated and hypermethylated DMPs and adjusted for multiple testing using Benjamini–Hochberg correction with an FDR<0.05. We used as background the TF-binding sites (TFBS) found in all tested CpGs from the Illumina EPIC array.

Functional enrichment analyses

To perform functional enrichment analyses (overrepresentation analyses (ORA)) with the EPIC array, we need to consider that some genes contain more probes than others. For this goal, we used the *gometh* function from the *missMethyl* R package (v1.36.0)[98] to obtain the GO:BP terms. For each ORA, we carefully selected a suitable background CpG list. To investigate the biological pathways associated with each tissue- and tissue-shared DMP, we used 754,054 positions from the array as background.

Gene expression quantification

Gene and transcript quantifications were based on the GENCODE 26 release annotation (https://www.gencodegenes.org/releases/26.html). We downloaded gene counts and TPM quantifications from the GTEx portal (https://gtexportal.org/home/datasets). We selected genes with protein-coding and lincRNA biotypes on the GTEx GENCODE v26 GTF. For the analyses involving expression data, we considered expressed genes per tissue (TPM ≥ 1 and ≥ 10 reads (unnormalized) in ≥ 20% of the tissue samples), excluding genes in the pseudoautosomal region (PAR). In total, we analyzed 22,967 genes (18,185 protein-coding genes and 4,782 lincRNAs) across tissues.

Hierarchical partition analysis

To calculate the independent relative contributions of tissue/donor and different demographic traits to DNA methylation variation, we used a variance partitioning approach using the R package variancePartition R (v1.32.5)[104]. Specifically, variance partitioning decomposes the model R2 using a linear mixed model to quantify the variation in DNA methylation attributable to each variable in the model. The model used to calculate the relative contribution of tissue and individual was:

M ~ PEER1 + PEER2 + PEER3 + PEER4 + PEER5 + (1 | SUBJID) + (1 | Tissue)

The model used to calculate the relative contribution of each demographic trait was as follows:

M ~ PEER1 + PEER2 + PEER3 + PEER4 + PEER5 + HardyScale + IschemicTime + sex + age + BMI + ancestry

We then applied the same methodology to calculate the independent relative contribution of tissue/donor and each demographic trait to the gene expression variation, using the same subset of tissues and samples. The models used are specified below:

(1) *gene counts ~ RIN + ExonicRate + IschemicTime + (1 | Donor) + (1 | Tissue)*

(2) *gene counts ~ RIN + PEER1 + PEER2 + ExonicRate + IschemicTime + HardyScale + ancestry + sex + age + BMI*

Genes or CpGs highly explained by individual were selected based on the R^2 of individual > 0.5. Genes or CpGs highly explained by tissue were selected based on the R^2 of tissue > 0.5.

Identification of *Cis*-driven Effects

To explore whether ancestry-related DMPs/DVPs are driven by *cis* genetic effects, we focused on those previously reported as mCpGs (CpGs with significant *cis*-mQTLs)[43]. We filtered variants by a MAF > 0.01 and a window of 500 kb around the CpG. We run two different models per CpG, excluding (H0) and including (H1) the variable Ancestry while correcting in both cases for the *cis*-mQTL variants:

H0: M values ~ HardyScale + IschemicTime + PEER1 + PEER2 + PEER3 + PEER4 + PEER5 + Sex + Age + BMI + cis-mQTL genotype {1…n}

H1: M values ~ HardyScale + IschemicTime + PEER1 + PEER2 + PEER3 + PEER4 + PEER5 + Sex + Age + BMI + cis-mQTL genotype {1…n} + Ancestry

We applied the *anova* function from the R stats package (v4.3.2) to compare the models using F tests. After multiple test corrections with the Benjamini–Hochberg method, if FDR < 0.05, we could reject the null hypothesis in which the two models are not significantly different and conclude that the ancestry effect is not solely driven by *cis* genetic effects. We considered the mCpGs associated with at least one independent *cis*-mQTL. We excluded *cis*-mQTLs if any individual was missing the genotype or if there were fewer than 3 genotyped individuals of each allelic combination. To avoid including linearly dependent *cis*-mQTLs, we computed the pairwise correlation matrix among the *cis*-mQTLs using the *findCorrelation* function in the caret R package (v6.0.94)[105] and excluded highly correlated *cis*-mQTLs (correlations > 0.85).

Enrichment of ancestry-related DMPs in *cis*-mQTLs

To investigate whether ancestry-related DMPs are overrepresented in mCpGs (CpGs with significant *cis*-mQTLs), we downloaded the latest analyses from Olivia et al.[43] (https://gtexportal.org/home/datasets). As indicated in the GTEx portal, to obtain the list of mCpGs per tissue, we selected *cis*-mQTLs with q values < 0.05. Then, for each tissue, we computed a two-tailed Fisher's exact test to test whether there were more ancestry-DMPs that were also mCpGs than expected if they were independent. We used as background the set of studied positions from the Illumina EPIC array. We determined a significant enrichment after correcting for multiple testing via the Benjamini–Hochberg method across tissues at an FDR < 0.05.

Enrichment of ancestry-related DMPs in regions with extreme methylation status

To assess whether African-DMPs were enriched in regions with extreme DNA methylation profiles, such as promoters or quiescent regions, we first generated a catalog with the coordinates of those regions. Briefly, we selected CpGs that, across donors, had average methylation levels for a specific tissue above 0.9 and CpGs that had average methylation levels below 0.1. We then computed a two-tailed Fisher's exact test to test whether there were more ancestry-DMPs located in those regions than expected (using as background the location of all the tested CpGs from the Illumina EPIC array). We determined a significant

enrichment after correcting for multiple testing via the Benjamini–Hochberg method across tissues at an FDR < 0.05.

Correlation between DNA methylation and gene expression

The probe–gene pairs were retrieved first from the EPIC v1.0 manifest b5. We considered a probe to be part of a pair for every gene annotated under "UCSC_RefGene_Name". For the probes annotated as promoters or enhancers for which no assigned genes were present, we annotated the closest gene using the function *matchGenes* from the R package *bumphunter* (v1.44.0)[103]. The total numbers of CpGs tested in promoters, enhancers, and gene bodies are 108,810, 25,287, and 370,343, respectively. We used the DNA methylation and gene expression residuals, after regressing out the effects of cell type composition (PEER factors) and technical covariates. We computed Pearson's correlation for the DMP-DEG pairs and corrected for multiple testing using the Benjamini–Hochberg method separately per genomic region (enhancer, promoter, and gene-associated). We considered correlations with an FDR < 0.05 to be significant. To compute the percentage of DEGs that significantly correlated with DMPs, we considered DEGs associated with at least one probe in the array. To increase our detection power, we repeated the analysis considering as DMP-DEG pairs all the DMPs significant at a p value nominal < 0.05.

Mediation analysis

To identify genes whose changes in expression with demographic traits were mediated by the DNA methylation of the associated CpGs, we used a regularized mediation analysis based on penalized structural equation modeling implemented in the R package *regmed* (v2.1.0)[56,106]. For each DEG for a particular tissue and demographic trait, we considered all the associated DMPs as mediators. To avoid including linearly dependent CpGs, we computed the pairwise correlation matrix among the DMPs using the *findCorrelation* function in the caret R package (v6.0.94)[105] and excluded highly correlated CpGs (correlations > 0.9). We followed the published pipeline to run mediation with one exposure and multiple mediators with a penalty "lambda" grid of *seq (from = 0.6, to = 0.01, by = -0.05)*. We then selected the best model with the function regmed.grid.bestfit and considered all those genes with at least one CpG having an alpha*beta different from 0 to be mediated.

Ancestry-specific chromatin regions

To address whether Ancestry-DMPs were located at genomic regions that had different chromatin states between African and European human populations, we first created a

catalog of CpGs with an annotation of whether the chromatin state of its location was the same ("Common region") or different ("Ancestry-specific") between these populations. Additionally, we focused on the "Ancestry-specific" regions and computed a two-tailed Fisher's exact test to test whether there were more ancestry-DMPs located in any of those regions than expected (using as background the location of all the tested CpGs from the Illumina EPIC array), separately for hypermethylated and hypomethylated positions. We determined a significant enrichment after correcting for multiple testing via the Benjamini–Hochberg method across tissues at an FDR < 0.05.

Differential peak calling between sexes

To assess whether sex-related DMPs are related to differential Polycomb repressive complex binding, we analyzed publicly available ChIP-Seq data for the H3K27me3 chromatin mark in heart and colon tissue: 4 heart, 3 colon sigmoid, and 4 colon transverse samples (male samples: GSM7247417, GSM7247418, GSM6347302, GSM6347303, GSM6347304, GSM4250725, GSM4250726, GSM4146448, GSM4146449, GSM4250562, GSM4250563, GSM4250564, and GSM4250565; female samples: GSM5669434, GSM5669435, GSM5669390, GSM5669389, GSM4247242, GSM4247243, GSM5112491, GSM5112492, GSM5112493, GSM5112494, and GSM2701098). We then used the differential peak calling method from the ChIPAtlas (Diff Analysis)[69], with default parameters.

Mapping and quality control of WGBS data

We downloaded fastq files for the WGBS data of lung GTEx samples from dbGaP under the accession number phs000424. We trimmed reads of their adapter sequences using Trim Galore (v0.6.6)[107] with a quality trimming cutoff of 30. We then aligned the trimmed reads to the hg38 build of the human genome using Bismark (v0.23.1)[108] with default parameters. We then used bismark_methylation_extractor to summarize the number of reads supporting a methylated cytosine and an unmethylated cytosine for every cytosine present in the reference genome. Consistent with previous studies[68], we included the parameters *"--ignore 2 --ignore_r2 3 --ignore_3prime 1 --ignore_3prime_r2 3"*. The final report file summarizes the evidence of methylation at each cytosine in the human reference genome.

DMR analysis

We used the R package *bsseq* to analyze the WGBS data[109]. Differentially methylated regions (DMRs) were identified as previously described[68], with a minimum of 70 CpGs required in a smoothing window. Briefly, we used BSmooth[110] to estimate CpG methylation

levels. We ran a ‘small’ smooth to estimate the methylation level in a small genomic region (smoothing over windows of at least 1 kb containing at least 70 CpGs). Following smoothing, we analyzed all CpGs that had a sequencing coverage of at least 1 in all samples for a total of 16,290,182 CpGs. Finally, we computed t-statistics and identified DMRs by thresholding the t-statistics following the *bsseq* pipeline. We considered as significant DMRs those with at least 3 CpGs in them.

References

1. Jin, B., Li, Y., and Robertson, K.D. (2011). DNA Methylation: Superior or Subordinate in the Epigenetic Hierarchy? Genes Cancer *2*, 607.

2. Jones, P.A., and Taylor, S.M. (1980). Cellular differentiation, cytidine analogs and DNA methylation. Cell *20*. https://doi.org/10.1016/0092-8674(80)90237-8.

3. CpG island methylation pattern in different human tissues and its correlation with gene expression (2009). Biochem. Biophys. Res. Commun. *383*, 421–425.

4. Smith, Z.D., and Meissner, A. (2013). DNA methylation: roles in mammalian development. Nat. Rev. Genet. *14*, 204–220.

5. Jones, P.A. (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat. Rev. Genet. *13*, 484–492.

6. Robertson, K.D. (2005). DNA methylation and human disease. Nat. Rev. Genet. *6*, 597–610.

7. Jones, P.A. (1999). The DNA methylation paradox. Trends Genet. *15*, 34–37.

8. Babenko, V.N., Chadaeva, I.V., and Orlov, Y.L. (2017). Genomic landscape of CpG rich elements in human. BMC Evol. Biol. *17*, 19–19.

9. Lister, R., Pelizzola, M., Dowen, R.H., David Hawkins, R., Hon, G., Tonti-Filippini, J., Nery, J.R., Lee, L., Ye, Z., Ngo, Q.-M., et al. (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. Nature *462*, 315.

10. Widschwendter, M., Jiang, G., Woods, C., Müller, H.M., Fiegl, H., Goebel, G., Marth, C., Müller-Holzner, E., Zeimet, A.G., Laird, P.W., et al. (2004). DNA hypomethylation and ovarian cancer biology. Cancer Res. *64*. https://doi.org/10.1158/0008-5472.CAN-04-0238.

11. Gama-Sosa, M.A., Slagel, V.A., Trewyn, R.W., Oxenhandler, R., Kuo, K.C., Gehrke, C.W., and Ehrlich, M. (1983). The 5-methylcytosine content of DNA from human tumors. Nucleic Acids Res. *11*, 6883.

12. Adorján, P., Distler, J., Lipscher, E., Model, F., Müller, J., Pelet, C., Braun, A., Florl, A.R., Gütig, D., Grabs, G., et al. (2002). Tumour class prediction and discovery by microarray-based DNA methylation analysis. Nucleic Acids Res. *30*, e21.

13. De Smet, C., De Backer, O., Faraoni, I., Lurquin, C., Brasseur, F., and Boon, T. (1996). The activation of human gene MAGE-1 in tumor cells is correlated with genome-wide demethylation. Proc. Natl. Acad. Sci. U. S. A. *93*, 7149.

14. Paul, D.S., Teschendorff, A.E., Dang, M.A.N., Lowe, R., Hawa, M.I., Ecker, S., Beyan, H., Cunningham, S., Fouts, A.R., Ramelius, A., et al. (2016). Increased DNA methylation variability in type 1 diabetes across three immune effector cell types. Nat. Commun. *7*. https://doi.org/10.1038/ncomms13555.

15. Tahara, T., Hirata, I., Nakano, N., Nagasaka, M., Nakagawa, Y., Shibata, T., and Ohmiya, N. (2017). Comprehensive DNA Methylation Profiling of Inflammatory Mucosa in Ulcerative Colitis. Inflamm. Bowel Dis. *23*. https://doi.org/10.1097/MIB.0000000000000990.

16. Hall, E., Dayeh, T., Kirkpatrick, C.L., Wollheim, C.B., Dekker, N.M., and Ling, C. (2013). DNA methylation of the glucagon-like peptide 1 receptor (GLP1R) in human pancreatic islets. BMC Med. Genet. *14*. https://doi.org/10.1186/1471-2350-14-76.

17. Yang, B.T., Dayeh, T.A., Volkov, P.A., Kirkpatrick, C.L., Malmgren, S., Jing, X., Renström, E., Wollheim, C.B., Nitert, and Ling, C. (2012). Increased DNA methylation and decreased expression of PDX-1 in pancreatic islets from patients with type 2 diabetes. Mol. Endocrinol. *26*. https://doi.org/10.1210/me.2012-1004.

18. Liu, Y., Aryee, M.J., Padyukov, L., Fallin, M.D., Hesselberg, E., Runarsson, A., Reinius, L., Acevedo, N., Taub, M., Ronninger, M., et al. (2013). Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. Nat. Biotechnol. *31*, 142–147.

19. Dekkers, K.F., van Iterson, M., Slieker, R.C., Moed, M.H., Bonder, M.J., van Galen, M., Mei, H., Zhernakova, D.V., van den Berg, L.H., Deelen, J., et al. (2016). Blood lipids influence DNA methylation in circulating cells. Genome Biol. *17*. https://doi.org/10.1186/s13059-016-1000-6.

20. Watson, C.J., Horgan, S., Neary, R., Glezeva, N., Tea, I., Corrigan, N., McDonald, K., Ledwidge, M., and Baugh, J. (2016). Epigenetic Therapy for the Treatment of Hypertension-Induced Cardiac Hypertrophy and Fibrosis. J. Cardiovasc. Pharmacol. Ther. *21*. https://doi.org/10.1177/1074248415591698.

21. Zaina, S., Heyn, H., Carmona, F.J., Varol, N., Sayols, S., Condom, E., Ramírez-Ruz, J., Gomez, A., Gonçalves, I., Moran, S., et al. (2014). DNA methylation map of human atherosclerosis. Circ. Cardiovasc. Genet. *7*. https://doi.org/10.1161/CIRCGENETICS.113.000441.

22. Horvath, S., and Raj, K. (2018). DNA methylation-based biomarkers and the epigenetic clock theory of ageing. Nat. Rev. Genet. *19*, 371–384.

23. Rakyan, V.K., Down, T.A., Maslau, S., Andrew, T., Yang, T.-P., Beyan, H., Whittaker, P., McCann, O.T., Finer, S., Valdes, A.M., et al. (2010). Human aging-associated DNA hypermethylation occurs preferentially at bivalent chromatin domains. Genome Res. *20*, 434.

24. Yuan, T., Jiao, Y., de Jong, S., Ophoff, R.A., Beck, S., and Teschendorff, A.E. (2015). An Integrative Multi-scale Analysis of the Dynamic DNA Methylation Landscape in Aging. PLoS Genet. *11*. https://doi.org/10.1371/journal.pgen.1004996.

25. Lu, A.T., Fei, Z., Haghani, A., Robeck, T.R., Zoller, J.A., Li, C.Z., Lowe, R., Yan, Q., Zhang, J., Vu, H., et al. (2023). Universal DNA methylation age across mammalian tissues. Nature Aging *3*, 1144–1166.

26. Christensen, B.C., Andres Houseman, E., Marsit, C.J., Zheng, S., Wrensch, M.R., Wiemels, J.L., Nelson, H.H., Karagas, M.R., Padbury, J.F., Bueno, R., et al. (2009). Aging and Environmental Exposures Alter Tissue-Specific DNA Methylation Dependent upon CpG Island Context. PLoS Genet. *5*, e1000602.

27. Horvath, S. (2013). DNA methylation age of human tissues and cell types. Genome Biol. *14*, 1–20.

28. Beynon, R.A., Ingle, S.M., Langdon, R., May, M., Ness, A., Martin, R.M., Suderman, M., Ingarfield, K., Marioni, R.E., McCartney, D.L., et al. (2022). Epigenetic biomarkers of ageing are predictive of mortality risk in a longitudinal clinical cohort of individuals diagnosed with oropharyngeal cancer. Clin. Epigenetics *14*, 1–13.

29. McCrory, C., Fiorito, G., Hernandez, B., Polidoro, S., O'Halloran, A.M., Hever, A., Ni Cheallaigh, C., Lu, A.T., Horvath, S., Vineis, P., et al. (2020). GrimAge Outperforms Other Epigenetic Clocks in the Prediction of Age-Related Clinical Phenotypes and All-Cause Mortality. J. Gerontol. A Biol. Sci. Med. Sci. *76*, 741–749.

30. Davegårdh, C., Hall Wedin, E., Broholm, C., Henriksen, T.I., Pedersen, M., Pedersen, B.K., Scheele, C., and Ling, C. (2019). Sex influences DNA methylation and gene expression in human skeletal muscle myoblasts and myotubes. Stem Cell Res. Ther. *10*, 1–17.

31. Liu, J., Morgan, M., Hutchison, K., and Calhoun, V.D. (2010). A Study of the Influence of Sex on Genome Wide Methylation. PLoS One *5*, e10028.

32. Singmann, P., Shem-Tov, D., Wahl, S., Grallert, H., Fiorito, G., Shin, S.-Y., Schramm, K., Wolf, P., Kunze, S., Baran, Y., et al. (2015). Characterization of whole-genome autosomal differences of DNA methylation between men and women. Epigenetics Chromatin *8*, 1–13.

33. Grant, O.A., Wang, Y., Kumari, M., Zabet, N.R., and Schalkwyk, L. (2022). Characterising sex differences of autosomal DNA methylation in whole blood using the Illumina EPIC array. Clin. Epigenetics *14*, 1–16.

34. Husquin, L.T., Rotival, M., Fagny, M., Quach, H., Zidane, N., McEwen, L.M., MacIsaac, J.L., Kobor, M.S., Aschard, H., Patin, E., et al. (2018). Exploring the genetic basis of human population differences in DNA methylation and their causal impact on immune gene regulation. Genome Biol. *19*. https://doi.org/10.1186/s13059-018-1601-3.

35. Galanter, J.M., Gignoux, C.R., Oh, S.S., Torgerson, D., Pino-Yanes, M., Thakur, N., Eng, C., Hu, D.,

Huntsman, S., Farber, H.J., et al. (2017). Differential methylation between ethnic sub-groups reflects the effect of genetic ancestry and environmental exposures. https://doi.org/10.7554/eLife.20532.

36. Giri, A.K., Bharadwaj, S., Banerjee, P., Chakraborty, S., Parekatt, V., Rajashekar, D., Tomar, A., Ravindran, A., Basu, A., Tandon, N., et al. (2017). DNA methylation profiling reveals the presence of population-specific signatures correlating with phenotypic characteristics. Mol. Genet. Genomics *292*, 655–662.

37. Bell, J.T., Pai, A.A., Pickrell, J.K., Gaffney, D.J., Pique-Regi, R., Degner, J.F., Gilad, Y., and Pritchard, J.K. (2011). DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. Genome Biol. *12*. https://doi.org/10.1186/gb-2011-12-1-r10.

38. Loyfer, N., Magenheim, J., Peretz, A., Cann, G., Bredno, J., Klochendler, A., Fox-Fisher, I., Shabi-Porat, S., Hecht, M., Pelet, T., et al. (2023). A DNA methylation atlas of normal human cell types. Nature *613*, 355–364.

39. Dozmorov, M.G. (2015). Polycomb repressive complex 2 epigenomic signature defines age-associated hypermethylation and gene expression changes. Epigenetics *10*, 484.

40. Hannum, G., Guinney, J., Zhao, L., Zhang, L., Hughes, G., Sadda, S., Klotzle, B., Bibikova, M., Fan, J.-B., Gao, Y., et al. (2013). Genome-wide Methylation Profiles Reveal Quantitative Views of Human Aging Rates. Mol. Cell *49*, 359.

41. Bergstedt, J., Azzou, S.A.K., Tsuo, K., Jaquaniello, A., Urrutia, A., Rotival, M., Lin, D.T.S., MacIsaac, J.L., Kobor, M.S., Albert, M.L., et al. (2022). The immune factors driving DNA methylation variation in human blood. Nat. Commun. *13*. https://doi.org/10.1038/s41467-022-33511-6.

42. Byun, H.-M., Siegmund, K.D., Pan, F., Weisenberger, D.J., Kanel, G., Laird, P.W., and Yang, A.S. (2009). Epigenetic profiling of somatic tissues from human autopsy specimens identifies tissue- and individual-specific DNA methylation patterns. Hum. Mol. Genet. *18*, 4808.

43. Oliva, M., Demanelis, K., Lu, Y., Chernoff, M., Jasmine, F., Ahsan, H., Kibriya, M.G., Chen, L.S., and Pierce, B.L. (2023). DNA methylation QTL mapping across diverse human tissues provides molecular links between genetic variation and complex traits. Nat. Genet. *55*, 112.

44. Melé, M., Ferreira, P.G., Reverter, F., DeLuca, D.S., Monlong, J., Sammeth, M., Young, T.R., Goldmann, J.M., Pervouchine, D.D., Sullivan, T.J., et al. (2015). The human transcriptome across tissues and individuals. Science. https://doi.org/10.1126/science.aaa0355.

45. Dor, Y., and Cedar, H. (2018). Principles of DNA methylation and their implications for biology and medicine. Lancet *392*. https://doi.org/10.1016/S0140-6736(18)31268-6.

46. Vogel, T.U., Evans, D.T., Urvater, J.A., O'Connor, D.H., Hughes, A.L., and Watkins, D.I. (1999). Major histocompatibility complex class I genes in primates: co-evolution with pathogens. Immunol. Rev. *167*. https://doi.org/10.1111/j.1600-065x.1999.tb01402.x.

47. Rakyan, V.K., Hildmann, T., Novik, K.L., Lewin, J., Tost, J., Cox, A.V., Dan Andrews, T., Howe, K.L., Otto, T., Olek, A., et al. (2004). DNA Methylation Profiling of the Human Major Histocompatibility Complex: A Pilot Study for the Human Epigenome Project. PLoS Biol. *2*, e405.

48. Wahl, S., Drong, A., Lehne, B., Loh, M., Scott, W.R., Kunze, S., Tsai, P.C., Ried, J.S., Zhang, W., Yang, Y., et al. (2017). Epigenome-wide association study of body mass index, and the adverse outcomes of adiposity. Nature *541*. https://doi.org/10.1038/nature20784.

49. Dick, K.J., Nelson, C.P., Tsaprouni, L., Sandling, J.K., Aïssi, D., Wahl, S., Meduri, E., Morange, P.E., Gagnon, F., Grallert, H., et al. (2014). DNA methylation and body-mass index: a genome-wide analysis. Lancet *383*. https://doi.org/10.1016/S0140-6736(13)62674-4.

50. Maher, A.C., Fu, M.H., Isfort, R.J., Varbanov, A.R., Qu, X.A., and Tarnopolsky, M.A. (2009). Sex differences in global mRNA content of human skeletal muscle. PLoS One *4*. https://doi.org/10.1371/journal.pone.0006335.

51. Landen, S., Voisin, S., Craig, J.M., McGee, S.L., Lamon, S., and Eynon, N. (2019). Genetic and epigenetic sex-specific adaptations to endurance exercise. Epigenetics *14*. https://doi.org/10.1080/15592294.2019.1603961.

52. Landen, S., Jacques, M., Hiam, D., Alvarez-Romero, J., Harvey, N.R., Haupt, L.M., Griffiths, L.R., Ashton, K.J., Lamon, S., Voisin, S., et al. (2021). Skeletal muscle methylome and transcriptome integration reveals profound sex differences related to muscle function and substrate metabolism. Clin. Epigenetics *13*, 1–20.

53. Chu, S.-K., and Yang, H.-C. (2017). Interethnic DNA Methylation Difference and Its Implications in Pharmacoepigenetics. Epigenomics. https://doi.org/10.2217/epi-2017-0046.

54. Ramos, E., Doumatey, A., Elkahloun, A.G., Shriner, D., Huang, H., Chen, G., Zhou, J., McLeod, H., Adeyemo, A., and Rotimi, C.N. (2014). Pharmacogenomics, ancestry and clinical decision making for global populations. Pharmacogenomics J. *14*. https://doi.org/10.1038/tpj.2013.24.

55. García-Pérez, R., Ramirez, J.M., Ripoll-Cladellas, A., Chazarra-Gil, R., Oliveros, W., Soldatkina, O., Bosio, M., Rognon, P.J., Capella-Gutierrez, S., Calvo, M., et al. (2022). The landscape of expression and alternative splicing variation across human traits. Cell genomics *3*. https://doi.org/10.1016/j.xgen.2022.100244.

56. Schaid, D.J., and Sinnwell, J.P. (2020). Penalized Models for Analysis of Multiple Mediators. Genet. Epidemiol. *44*, 408.

57. Slieker, R.C., van Iterson, M., Luijk, R., Beekman, M., Zhernakova, D.V., Moed, M.H., Mei, H., van Galen, M., Deelen, P., Bonder, M.J., et al. (2016). Age-related accrual of methylomic variability is linked to fundamental ageing mechanisms. Genome Biol. *17*, 1–13.

58. A global reference for human genetic variation (2015). Nature *526*, 68–74.

59. Feil, R., and Fraga, M.F. (2012). Epigenetics and the environment: emerging patterns and implications. Nat. Rev. Genet. *13*. https://doi.org/10.1038/nrg3142.

60. Aracena, K.A., Lin, Y.-L., Luo, K., Pacis, A., Gona, S., Mu, Z., Yotova, V., Sindeaux, R., Pramatarova, A., Simon, M.-M., et al. (2024). Epigenetic variation impacts individual differences in the transcriptional response to influenza infection. Nat. Genet. *56*, 408–419.

61. Tissue-specific contribution of macrophages to wound healing (2017). Semin. Cell Dev. Biol. *61*, 3–11.

62. Koh, T.J., and DiPietro, L.A. Inflammation and wound healing: The role of the macrophage. Expert Rev. Mol. Med. *13*, e23.

63. Mosser, D.M., and Edwards, J.P. (2008). Exploring the full spectrum of macrophage activation. Nat. Rev. Immunol. *8*, 958–969.

64. Chen, W., Lu, C., Hirota, C., Iacucci, M., Ghosh, S., and Gui, X. (2016). Smooth Muscle Hyperplasia/Hypertrophy is the Most Prominent Histological Change in Crohn's Fibrostenosing Bowel Strictures: A Semiquantitative Analysis by Using a Novel Histological Grading Scheme. J. Crohns. Colitis *11*, 92–104.

65. Randolph, H.E., Fiege, J.K., Thielen, B.K., Mickelson, C.K., Shiratori, M., Barroso-Batista, J., Langlois, R.A., and Barreiro, L.B. (2021). Genetic ancestry effects on the response to viral infection are pervasive but cell type-specific. Science *374*, 1127.

66. Liu, Y., Zhou, Q., Zhong, L., Lin, H., Hu, M.-M., Zhou, Y., Shu, H.-B., and Li, S. (2018). ZDHHC11 modulates innate immune response to DNA virus by mediating MITA–IRF3 association. Cell. Mol. Immunol. *15*, 907–916.

67. Liu, E., Sun, J., Yang, J., Li, L., Yang, Q., Zeng, J., Zhang, J., Chen, D., and Sun, Q. (2021). ZDHHC11 Positively Regulates NF-κB Activation by Enhancing TRAF6 Oligomerization. Frontiers in Cell and Developmental Biology *9*. https://doi.org/10.3389/fcell.2021.710967.

68. Rizzardi, L.F., Hickey, P.F., Idrizi, A., Tryggvadóttir, R., Callahan, C.M., Stephens, K.E., Taverna, S.D., Zhang, H., Ramazanoglu, S., Hansen, K.D., et al. (2021). Human brain region-specific variably methylated regions are enriched for heritability of distinct neuropsychiatric traits. Genome Biology *22*, 1–26.

69. Oki, S., Ohta, T., Shioi, G., Hatanaka, H., Ogasawara, O., Okuda, Y., Kawaji, H., Nakaki, R., Sese, J., and Meno, C. (2018). ChIP-Atlas: a data-mining suite powered by full integration of public ChIP-seq data. EMBO Rep. https://doi.org/10.15252/embr.201846255.

70. Migeon, B.R. (2019). The Non-random Location of Autosomal Genes That Participate in X Inactivation. Frontiers in Cell and Developmental Biology *7*. https://doi.org/10.3389/fcell.2019.00144.

71. Chu, C., Zhang, Q.C., da Rocha, S.T., Flynn, R.A., Bharadwaj, M., Calabrese, J.M., Magnuson, T., Heard, E., and Chang, H.Y. (2015). Systematic discovery of Xist RNA binding proteins. Cell *161*, 404–416.

72. Pullirsch, D., Härtel, R., Kishimoto, H., Leeb, M., Steiner, G., and Wutz, A. (2010). The Trithorax group protein Ash2l and Saf-A are recruited to the inactive X chromosome at the onset of stable X inactivation. Development *137*, 935–943.

73. Nugent, B.M., O'Donnell, C.M., Epperson, C.N., and Bale, T.L. (2018). Placental H3K27me3 establishes female resilience to prenatal insults. Nat. Commun. *9*. https://doi.org/10.1038/s41467-018-04992-1.

74. Seale, K., Horvath, S., Teschendorff, A., Eynon, N., and Voisin, S. (2022). Making sense of the ageing methylome. Nat. Rev. Genet. *23*. https://doi.org/10.1038/s41576-022-00477-6.

75. Jain, N., Li, J.L., Tong, L., Jasmine, F., Kibriya, M.G., Demanelis, K., Oliva, M., Chen, L.S., and Pierce, B.L. (2024). DNA methylation correlates of chronological age in diverse human tissue types. Epigenetics Chromatin *17*, 25.

76. Yusipov, I., Bacalini, M.G., Kalyakulina, A., Krivonosov, M., Pirazzini, C., Gensous, N., Ravaioli, F., Milazzo, M., Giuliani, C., Vedunova, M., et al. (2020). Age-related DNA methylation changes are sex-specific: a comprehensive assessment. Aging *12*, 24057.

77. McCartney, D.L., Zhang, F., Hillary, R.F., Zhang, Q., Stevenson, A.J., Walker, R.M., Bermingham, M.L., Boutin, T., Morris, S.W., Campbell, A., et al. (2019). An epigenome-wide association study of sex-specific chronological ageing. Genome Med. *12*, 1–11.

78. Boix, C.A., James, B.T., Park, Y.P., Meuleman, W., and Kellis, M. (2021). Regulatory genomic circuitry of human disease loci by integrative epigenomics. Nature *590*, 300–307.

79. Quach, H., Rotival, M., Pothlichet, J., Loh, Y.E., Dannemann, M., Zidane, N., Laval, G., Patin, E., Harmant, C., Lopez, M., et al. (2016). Genetic Adaptation and Neandertal Admixture Shaped the Immune System of Human Populations. Cell *167*. https://doi.org/10.1016/j.cell.2016.09.024.

80. Pennington, R., Gatenbee, C., Kennedy, B., Harpending, H., and Cochran, G. (2009). Group differences in proneness to inflammation. Infect. Genet. Evol. *9*. https://doi.org/10.1016/j.meegid.2009.09.017.

81. Augui, S., Nora, E.P., and Heard, E. (2011). Regulation of X-chromosome inactivation by the X-inactivation centre. Nat. Rev. Genet. *12*, 429–442.

82. Gendrel, A.-V., Apedaile, A., Coker, H., Termanis, A., Zvetkova, I., Godwin, J., Tang, Y.A., Huntley, D., Montana, G., Taylor, S., et al. (2012). Smchd1-dependent and -independent pathways determine developmental dynamics of CpG island methylation on the inactive X chromosome. Dev. Cell *23*, 265–279.

83. Oliva, M., Muñoz-Aguirre, M., Kim-Hellmuth, S., Wucher, V., Gewirtz, A.D.H., Cotter, D.J., Parsana, P., Kasela, S., Balliu, B., Viñuela, A., et al. (2020). The impact of sex on gene expression across human tissues. Science *369*. https://doi.org/10.1126/science.aba3066.

84. Lau-Corona, D., Bae, W.K., Hennighausen, L., and Waxman, D.J. (2020). Sex-biased genetic programs in liver metabolism and liver fibrosis are controlled by EZH1 and EZH2. PLoS Genet. *16*. https://doi.org/10.1371/journal.pgen.1008796.

85. O'Hagan, H.M. (2014). Chromatin Modifications during Repair of Environmental Exposure-Induced DNA Damage: A Potential Mechanism for Stable Epigenetic Alterations. Environ. Mol. Mutagen. *55*, 278.

86. Mu, W., Starmer, J., Fedoriw, A.M., Yee, D., and Magnuson, T. (2014). Repression of the soma-specific transcriptome by Polycomb-repressive complex 2 promotes male germ cell development. Genes Dev. *28*, 2056.

87. Wang, Y., Wang, W., Cheng, K., Geng, K., Liang, J., Wang, P., Zhang, J., Niu, S., Jia, L., Zhang, S., et al. (2022). Polycomb subunit Pcgf2 mediates ovulation and fertility through transcriptional regulation progesterone receptor. Frontiers in cell and developmental biology *10*. https://doi.org/10.3389/fcell.2022.1010601.

88. Iron overload increases the sensitivity of endometriosis stromal cells to ferroptosis via a PRC2-independent function of EZH2 (2024). Int. J. Biochem. Cell Biol. *169*, 106553.

89. Oakes, C.C., La Salle, S., Smiraglia, D.J., Robaire, B., and Trasler, J.M. (2007). A unique configuration of genome-wide DNA methylation patterns in the testis. Proc. Natl. Acad. Sci. U. S. A. *104*. https://doi.org/10.1073/pnas.0607521104.

90. Knight, A.K., Spencer, J.B., and Smith, A.K. (2024). DNA methylation as a window into female reproductive aging. Epigenomics *16*. https://doi.org/10.2217/epi-2023-0298.

91. Uysal, F., and Ozturk, S. (2020). The loss of global DNA methylation due to decreased DNMT expression in the postnatal mouse ovaries may associate with infertility emerging during ovarian aging. Histochem. Cell Biol. *154*. https://doi.org/10.1007/s00418-020-01890-w.

92. Sukur, G., Uysal, F., and Cinar, O. (2023). High-fat diet induced obesity alters Dnmt1 and Dnmt3a levels and global DNA methylation in mouse ovary and testis. Histochem. Cell Biol. *159*. https://doi.org/10.1007/s00418-022-02173-2.

93. The GTEx Consortium (2020). The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science *369*, 1318.

94. Alexander, D.H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. Genome Res. *19*, 1655.

95. limma Bioconductor. http://bioconductor.org/packages/limma/.

96. Du, P., Zhang, X., Huang, C.-C., Jafari, N., Kibbe, W.A., Hou, L., and Lin, S.M. (2010). Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. BMC Bioinformatics *11*, 1–9.

97. Ramirez, J.M., Ribeiro, R., Soldatkina, O., Moraes, A., García-Pérez, R., Ferreira, P.G., and Melé, M. (2024). The molecular impact of cigarette smoking resembles aging across tissues. bioRxiv. https://doi.org/10.1101/2024.03.14.585016.

98. missMethyl Bioconductor. http://bioconductor.org/packages/missMethyl/.

99. Kim-Hellmuth, S., Aguet, F., Oliva, M., Muñoz-Aguirre, M., Kasela, S., Wucher, V., Castel, S.E., Hamel, A.R., Viñuela, A., Roberts, A.L., et al. (2020). Cell type specific genetic regulation of gene expression across human tissues. Science (New York, N.Y.) *369*, eaaz8528.

100. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans (2015). Science (New York, N.Y.) *348*. https://doi.org/10.1126/science.1262110.

101. Ernst, J., and Kellis, M. (2012). ChromHMM: automating chromatin-state discovery and characterization. Nat. Methods *9*, 215–216.

102. Roadmap Epigenomics Consortium, Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Kheradpour, P., Zhang, Z., Heravi-Moussavi, A., Liu, Y., et al. (2015). Integrative analysis of 111 reference human epigenomes. Nature *518*, 317.

103. Jaffe, A.E., Murakami, P., Lee, H., Leek, J.T., Fallin, Feinberg, A.P., and Irizarry, R.A. (2012). Bump hunting to identify differentially methylated regions in epigenetic epidemiology studies. Int. J. Epidemiol. *41*. https://doi.org/10.1093/ije/dyr238.

104. Hoffman, G.E., and Schadt, E.E. (2016). variancePartition: interpreting drivers of variation in complex gene expression studies. BMC Bioinformatics *17*, 1–13.

105. Kuhn, M. (2008). Building Predictive Models in R Using the caret Package. J. Stat. Softw. *28*, 1–26.

106. Schaid, D.J., Dikilitas, O., Sinnwell, J.P., and Kullo, I.J. (2022). Penalized mediation models for multivariate data. Genet. Epidemiol. *46*. https://doi.org/10.1002/gepi.22433.

107. Babraham Bioinformatics - Trim Galore! http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/.

108. Krueger, F., and Andrews, S.R. (2011). Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics (Oxford, England) *27*. https://doi.org/10.1093/bioinformatics/btr167.

109. Hansen, K.D. (2024). The bsseq User's Guide. https://www.bioconductor.org/packages/devel/bioc/vignettes/bsseq/inst/doc/bsseq.html.

110. Hansen, K.D., Langmead, B., and Irizarry, R.A. (2012). BSmooth: from whole genome bisulfite sequencing reads to differentially methylated regions. Genome Biology *13*, 1–10.

List of supplementary tables:

Supplementary Table 1: GO enrichment results for A) tissue-variable CpGs; B) tissue-variable CpGs per genomic location; C) individual-variable CpGs; and D) individual-variable CpGs per genomic location.

Supplementary Table 2: GO enrichment results for correlated DEGs correlated with DMPs in the ovary with age.

Supplementary Table 3: Correlation analysis between DMPs and DEGs per tissue and trait. Significantly correlated DMP–DEG pairs are included.

Supplementary Table 4: Mediation results. Table depicting DEGs and the number of *cis*-associated DMPs that significantly mediate gene expression changes.

Supplementary Table 5: GO enrichment results for ancestry-related DMPs located in enhancers.

Supplementary Table 6: GO enrichment results for sex-related DMPs. Hyper: Higher methylation in females, Hypo: Higher methylation in males.

Supplementary Table 7: Enrichment of TFBSs in sex-DMPs A) per tissue and B) tissue-shared (DMPs shared between 2 or more tissues). Hyper: Higher methylation in females, Hypo: Higher methylation in males.

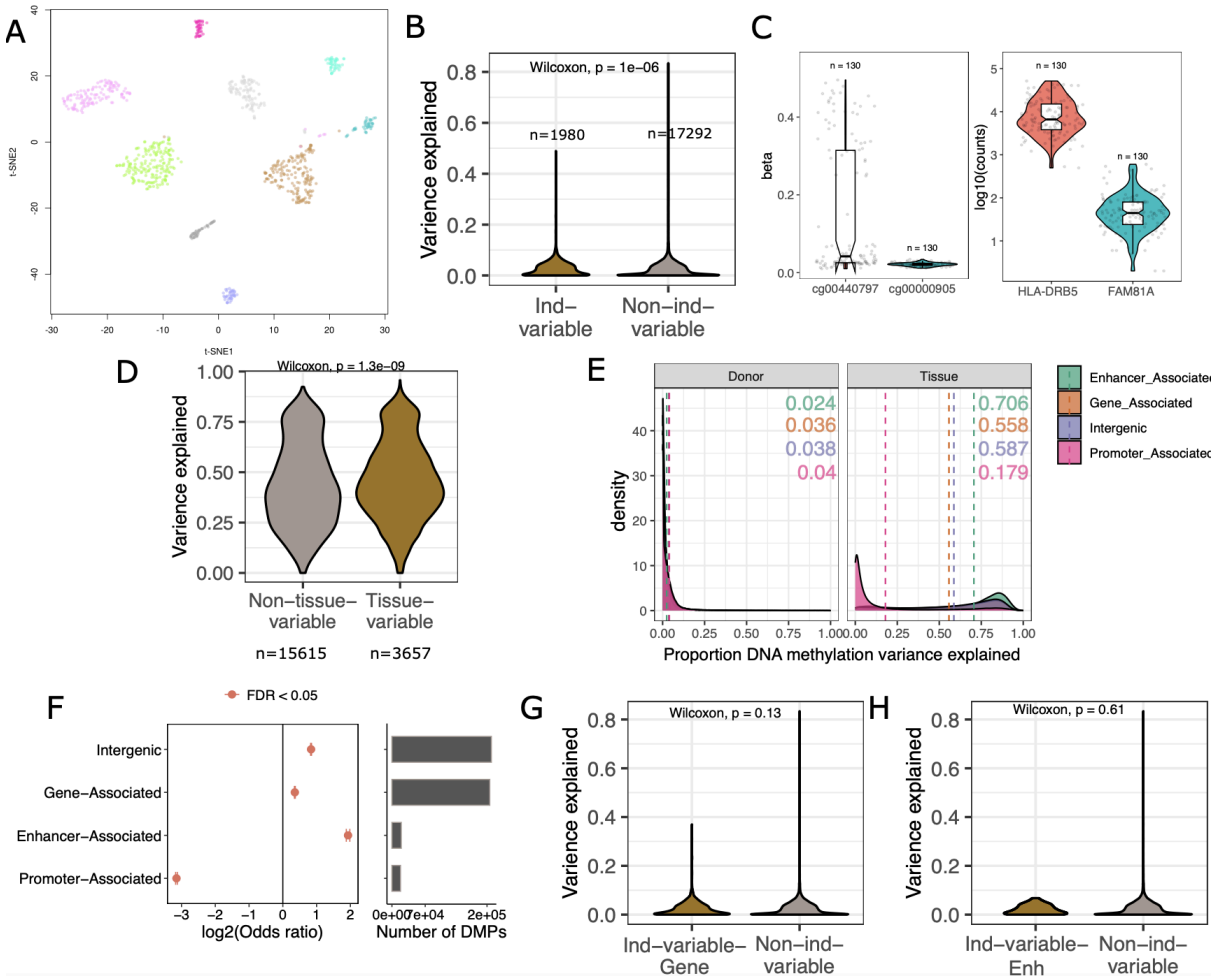
Supplementary Table 8: GO enrichment results for the TFs enriched among hypermethylated sex-DMPs shared between 2 or more tissues.

Supplementary Table 9: Enrichment of TFBSs in age-DMPs per tissue and direction of methylation change. Hyper: Increased methylation in older individuals; Hypo: Increased methylation in young individuals.

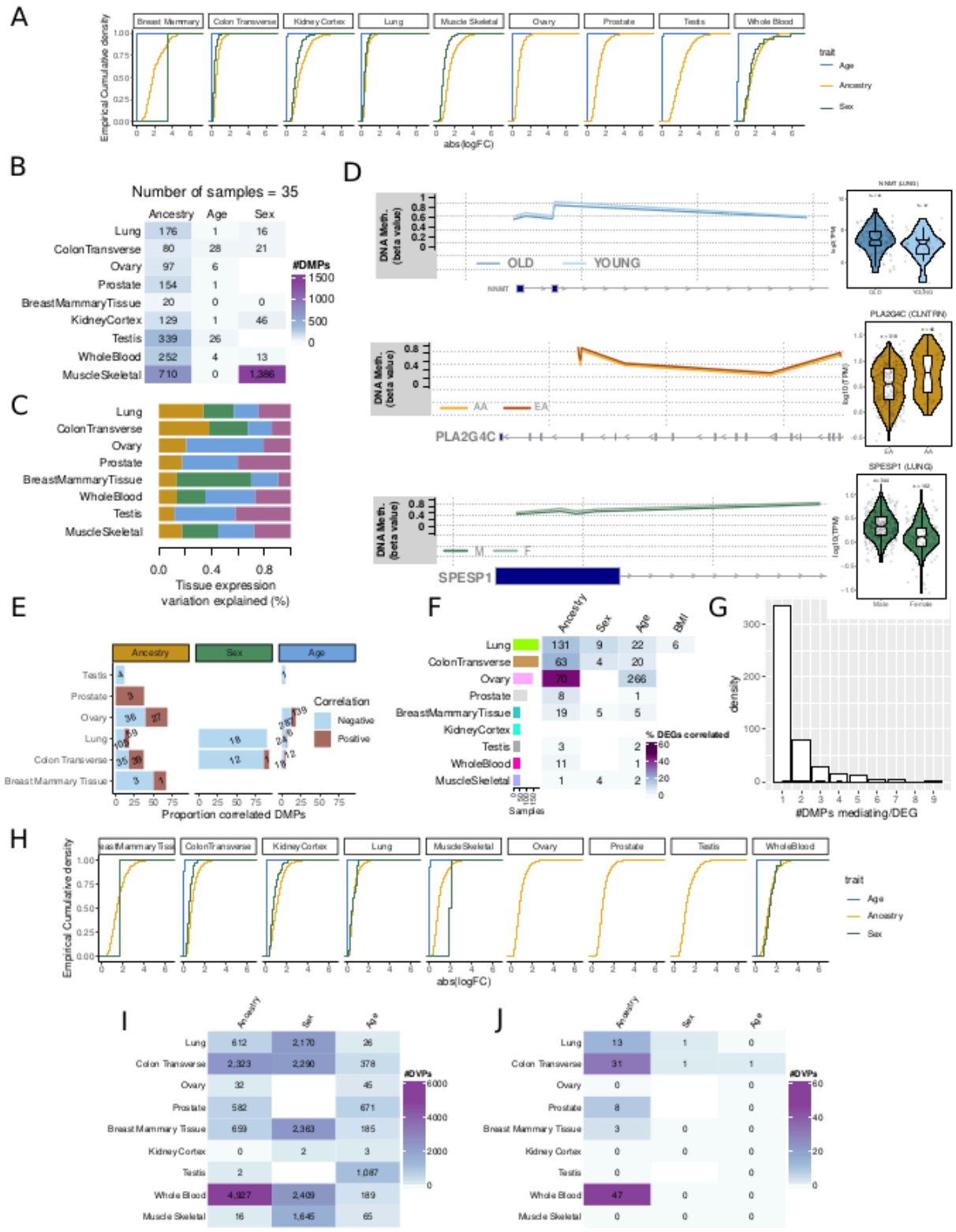
Supplementary Table 10: GO enrichment results for age-related DMPs per A) tissue and B) tissue and genomic location. Hyper: Increased methylation in older individuals; Hypo: Increased methylation in young individuals.

Supplementary Table 11: GO enrichment results for tissue-shared age- hypermethylated DMPs (more methylated in older individuals in 3 or more tissues).

Supplementary Figures:



<u>Figure S1.</u> Overview of DNA methylation across tissues and individuals. A. t-SNE showing the clustering of the DNA methylation data by tissue. **B.** Proportion of gene expression variation explained for genes individually-variable (brown) and not individually-variable (gray) at the DNA methylation level. **C.** Examples of 1 individually-variable CpG (cg00440797) and one non-individually variable CpG (cg00000905), and their corresponding genes (HLA-DRB5 and FAM81A, respectively). **Left.** Beta values for each CpG in lung. **Right.** log10(counts) for each gene in lung. The number of individuals per group are indicated. **D**.Proportion of gene expression variation explained for genes tissue-variable (brown) and not tissue-variable (gray) at the DNA methylation level. **E.** Proportion of DNA methylation variation explained by individual (Left) and tissue (Right). The distributions are clustered by the genomic location of the CpGs. Median numbers for each genomic location are shown in each facet. **F. Left.** Enrichment of each genomic location on tissue variable CpGs. X-axis represents the log2(OR) and dots are colored according to the significance of the enrichment (two-tailed Fisher's exact test). Confidence interval is shown. **Right.** Number of tissue variable CpGs associated with each genomic location. G. Proportion of gene expression variation explained for genes individually-variable (brown) and not individually-variable (gray) at the DNA methylation level of their gene bodies. H. Proportion of gene expression variation explained for genes individually-variable (brown) and not individually-variable (gray) at the DNA methylation level of their enhancers.



<u>Figure S2:</u> DNA methylation variation across tissues and demographic traits. A. Effect sizes of DMPs per demographic trait. **B.** Mean number of DMPs per tissue and demographic trait for 50 random permutations when downsampling to an equal number of samples across tissues=35 . **C.** Proportion of total tissue gene expression variation explained by each demographic trait. **D.** Examples of correlated DEGs with at least one DMP. log10(TPM) values for gene expression are shown in violin plots. Beta values for DNA methylation of correlated DMPs are plotted according to their genomic location. The gene structure is shown below each plot. **E.** Proportion and number of DMPs correlated with DEGs per tissue and demographic trait., colored by the direction of correlation. **F.** Number of DEGs correlated with at least one DMP, DMPs selected

based on nominal p.value < 0.05. Cells are colored by the proportion of DEGs correlated out of the total number of DEG with a probe. **G.** Number of DMPs mediating gene expression changes per each DEG. **H.** Effect sizes of DVPs per demographic trait and tissue. **I.** Number of DVG per trait and tissue. **J.** Number of DVG per trait and tissue that have an associated DVP.

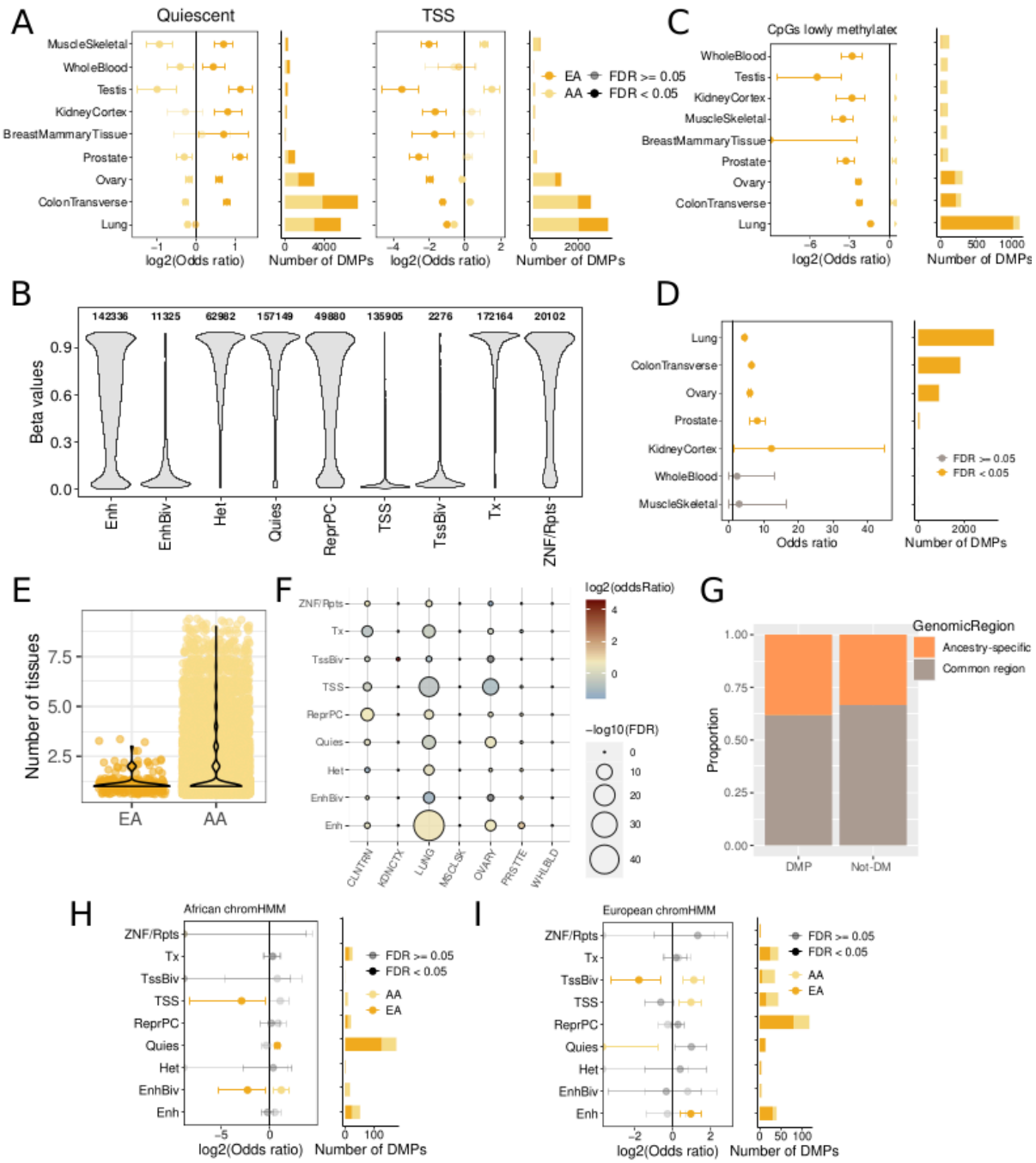


Figure S3. Ancestry-associated DNA methylation changes. A. Ancestry-DMPs enrichment at 2 chromatin states across tissues: quiescent and transcription start sites. Left. Enrichment and confidence interval of each tissue-DMPs. Dots are colored according to the direction of DNA methylation change and shaded according to the significance of the enrichment (two-tailed Fisher's exact test). Right. Number of DMPs within each chromatin state per tissue. **B.** Representation of the different chromatin states analyzed throughout the paper and their associated DNA methylation status. **C.** Enrichment of ancestry-DMPs in CpGs not methylated in the genome. Legend is shared with panel **A**. **D.** Enrichment of Ancestry-DMPs in mQTLs across tissues. Left. Enrichment and confidence interval of each tissue-DMPs. Dots are colored according to the significance of the

enrichment (two-tailed Fisher's exact test). Right. Number of DMPs with mQTLs within each tissue. **E.** Number of tissues where ancestry-DVPs are found. **F.** Enrichment of cis-driven Ancestry-DMPs at different chromatin states across tissues. **G.** Proportion of DMPs and CpGs not DM (Not DM) overlapping Ancestry-specific genomic regions (CpGs located in a chromatin region differentially annotated between population by chromHMM). **H.** Enrichment of Ancestry-DMPs in an African-descent individual at multiple chromatin states (see Methods). **I.** Enrichment of Ancestry-DMPs in a European-descent individual at multiple chromatin states (see Methods).

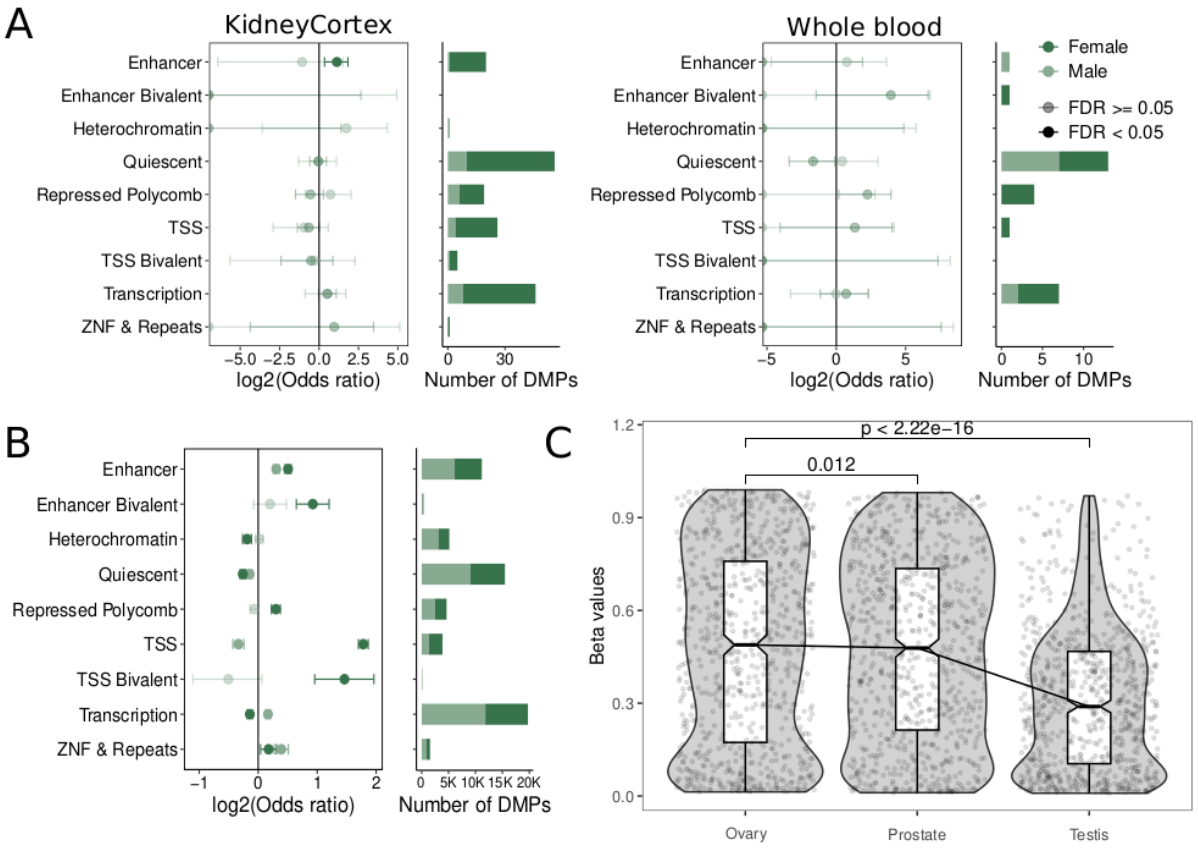


Figure S4. Sex-associated DNA methylation changes. A. Sex-DMPs enrichment at different chromatin states across tissues. Left. Enrichment and confidence interval of each tissue-DMPs. Dots are colored according to the direction of DNA methylation change and shaded according to the significance of the enrichment (two-tailed Fisher's exact test). Right. Number of DMPs within each chromatin state per tissue. **B.** Sex-DMRs enrichment at different chromatin states in Lung. Enrichment and confidence interval of each tissue-DMPs. Dots are colored according to the direction of DNA methylation change and shaded according to the significance of the enrichment (two-tailed Fisher's exact test). Right. Number of DMPs within each chromatin state per tissue. **C.** DNA methylation status (Beta values) of shared DMPs in the sex-related tissues. P-values correspond to Wilcoxon-test with paired data.

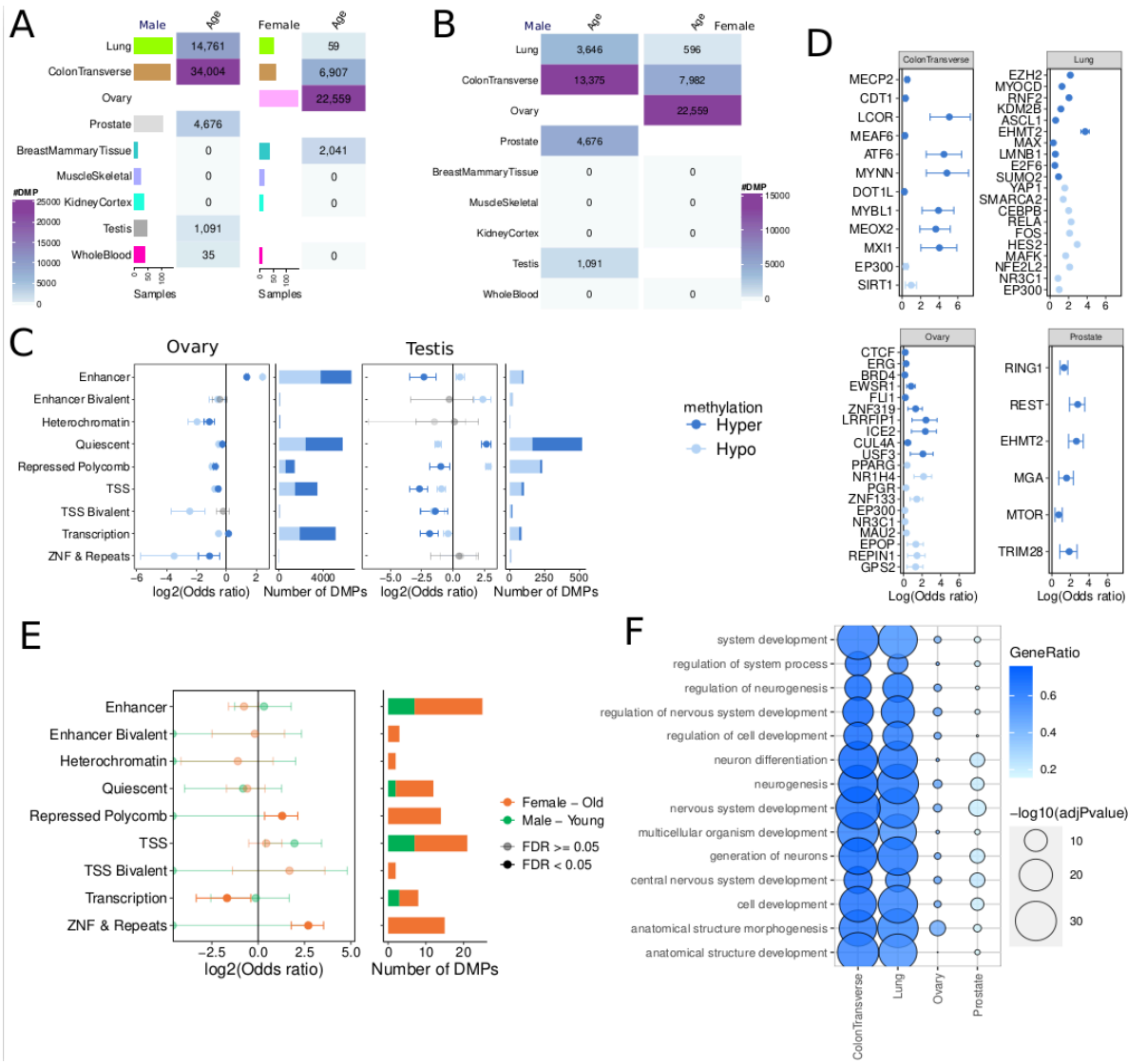


Figure S5. Age-associated DNA methylation changes. A. Number of age-DMPs per tissue in male and female individuals. Left barplot represents the number of samples for every tissue. **B.** Number of age-DMPs per tissue in male and female individuals, downsampling to the same number of individuals per sex-tissue pair. Left barplot represents the number of samples for every tissue. **C.** Age-DMPs enrichment at chromatin states for Ovary and Testis. **Left.** Enrichment OR and confidence intervals. Dots are colored according to the direction of DNA methylation change. Non-significant enrichments are colored in gray tones (two-tailed Fisher's exact test). **Right**. Number of DMPs within each chromatin state per tissue. **D.** Top significant Transcription Factor Binding Sites (TFBS) enriched at age-DMPs for each tissue (ordered by FDR and direction of methylation change). **E.** Enrichment of sex-age additive effects across chromatin states in Lung. Left. Enrichment of DMPs. Dots are colored according to the direction of DNA methylation change and significance (two-tailed Fisher's exact test). Confidence interval is shown. Right. Number of DMPs within each chromatin state per tissue. **F.** Enrichment for the 14 GO biological-process pathways enriched in 4 tissues.

The impact of sex, age, and genetic ancestry on DNA methylation across tissues

Winona Oliveros[1,2], Jose Miguel Ramirez[1,2], Marta Melé[1,3*]

Affiliations

[1]Department of Life Sciences, Barcelona Supercomputing Center (BCN-CNS), Barcelona, Catalonia, 08034, Spain
[2]Universitat de Barcelona, Barcelona, Catalonia, 08034, Spain
[3] Lead contact
*Correspondance: marta.mele@bsc.es