Enhanced epigenetic profiling of classical human monocytes reveals a specific signature of healthy aging in the DNA methylome

Irina Shchukina1,2,3, Juhi Bagaitkar4,5, Oleg Shpynov1,3,13, Ekaterina Loginicheva1, Sofia Porter1, Denis A. Mogilenko1, Erica Wolin6, Patrick Collins1, German Demidov1,5,6, Myktya Artomov7,8, Konstantin Zaitsev2,3, Sviatoslav Sidorov9, Christina Camell9, Monika Bambouskova1, Laura Arthur1, Amanda Swain1, Alexandra Panteleeva1, Aleksei Dievskii1, Evgeny Kurbatsky3, Petr Tsurinov1,3, Roman Chernyatchik1,3, Vishwa Deep Dixit9, Marko Jovanovic4, Sheila A. Stewart10, Mark J. Daly7,8,11, Sergey Dmitriev3, Eugene M. Oltz1 and Maxim N. Artyomov2,12

The impact of healthy aging on molecular programming of immune cells is poorly understood. Here we report comprehensive characterization of healthy aging in human classical monocytes, with a focus on epigenomic, transcriptomic and proteomic alterations, as well as the corresponding proteomic and metabolomic data for plasma, using healthy cohorts of 20 young and 20 older males (~27 and ~64 years old on average). For each individual, we performed enhanced reduced representation bisulfite sequencing-based DNA methylation profiling, which allowed us to identify a set of age-associated differentially methylated regions (DMRs)—a novel, cell-type-specific signature of aging in the DNA methylome. Hypermethylation events were associated with H3K27me3 in the CpG islands near promoters of lowly expressed genes, while hypomethylated DMRs were enriched in H3K4me1-marked regions and associated with age-related increase of expression of the corresponding genes, providing a link between DNA methylation and age-associated transcriptional changes in primary human cells.
transcriptional data highlighted the connection between age-associated transcriptional alterations and hypomethylated DMRs, while hypermethylated DMRs were not readily associated with transcriptional changes. Together with parallel profiling of plasma proteins and unbiased metabolic profiling from the same individuals, the compendium of data collected in this work comprises a comprehensive aging data resource obtained under stringent inclusion criteria. Easily accessible visualization and exploration of all data from this study are available online at https://artemyovlab.wustl.edu/aging/.

**Results**

**Study design, cohort characteristics and systemic age-associated changes.** It has been reported that genetic factors, race, sex, body mass and lifestyles can substantially affect the outcomes of human studies focused on various aspects of aging. Here we used stringent inclusion criteria to eliminate the effect of confounding variables such as inherited genetic traits, sex, environmental stressors and inflammatory disorders. Blood was collected from healthy young (24–30 years old; n = 20) and older (57–70 years old; n = 20) white males. All donors were non-smokers, with a healthy body mass ratio (body mass index (BMI) < 30) and self-reported absence of underlying inflammatory conditions, acute viral infections and cancer. All plasma cytokines measured by a bioplex assay were within normal ranges (Extended Data Fig. 1a and Supplementary Table 1). One cytokine that stood out was interleukin-8 (IL-8), which was strongly statistically increased with age even for the considered cohort size (Extended Data Fig. 1a). Complete blood counts (CBCs) and differential analysis of whole blood (Fig. 1b, Extended Data Fig. 1b and Supplementary Table 1) showed a significant decrease in red blood cell counts with age, consistent with previous reports. Total white blood cell counts were similar between the young and older cohorts, but we saw significant changes in white blood cell differentials: total lymphocyte counts decreased with age, while myeloid cell counts were higher, consistent with the previously described age-associated shift towards the myeloid lineage (Fig. 1b and Extended Data Fig. 1b).

To characterize the differences in the environment that bathes circulating monocytes, we performed proteomic and metabolomic profiling of plasma (Extended Data Figs. 2 and 3). Principal component analysis (PCA) and hierarchical clustering revealed moderate separation between the two cohorts for both datasets (Extended Data Figs. 2a,b and 3a,b). Statistical testing resulted in 39 significantly
different metabolites (false discovery rate (FDR) < 0.05; Extended Data Fig. 2c and Supplementary Table 3) and 53 significantly different proteins (FDR < 0.05; Extended Data Fig. 2c and Supplementary Table 4). Overall, age-associated protein signature corroborated the recently published reports by the Ferrucci group24,25, for example, sclerostin and growth differentiation factor-15 (GDF-15) were among the most distinct proteins, and we additionally validated them and three other targets by enzyme-linked immunosorbent
We identified a number of pathways significantly altered with age (Supplementary Table 8b), including an increase in cytokine signaling pathways, a decrease in the oxidative phosphorylation pathway and a decrease in multiple protein translation pathways (Fig. 2e and Extended Data Fig. 4c). The latter led us to an idea that more profound changes between age groups might be present at the level of proteome. Therefore, we subjected monocytes to proteomic profiling, which detected significant age-associated changes in 134 proteins (Fig. 2f and Supplementary Table 9). We found an increase in protein levels of natriuretic peptide receptor (NPR2), cytokine interleukin-18 (IL-18), myeloperoxidase (MPO) and Toll-like receptor chaperon heat shock protein 90 beta family member 1 (HSP90B1) in the older cohort, suggesting that the baseline condition of monocytes might be skewed towards a more pro-inflammatory state in older individuals. Analysis of age-associated changes in gene expression corresponding to the significant proteins revealed that the majority of the identified protein level alterations could not be explained by a shift in expression, suggesting age-associated disturbance of post-transcriptional regulation (Extended Data Fig. 4f).

The difference in monocytes’ proteome suggested that despite similarity of their transcriptional profiles older and young monocytes might respond differently to activating stimuli. We tested this hypothesis using lipopolysaccharide (LPS) stimulation of monocytes from young (n=7) and older (n=7) donors, and macrophages in vitro differentiated from these monocytes. We collected and sequenced RNA from these four groups of samples (Extended Data Fig. 4d and Supplementary Table 10). While we observed known transcriptional signatures of monocyte differentiation and activation, we again could not detect any significant changes between activated cells from different age groups (Extended Data Fig. 4e). Similar to baseline monocyte profiling, this does not necessarily imply the absence of the age-associated changes but indicates that their absolute magnitude is probably low and requires bigger cohorts to be detected. Taken together, our data show that transcriptional profiles of monocytes do not change considerably along healthy aging, while larger magnitudes of changes are observed on the protein level.

**Identification of DMRs associated with healthy aging.** A number of studies have established a robust relationship between age and DNA methylation using DNA methylation arrays in various cell types or whole tissues\(^\text{30,31}\). These observations were underscored by small-magnitude age-associated changes.

**Protein levels change substantially while transcripts levels are more robust during healthy aging.** To explore intracellular signatures, we profiled pure CD14\(^+\)CD16\(^-\) monocytes (>98% purity; Extended Data Fig. 4a) from young and older individuals using deep RNA sequencing (RNA-seq) (Supplementary Table 6). PCA of these data showed no clear separation of older and young transcripts and differential expression analysis revealed few significantly changing genes (Fig. 2a,b and Supplementary Table 7). One possible reason for this could be that cohort sizes in our study were limiting the power to detect changes of small absolute magnitude. Thus, we re-analysed a large publicly available dataset generated as a part of the Multi-Ethnic Study of Atherosclerosis (MESA) study, which profiled purified monocyte samples from over 1,200 donors between the ages of 44 and 83 years\(^\text{27,28}\). Consistent with the initial report\(^\text{28}\), we identified 4,549 statistically significant differentially expressed genes (Fig. 2c and Supplementary Table 8a). Yet, fold changes of gene expression between middle-aged and older MESA donors were very small—most changes being around 1% of the average expression level of the corresponding gene (Fig. 2d). Our downsampling simulation showed that to detect changes of this magnitude one requires a dataset with at least ~100 donors per group (Extended Data Fig. 4b). Therefore, we conclude that both our data and public data demonstrated the relative stability of transcriptional landscape in human monocytes characterized by small-magnitude age-associated changes.

**Resource**
by the development of DNA methylation-based algorithms for age prediction\textsuperscript{10,28–34}. Yet, DNA methylation array technology is limited in ability to find dedicated regions undergoing age-related DNA methylation or demethylation. Here we used eRRBS\textsuperscript{35}, which allows for identification of continuous regions by sequencing many closely located cytosines, providing deeper insight into the DNA methylation landscape. Individual eRRBS libraries for each sample were sequenced at 70 ± 10 million reads depth (Extended Data Fig. 5a and Supplementary Table 11) and yielded ~3 million well-covered CpGs (with mean coverage ≥10 reads across all samples). This coverage represents ~10% of all CpGs in the human genome, including 24,127 CpG islands (84% of all islands, Fig. 3a). To understand global age-associated changes in methylation profiles, we first compared average levels of cytosine methylation within donors of the two age groups (Fig. 3b–d). While no difference was observed in overall methylation of CpGs outside of the CpG islands, methylation within the CpG islands significantly increased in older donors (Fig. 3b). This result is concordant with previous data showing that CpG islands tend to gain methylation with age\textsuperscript{11,36,37}. Another global change that we confirmed was increased variability in DNA
methylations with age\(^{31-42}\) (Fig. 3e). Finally, we have obtained highly consistent results for the two age groups using both Horvath\(^{32}\) and Hannum\(^{31}\) models, even though a number cytosines had to be imputed to use these approaches (Extended Data Fig. 5b).

eRRBS, as opposed to widespread DNA methylation array technology, allows for identification of DMRs composed of multiple concordantly changing cytosines, which is more likely to identify biologically relevant regions. We used the MethPipe pipeline\(^{41}\) to perform a genome-wide comparison of the methylation between the two groups, which yielded 1,160 DMRs (Fig. 3f and Supplementary Table 12). Approximately half of the regions were hypermethylated with age and were significantly enriched in CpG islands (Fig. 3f and Extended Data Fig. 5c), consistent with our observations on the global level (Fig. 3b). Other regions lost methylation with age and were typically located outside the CpG islands. These findings indicated the presence of multiple demethylation events that accompanied healthy human aging and were equally as characteristic as a gain of methylation in CpG islands.

Next, we sought to establish the robustness of the identified region-based signature by analysing the behaviour of these regions in publicly available datasets. Two whole-genome bisulfite sequenc- ing (WGBS) datasets focused on aging of immune cells were published previously\(^{36,44}\). In both cases, group sizes were limited to one or two samples per group, which precluded statistically sound discovery of an aging signature based on these data. However, the datasets could still be used for validation of aging DMRs identified from our data. First, we compared detected DMRs from our dataset to the dataset for purified classical monocytes from cord blood (n = 2) and from venous blood from older donors (n = 4, age 60–70 years old) available from the Blueprint Consortium\(^{35,54}\). We find that methylation changes in the DMRs that we identified were highly consistent with differences in the same regions observed in the Blueprint dataset. For instance, a DMR located within the GRM2 promoter encompassed 50 CpGs and showed a significant difference in average methylation between cohorts in our data (Fig. 3g), top as well as in WGBS data from the Blueprint Consortium (Fig. 3g, bottom). High reproducibility was also a case when all DMRs were considered: we found that the changes in DMRs were consistent (Fig. 3g) and highly correlated (Extended Data Fig. 5d) between our dataset and the Blueprint dataset. Similar results were obtained when we compared our signature against previously published WGBS samples from a newborn and a 103-year-old centenarian\(^{36}\) (Extended Data Fig. 5e).

Lastly, we compared our signature to the MESA dataset\(^{4,27}\) generated using DNA methylation arrays (Extended Data Fig. 5f). Out of 1,160 DMRs, only a minute fraction had three or more CpGs covered in a widely used Infinium 450k methylation array (used in MESA study), meaning that the signature that we identified could not have been found using array-based profiling techniques (Fig. 3h).

Still, even despite technological differences, PCAs on cytosines located within the DMRs showed a very clear separation by age in the MESA data (Fig. 3i, left). Methylation of the most cytosines located inside DMRs significantly correlated with donor age, and the directionality of the age-associated methylation changes in the MESA dataset matched differences observed in our dataset (Fig. 3i, right). Thus, here we report a novel set of age-associated differentially methylated regions and validate the robustness of this signature across different studies.

Age-related loss and gain of methylation follow distinct chromatin clues. To understand the relationship between identified age-associated changes in methylation patterns and other chromatin features, we focused on the next layer of epigenetic regulation and characterized five post-translational modifications of histone 3 (H3) tails (H3K4me3, H3K4me1, H3K27ac, H3K27me3 and H3K36me3) for monocytes from the young and older groups (Fig. 4a,b and Supplementary Table 13). To generate data for histone modifications for each donor, we optimized the ULI-ChIP–seq protocol\(^{57}\). While the ULI-ChIP–seq allows for robust peak calling within individual samples, the data obtained by this method are considerably more variable between samples when compared with classical ChIP–seq approaches\(^{31}\) (Extended Data Fig. 6a,b). To address this limitation, we developed a new computational approach, called SPAN, inspired by a semi-supervised method described by Hocking et al.\(^{16}\) (Fig. 4a, Extended Data Fig. 7 and Supplementary Information). Peaks called by SPAN were significantly more consistent than output of classical peak callers, as shown in Extended Data Fig. 6 by various quality control metrics. We also evaluated overlap between consensus peaks and existing ChromHMM annotation for CD14\(^{+}\) monocytes to show that generated chromatin profiles agreed with established functional roles of the profiled histone marks\(^{45}\) (Fig. 4c).

Having ensured accurate peak calling, we investigated the age-associated changes in our datasets. PCA plots for each of the histone marks (Fig. 4d) showed an absence of global differences between the two cohorts. We used a number of differential ChIP–seq tools, which detected either no differences or a small number of differences that were not reproducible between the tools and were not confirmed by visual inspection (Supplementary Fig. 1). Next, we sought to leverage the multiomics nature of our dataset and asked whether the chromatin landscape relates to identified
age-associated DMRs. We started by comparing DMRs from our dataset (Fig. 3) to genomic annotation defined by ENCODE ChromHMM for CD14+ monocytes (Fig. 4c and Extended Data Fig. 8a,b) and found that DMRs were highly enriched in bivalent states and Polycomb-repressed regions. Conversely, heterochromatin and quiescent states were significantly unlikely to host a DMR. Enrichment analysis of DMRs against consensus peak sets revealed overrepresentation of DMRs in regions marked with H3K27me3, H3K4me3 and H3K4me1 modifications (Fig. 4f), with distinct characteristics for up and down DMRs (Fig. 4g). Indeed, hypomethylated DMRs (down DMRs) showed the highest enrichment in non-bivalent H3K4me1-marked regions (Fig. 4h).
regions were usually located outside the CpG islands (Fig. 4i, left half of volcano plot), and the overall H3K4me1 signal was significantly higher in hypomethylated regions (Extended Data Fig. 8c). On the contrary, hypermethylated DMRs (up DMRs) typically co-localized with H3K27me3-marked CpG islands (Fig. 4j,k, right half of volcano plot). General H3K27me3 levels were significantly higher in hypermethylated regions compared with hypomethylated regions (Extended Data Fig. 8c; see Extended Data Fig. 8c,d for the remaining chromatin marks). Overall, our data demonstrate that despite the evident age-related re-organization of DNA methylation, the histone modifications H3K4me3, H3K4me1, H3K27ac, H3K27me3 and H3K36me3 are stable and do not undergo drastic re-arrangement with age in the basal state. At the same time, age-associated changes in DNA methylation are enriched within specific chromatin features, distinct for up and down DMRs.

DNA hypomethylation is directly linked to age-associated transcriptional changes. Significant fractions of both up and down DMRs were located in gene promoters (~60% and ~40% respectively; Fig. 5a). Therefore, we wanted to evaluate the regulatory potential of identified regions and their impact on expression level of the corresponding genes. Strikingly, while genes with hypomethylated promoters obeyed regular distribution of gene expression, genes with hypermethylated promoters were either not expressed or expressed at low levels in all samples (Fig. 5b). This lack of expression of genes with hypermethylated regions (up DMRs) in their promoters is consistent with the abundance of repressive H3K27me3 mark in up DMRs (Fig. 4k). These observations also suggested that hypomethylated DMRs were more likely to have a functional impact on gene expression than hypermethylated DMR. Specifically, we reasoned that age-related hypermethylation of promoters associated with up DMRs could only further decrease expression of the corresponding genes that are already lowly expressed, indicating unlikely functional consequences of this decrease. However, genes with promoters that lose DNA methylation with age (down DMRs) would increase their expression levels, indicating possible functional impact of these regions (Fig. 5c, left). To test this hypothesis, we determined whether genes with hypo- or hypermethylated DMR intersecting their promoter region (−10 kb; +3 kb) around transcription start site (TSS) were significantly enriched within up- or downregulated genes derived from the highly powered MESA transcriptomic dataset (Fig. 2c). Strikingly, gene set enrichment analysis (GSEA) indeed showed that hypomethylated DMRs were significantly upregulated with age in the MESA cohort, while no significant enrichment was observed for genes with hypermethylated promoter regions (Fig. 5c, right).

Next, we wanted to ask whether up and down DMRs tended to co-localize with any specific transcription factors. We performed enrichment analysis comparing DMRs to binding sites of 485 transcription factors annotated by the ReMap atlas, while adjusting for various cell types and tissues is required to make a solid conclusion. Interestingly, we identified four transcription factors that were significantly co-localized with up DMRs after adjustment for CpG islands (Fig. 5d and Supplementary Table 14). Among them, Jumonji (encoded by JARID2) and DEAD box protein 5 (DDX5) are known to be associated with Polycomb repressive complex 2 (PRC2), which is consistent with co-localization of up DMRs with H3K27me3 mark. But surprisingly, we also identified two novel transcription factors from one family—methyl-CpG-binding domain protein 2 (MBD2) and MBD3—that had not been linked to age-related DNA hypermethylation previously (Fig. 5e). All detected proteins were highly expressed in monocytes (Extended Data Fig. 8e), making them strong candidates for potential regulators of age-associated DNA methylation gain.

Together with enrichment of the DMRs in the specific chromatin signatures, these comparisons highlighted a distinct regulation as well as functional importance of the DMRs identified in our analysis. We propose the following hypothesis to describe the relationship between age-related DNA methylation, its regulation and its physiological impact on cellular function (Fig. 5f). Hypermethylated regions are enriched within CpG islands and are often characterized by a bivalent signal. These regions are associated with silent genes, and, consequently, their hypermethylation has little impact on the already absent expression. In contrast, hypomethylated regions are generally outside of CpG islands and enriched in regions marked by H3K4me1 alone. Hypomethylated DMRs are associated with age-related upregulation of their corresponding genes.

Up and down DMRs behave distinctly in different physiological and clinical contexts. Next, we investigated the behaviour of up and down DMRs in a variety of physiological and clinical settings related to aging. As most of the public datasets were generated using DNA methylation array technology, we were able to retrieve methylation levels of only a fraction of DMRs in this analysis, as discussed above (Fig. 3h). Nevertheless, we found that average methylation level of cytosines from up and down DMRs that were covered by the array was able to accurately capture the difference between young and old twin pairs8 (Fig. 6a). Furthermore, this metric correlated significantly with age of donors from MESA dataset (Fig. 6b, \( P < 0.001 \) for both), indicating that methylation of these regions continued to change with age even in a population that was older than our studied cohort. However, the behaviour of up and down DMRs was different when we looked at the data from bulk brain tissue of a healthy aging cohort10: average methylation of up DMRs was still associated with age, down DMRs failed to show significant correlation with age (Fig. 6c). While more comprehensive analysis of various cell types and tissues is required to make a solid conclusion, this observation suggests differential tissue/cell-type specificity of up and down DMRs.

To test whether up and down DMRs were relevant for the phenotype of accelerated aging, we compared DMR methylation in human immunodeficiency virus (HIV)-positive donors and matching...
controls. Strikingly, while age was robustly associated with both up and down DMRs, age acceleration in individuals with HIV was evident only for hypermethylated up DMRs (Fig. 6d, left). Comparison of age- and sex-adjusted levels of mean methylation allowed the quantification of this difference, confirming that up DMRs were more methylated in patients with HIV irrespective of donors age (Fig. 6d, right). Further analysis of published clinical phenotypes revealed a novel association with age acceleration: we find that the methylome of lungs of patients with asthma demonstrate age acceleration in a similar manner as observed for patients with HIV (Fig. 6e).

**Age-associated methylome–transcriptome conjecture**

**Down DMRs enrichment**

adjusted for H3K4me1

**Up DMRs enrichment**

adjusted for CpG islands

MBD2

DDX5

MBD3

JARID2

EZH2

---

**TFs ordered by DMRs enrichment**

Fold change (DMR/random)

0

0.1

0.2

0.3

0.4

0

50

100

150

TFs ordered by DMRs enrichment

---

**Age-associated methylome–transcriptome conjecture**

Up and down DMRs follow distinct distribution patterns across the genome

Up and down DMRs have different functional output

Potential regulators
Notably, we did not observe any age acceleration or delay in either brain or blood of patients with Alzheimer’s disease (Extended Data Fig. 9a).

Next, we asked whether changes in DNA methylation driven by lifestyle choices were also affected by age. We focused on obesity and smoking phenotypes due to available RRBS data for these features that enabled investigation of corresponding signature in our dataset. Smoking did not demonstrate any connection with age-associated DMRs (Extended Data Fig. 9c). However, comparing signatures of obesity reported by Day et al. to our dataset, we were able to observe a non-trivial relationship between DNA methylation, aging and BMI. Based on their RRBS data, Day and co-authors identified 170 obesity DMRs that were well covered in our eRRBS data (Fig. 6f). Importantly, obesity- and age-associated DMRs were distinct sets of genomic regions, and no change in methylation in age-associated DMRs between lean and obese groups was observed (Extended Data Fig. 9b). Next, we looked at the obesity-associated DMRs and leveraged variability of BMI in our cohort to compare it with our data. Even though none of our donors were obese, we saw a consistent increase in mean methylation of obesity up DMRs in the overweight group, but only in the case of older donors (Fig. 6g).

Consistently, when mean methylation of obesity-associated up and down DMRs was plotted against BMI, it was evident that obesity up DMRs were significantly more methylated in the older donors from our cohort (Fig. 6h). This suggests that the methylation of healthy young individuals is more robust to small variations in lifestyle (for example, weight change), while in older individuals even mild BMI differences can impact the DNA methylome.

Hypomethylated regions are genetically linked to a number of conditions. Lastly, we compared age-associated DMRs against medical and population genetic databases, such as the UK Biobank. We evaluated overrepresentation of DMRs in sets of single nucleotide polymorphisms (SNPs) linked to 34 clinical phenotypes through genome-wide associated studies (GWAS). While the overall distribution of DMRs across the chromosomes was fairly random, we found DMRs to be significantly enriched in four sets of SNPs (Fig. 7a,b). Only down DMRs showed enrichment in physiological phenotypes (Fig. 7b), which was consistent with our previous observation that hypomethylated regions had more functional potential in contrast to hypermethylated DMRs, associated with predominantly silent genes. Specifically, down DMRs were significantly overrepresented in SNPs associated with asthma, level of glycated haemoglobin (Hb), total protein in blood and multiple sclerosis (MS).

The majority of SNPs associated with significant phenotypes were located within the human leukocyte antigen (HLA) locus on the sixth chromosome (Fig. 7c,d). This genomic region contains multiple genes regulating immune response, including genes that encode antigen processing and presentation complexes. We confirmed statistically significant enrichment of down but not up DMRs in the HLA locus by random simulations (Fig. 7e). One of the down DMRs resided directly in an exon of the HLA-DQB1 gene, encoding a part of the major histocompatibility complex class II (Fig. 7f).

Discussion

We generated and analysed data describing transcriptomic, proteomic and epigenetic changes in human CD14+CD16− monocytes during physiological aging using a stringently selected healthy male cohort. We show that in the absence of other inflammatory conditions, aged monocytes are associated with cell-intrinsic alterations in DNA methylation patterns yet are not associated with any dramatic rearrangements in their transcriptional or chromatin profiles for five common chromatin marks (H3K4me3, H3K27me3, H3K4me1, H3K27ac and H3K36me3). This result is consistent with the absence of a global change in the abundance of various histone modifications in monocytes that was recently shown using mass cytometry (CyTOF). Importantly, co-analysis of our data with the MESA transcriptional dataset that includes 1,200 monocyte samples showed that while transcriptional differences definitely accompanied aging, their magnitude was very small and required high statistical power to be detected, at least in the case of classical monocytes. It is feasible that larger cohort studies might also uncover statistically significant differences in histone modifications between ages, yet it is fair to conclude that the absolute magnitude of such changes would probably be very moderate. Furthermore, age is known to be sex dysmorphic, suggesting that more profound changes could be identified between female cohorts by future studies.

Utilization of next-generation sequencing-based technology (eRRBS) allowed us to establish a pattern of age-associated changes in DNA methylation and define DMRs that change with age. In contrast to previous studies that focused on DNA methylation array data and, therefore, identified sets of single-stand CpGs associated with age, we identified age-associated regions that represented concordant changes across multiple neighbouring cytosines. This strategy yielded highly robust and physiologically relevant regions as shown by their ability to clearly capture age-specific variations in multiple independent published datasets (Figs. 3 and 6). We found that hypo- and hypermethylation regions were nearly equally frequent in our data and that they differed significantly in their genomic locations, chromatin profiles and relation to the transcriptional activity. Hypermethylated regions were found to be strongly associated with H3K27me3-marked CpG islands residing in the promoters of silenced genes. This pattern matches a previously proposed hypothesis that Polycomb is involved in age-associated hypermethylation. In addition, we identified MBD2 and MBD3 as new putative regulatory candidates strongly associated with up DMRs. We found hypomethylated DMRs to be highly enriched in non-bivalent regions carrying the H3K4me1 mark, consistent with comparison of mesenchymal stem cells age-associated signature versus cell line ChIP-seq data. This observation might suggest indirect regulatory function of down DMRs and provide a new insight into the role of DMRs in physiological aging.
into possible mechanisms of age-associated methylation loss that are yet to be revealed. Unlike CpG islands, the H3K4me1 modification marks sites that often co-localize with cell-type-specific regulatory regions\(^6\). Accordingly, this suggests that down DMR might be more cell-type specific. Altogether, these observations underscore the existence of orthogonal processes (global and cell-type specific) establishing age-related DNA methylation patterns as well as the importance of profiling pure cell populations.

![Graphs and data tables](image.png)

**Twin study (GSE52114 whole blood)**

- **Up DMRs**
  - Mean DMR methylation: 0.38
  - Status: HIV–, HIV+
  - Age-adjusted DMR methylation: 0.40

- **Down DMRs**
  - Mean DMR methylation: 0.42
  - Status: HIV–, HIV+ Asthma
  - Control

**HIV (GSE67705 whole blood)**

- **Up DMRs**
  - Mean DMR methylation: 0.46
  - Status: Healthy brain (GSE66351 bulk tissue)

- **Down DMRs**
  - Mean DMR methylation: 0.42

**Asthma (GSE85666 lung epithelial cells)**

- **Up DMRs**
  - Mean DMR methylation: 0.49

- **Down DMRs**
  - Mean DMR methylation: 0.47

**Healthy brain (GSE66351 bulk tissue)**

- **Up DMRs**
  - Mean DMR methylation: 0.50

- **Down DMRs**
  - Mean DMR methylation: 0.47

**Obesity RRBS profiling**

Day et al.\(^7\) (GSE73303)

- Lean (n = 11) Obese (n = 9)

- 170 Obesity DMRs 149/170 covered in our dataset

- Compare their methylation to BMI in our young and older groups

**MESA (GSE56046 monocytes)**

- **Up DMRs**
  - Mean DMR methylation: 0.44
  - Status: Control

- **Down DMRs**
  - Mean DMR methylation: 0.27

**NATuRE AGING | VOL 1 | JANUARY 2021 | 124–141 | www.nature.com/nataging**
Fig. 7 | Down DMRs are enriched in genetic susceptibilities loci for asthma, MS, glycated Hb and total protein. a, Manhattan plot of chromosomal locations and corresponding $-\log_{10} P$ values for CpGs compared between older and young donors as in Fig. 3f. b, Quantile–quantile plot for DMR enrichment analysis against GWAS datasets. Each dot represents a phenotype compared against down or up DMRs. c, Top hits of DMR enrichment analysis. MS, multiple sclerosis; AMD, age-related macular degeneration; LDL, low-density lipoproteins. d, Manhattan plots of SNPs associated with significant phenotypes and located within the HLA locus. Green vertical bars highlight down DMRs located in this locus. e, Enrichment of DMRs in the HLA locus. The histogram shows the distribution of simulated intersection sizes (10,000 random simulations). f, Visualization of DMR intersecting the HLA-DQB1 exon. Each dot represents a CpG and the green vertical bar highlights the DMR. $P$ values in b, c and e are calculated by a one-sided enrichment test as described in Methods. $P$ values in d are retrieved from the GWAS catalogue.
Most importantly, hypo- and hypermethylated DMRs also showed a striking difference in their regulatory impact on transcriptional activity of the associated genes. While genes associated with hypomethylated regions showed a normal expression level distribution, hypermethylated regions were predominantly linked to repressed genes. Accordingly, since methylation gain would only lead to further decrease in transcription, we observed no downstream transcriptional effect of the hypermethylated regions. The hypomethylated regions, however, showed significant enrichment among the genes upregulated with age as defined through re-analysis of data from the large independent cohort (MESA). This observation establishes the functional output of age-associated hypomethylation of DNA, proposes that these down DMRs have a more direct effect on the transcriptional state of the cell and explains the mechanism behind upregulation of a subset of age-related genes. With respect to age-associated hypermethylation of already silenced genes, this process can serve as a protective mechanism that allows cells to avoid tumourigenic transformation.\(^5\)

Enrichment of identified hypomethylated DMRs among functionally important regions further underscores their regulatory potential. We observed overrepresentation of down DMRs in the HLA region, suggesting an effect of methylation changes on antigen presentation. This enrichment also drives significant intersection of down DMRs with SNPs associated with asthma, MS and other phenotypes. The gradual nature of age-associated changes of DNA methylation might explain why diseases such as MS and adult onset asthma develop later in life. Interestingly, in the case of asthma, only half of the SNP–DMR intersections were explained by demethylation of the HLA locus, revealing a more robust link between DNA methylation and the disease. We also showed that asthma was associated with acceleration of the methylation clock in lung epithelial cells. Surprisingly, down DMRs were not sensitive to this accelerated methylation and the disease. We also showed that asthma was associated with acceleration of the methylation clock in lung epithelial cells. Surprisingly, down DMRs were not sensitive to this acceleration, suggesting that asthma itself does not drive changes in methylation of down DMRs, but age-related loss of methylation in these cells to avoid tumourigenic transformation.\(^5\)

Plasma profiling using metabolomics. For each donor, 250 μl of frozen plasma was shipped on dry ice to Metabolon (http://www.metabolon.com) for liquid chromatography-tandem mass spectroscopy (LC-MS); 734 peaks were annotated. Pre-processing was performed by Metabolon: peaks were quantified using area under the curve and each compound was corrected in run-day blocks by registering the medians to equal one and normalizing each data point proportionately (termed the ‘block correction’) to correct variation resulting from instrument inter-day tuning differences. DAY20 was excluded from the analysis as an outlier based on PCA. Significant differences in metabolites between cohorts were determined using two-sided Mann–Whitney U-test. P-values were adjusted for multiple testing using the Benjamini–Hochberg method.

Bioelectrical and ELISA. Plasma was profiled using a MesoScale V-plex Pro-inflammatory Panel I Human kit (IFN-γ, IL-1β, IL-2, IL-4, IL-6, IL-8, IL-10, IL-12p70, IL-13, TNF-α and statistical differences in fold change were determined by two-sided Mann–Whitney U-test. ELISA kits were used to determine GDF-15, Sclerostin (RnD), Osteomodulin (Aviva), Notch1 (ThermoFisher) and sCD86 (Abcam) as per the manufacturer’s instructions.

PLASMA profiling using proteomics. For each donor, 500 μl of frozen plasma was shipped on dry ice to the Genome Technology Access Center (GTAC) core facility at Washington University in St. Louis for high-density protein expression analysis via SomaScan assay.\(^58\). Profiles for ~1,300 analytes were acquired. Pre-processing was performed by GTAC core facility at Washington University in St. Louis: raw relative fluorescence units measurements for every SOMAmer reagent were normalized subsequently with hybridization normalization, plate scaling, median and calibrator normalization and transformed in log2, scale. Differential analysis was done for young versus older groups using quantile-normalized data for all 40 samples. For differential analysis, functions limFit and e湾es from the limma package (v3.43.5) were used. P-values were adjusted for multiple testing using the Benjamini–Hochberg method.

Correlation analysis (Extended Data Fig. 3e) used significant plasma proteins and all profiled plasma metabolites that had non-zero variance. Pheatmaps of absolute values of Spearman correlation coefficients were plotted using Pheatmap R package (v1.0.12).

RNA-seq data for young versus older monocytes. Total RNA was isolated from snap-frozen monocyte pellets using Qiagen’s RNeasy Mini Kit according to the manufacturer’s protocol. RNA concentration and integrity were assessed by Agilent 2200 Tape Station. RNA samples were submitted to BGI for long non-coding RNA sequencing and small RNA sequencing. After library construction, paired-end 100 base pair reads were generated on the DNBSEQ platform.

FASTQ files for each sample were aligned to the hg19 genome (Gencode, release 28) using STAR (v2.6.1b) with the following parameters: STAR –genomicDir $GENOME_DIR –readFilesIn $WORK_DIR/FILE_1 $WORK_DIR/FILE_2 –runThreaded $readFilesCommand zcat –outFilterMultimapNmax 15 –outFilterMismatchNmax 6 –outReadsUnmapped Fastx –outSAMstrandField intronMotif –outSAMtype BAM SortedByCoordinate –outFileNamerPrefix .\(^5\).

Quality control for each sample was performed by FastQC (v0.11.3) and Picard tools (v2.18.4). Quantification was done using htseq-count function from HTSeq framework (v0.6.1): htseq-count -f bam -r pos -s no -t exon SSBAM $ANNOTATION\^ OUTPUT.

Raw counts were normalized before PCA via getVarianceStabilizedData function from DeSeq2 package (v1.24.0). Differential expression analysis was done using DESeq function from DESeq2 with default settings. The following design was used in the analysis: gene = age + batch + PC1 + PC2 + PC3. PC1, PC2 and PC3 are the three main principal components explaining genetic variability in the cohort.\(^5\)

Gene lists were retrieved from the CHIP-seq raw reads as described in Supplementary Information and Supplementary Fig. 2. The significance threshold was set to an adjusted P < 0.05.
Publicly available transcriptional data. The transcriptomic dataset for the MESA cohort was re-analysed. GSE56045 contains normalized expression values for purified human monocytes. Differential analysis was performed using the limma package (v3.9.4.5). We corrected for confounding variables by including chip and race–gender–site parameters into model design: gene = age + chip + race–gender–site. Age was used as continuous variable, no separation into age groups was performed. P values were adjusted for multiple testing using the Benjamini–Hochberg method. The significance threshold was set to FDR < 0.05. Gene set enrichment analysis via the fgsea R package was performed. The significance threshold was set to FDR < 0.05. Gene set enrichment analysis via the fgsea R package was performed.

RNA-seq data for differentiation and stimulation of monocytes. Differentiation of primary human CD14+/CD16+ monocytes into resting macrophages occurred after seven days of culture in RPMI media supplemented with 11 mM glucose, 2 mM glutamine and 10% fetal calf serum in the presence of 50 ng/ml of M-CSF (Peprotech, catalogue number 300-25). 5 × 10⁶ resting macrophages were activated with 10 μg/ml of lipopolysaccharides (LPS) from Escherichia coli (Sigma, O111:B4) or mock for 24 h. RNA was isolated from monocytes and macrophages using AllPrep DNA/RNA Mini Kit (Qiagen, catalogue number 80204) and treated with RNase-free DNase (Qiagen, catalogue number 79254). Libraries were prepared as described previously. Briefly, complementary DNA was synthesized using custom oligo(dT) primers with a barcoded adaptor-linker sequence (CCCTACGAGACTTCTGCAGTCT-CCCCCTTTTTT). Barcoded cDNA was pooled together before downstream quantitative PCR analysis and the RNA–DNA hybrids were degraded by consecutive acid–alkali treatment. A second sequencing linker (AGATCGGAACGGCAGCACCCTGTCG) was ligated via T4 ligase (NEB), followed by clean up with SPRI beads (Beckman–Coultier). The libraries were amplified by 12 cycles of PCR and cleaned up with SPRI beads, yielding strand-specific RNA-seq libraries. Data were sequenced via HiSeq 2500 40bp × 10bp paired-end sequencing. Files obtained from the sequencing centre were demultiplexed using the fastq-mxl tool. Fastq files for each sample were aligned to the hg19 genome (Gencode, release 28) using STAR (v2.6.1b) with the following parameters: STAR –genomicDir SGENOME_DIR –readFilesIn SREADDIR/SIFILE –runReadParallel 4 –outFilterMultimapNmax 15 –outFilterMismatchNmax 6 –outSAMunmapped Fastx –outSAMstrandField intronMotif –outSAMType BAM SortedByCoordinate –outFileNamePrefix S. Quality control for each sample was performed using FastQC (v0.11.3) and Picard tools (v2.18.4). Alignment reads were quantified using a HTSeq-based quantp3 script (available at https://github.com/clab/quantp3/) to account for sequencing-specific bias and low level of sequence specificity close to 3' end. DESeq2 (v1.24.0) was used for analysis of differential gene expression.

Monocyte proteomic data. For proteomic analysis, we used an independent cohort of donors that was recruited using the same inclusion criteria as in the rest of the study (n = 11 samples per group). For each donor, snap-frozen monocytes (1 × 10⁶ per sample) were processed on dry ice and subjected to LC–MS analysis. Details of Biognosys Discovery protein profiling pipeline (HRM ID + mass spectrometry) can be found on their website (https://biognosys.com/technology/discovery-proteomics). On average, 5,580 proteins were quantified per sample. In total, 5,804 proteins represented by 74,348 peptides were quantified per sample. In total, 5,804 proteins represented by 74,348 peptides were quantified. The fold change of each peptide ion variant was estimated as average abundance of peptide ion variant across biological replicates in the older group/average abundance of peptide ion variant across biological replicates in the young group. The values were then log-transformed and fold changes of all peptides belonging to the same protein were compared with zero using a two-sided paired t-test. Multiple testing correction was performed as described in Storey et al. We removed major histocompatibility complex proteins from the analysis due to high similarity between the variants that could not be accurately resolved using this data. For visualization purposes (Fig. 2g), we used protein intensities estimated as an average over all biological replicates of a donor for each sample. More than 500,000 cytosines were covered with at least ten reads in one sample. More than 500,000 cytosines were covered with at least ten reads in each of samples, in accord with field standards (Fig. 3a). We removed all cytosines with insufficient coverage: average coverage across all 40 samples was required to be greater than or equal to ten reads. After filtration, 2,808,448 CpG cytosines remained in the analysis (we refer to this set as ‘covered CpG cytosines’, and all the downstream analysis used only these cytosines). Overall, for approximately 84% of CpG islands, at least one cytosine in CpG context was covered in the experiment.

CpG island methylation analysis of monocyte eRRBS data. To investigate global changes in the methylene, we averaged methylation levels of all cytosines in CpG context inside and outside CpG islands for each donor (Fig. 3b). Two-way analysis of variance (ANOVA) (~age + batch) was used to calculate P values. The mean methylation level of each CpG island refers to a mean of methylation levels of all covered cytosines within the island. We included age as a continuous variable in methylation analysis. We used multiple methylation clock models, methylation levels of non-covered in our dataset CpGs were imputed as an average methylation within a [-100 kb; +100 kb] window around a CpG. Methylation was set to zero if imputation was not possible (no CpGs were covered in the window.)

Differential methylation analysis of monocyte eRRBS data. We used the Methpipe pipeline (v3.4.3) to find DMRs44. Initial per-donor methylation files were converted to Methpipe metchnoot format and merged into a proportion table using merge-methcats function. Design table included intercept, age and batch. For each cytoine, a linear model was fit by radmeth regression function (radmeth regression -factor age -OUTPUT_CYTO_DIRECTION_TBL_PROP_TBL, and significance combining and adjustment for multiple testing was performed by radmeth adjust command (radmeth -adjust -bins 1:50:1 OUTPUT_CYTO>ADJ_ CYTO). Significantly different cytosines were merged into regions by radmeth: merge -cytosines (OUTPUT_CYTO>DMRS). Resolving regions were subjected to further filtration: number of cytosines within the region ≥3 and absolute value of difference ≥0.025.

For each DMR, a combined P value was calculated by the Fisher method (function fisher.method from BisrrRNA package v0.2.2) from Methpipe P values for cytosines located within the region. Combined P values were used for visualization purposes (plots, if ~adj > 0.05). P values were transformed using the CHIPEeker package (v1.20.0). A promoter was defined as [-10 kb; +3 kb] relative to the TSS. To find DMRs that intersect CpG islands we used bedtools2 (v2.25.0) intersect function.

Comparison of monocyte eRRBS data with public datasets. The publicly available Blueprint dataset EGA:AD0001002233 was used for DMR validation. The methylation signal in BigWig format was downloaded from the Blueprint online portal for classical CD14+CD16– monocytes from cord blood (n = 2) and older donors (n = 4, 60–70 years old). Bedtools2 (v2.25.0) map function was used to calculate mean methylation level for each DMR in Blueprint samples (bedtools map BEDS bed -b BP SAMPLE.bedGraph -c 0 -o mean)45. IGV (v2.3.72) was used for BigWig-file visualisation (Fig. 3c).

WGBS data for cord blood and a 103-year-old centenarian previously published by Heyn et al. was downloaded from the GEO database (GSE31263) in BED format. DMR methylation (Extended Data Fig. 5e) was estimated as mean methylation of all CpGs residing inside DMR and covered in WGBS sample. WGBS methylation profile for MIR155 was downloaded from GEO database GSE56046. M values were transformed into beta-values by m2beta function from lumi package (v2.26.4). The batch effect stemming from chip and race–gender–site variables was removed by ComBat function from sva package (v3.22.0). For Fig. 3i, we only used methylation levels of CpGs residing within DMRs. In Fig. 3i (right), we used library preparation protocols (ten young and ten older in each batch). Therefore, we account for a possible batch effect (Extended Data Fig. 5k) in all the following analysis.
ULI-ChIP-seq. Aliquots of 100,000 CD14⁺CD16⁻ monocytes were thawed on ice for 5 min then immediately resuspended in 20 µl of EZ Nuclear Isolation Buffer (Sigma-Aldrich), and incubated for 5 min on ice. Samples were digested using 2 units/µl MNase in 20 µl MNase Digestion Buffer (NEB) for 5 min at 37°C. Reactions were stopped by the addition of 10 µM EDTA and 0.1% Triton/0.1% deoxycholate (final concentration). Isolated chromatin was incubated on ice for 15 min, followed by vortexing on a medium setting for 30 s. The volume was adjusted to 200 µl with Complete IP Buffer (20 mM Tris-HCl pH 8.0, 2 mM EDTA, 150 mM NaCl, 0.1% Triton X-100, 10 mM sodium butyrate, 1X protease inhibitor cocktail, 1 mM PMSF) and incubated for 1 h at 4°C on a gentle rocking platform. Ten percent of the total chromatin was removed to assess digestion efficiency and to use as an input control.

Chromatin for immunoprecipitation was pre-cleared using Protein A Dynabeads (Invitrogen) for 1–4 h at 4°C and subjected to immunoprecipitation overnight at 4°C (overnight at 0.3 µg input DNA/million monocytes). A 0.3 million CD14⁺ monocytes were thawed on ice for 5 min. ULI-ChIP-seq libraries depth (RPM) in weak consensus peaks. Second, we ran differential analysis tools, described in ref. 78. DiffBind (v2.4.8), MACS2 bdgdiff (v2.1.1) and diffReps (v1.55.6). We also added ChIPDiff (version from 27 March 2008) to the comparison. As Macs2 bdgdiff does not support replicated data, we pooled samples into a single BAM file. A single pooled control for all young and older samples was used with all tools. See Supplementary Information for exact setting used for differential analysis.

Overrepresentation analysis. We used the ENCODE 18-chromosomes ChromHMM annotation (ENCSCR071CD, CD14⁺ monocytes) for overrepresentation analysis. As our visual inspection using eRRBS technology, only a fraction of the genome was covered, and we had to account for incomplete background in our analysis. First, a RED file with all covered regions was generated by merging filtered CpGs located closer than read length (50 bp) and expanding regions to be no shorter than read length. Next, bedtools2 (v2.25.0) function shuffle was used to simulate 100,000 sets of non-overlapping regions with length distribution that matched DMRs and located within 0.05 µg of the genome. For each set, we calculated the real number of DMRs intersecting the state as well as intersection size with each simulated set (bedtools intersect –a RegionSet –b ChromHMMState –u). The expected size of intersection for each state was estimated as the mean of simulated intersection sizes. Corresponding fold change (DMRs/random) was defined as (real intersection/expected intersection). Significance of over- and underrepresentation for each ChromHMM state was calculated as (number of simulated sets with intersection higher than real/100,000) and (number of simulated sets with intersection lower than real/100,000), respectively (Extended Data Fig. 5c). The Benjamini–Hochberg method was used to adjust for multiple testing.

Overrepresentation analysis against peak sets was performed similarly. Weak consensus peaks (peaks confirmed by at least two samples) for each histone modification profiled in our study were used in this analysis.

PCA of ChromHMM states was performed similarly to PCA of CpG islands mean methylation (described in the eRRBS methods section): mean methylation level of a region refers to a mean methylation level of genes within the region. Values were centred and scaled for PCA (Extended Data Fig. 8b).

Comparison of ChIP-seq signal. ChIP-seq signal intensities in DMRs were estimated using DiffBind package (v2.4.8). We used counts = dba.count(dba$sampleSheet = sheet), FragmentSize = 125, bimoveDuplicates = TRUE), counts = dba.count(counts, peaks=NULL, Score=DBA_SCORE_TMM_MINUS_FULL, tmm_normalized = dba.peaks(k27ac_counts, bRetrieve = TRUE, DataType = DBA_DATA_FRAME) commands to retrieve TMM-normalized signal values for DMRs. A two-sided Mann–Whitney U-test was performed to calculate the P value between hypo- and hypermethylated DMRs.

DMR-associated expression patterns and enrichment in MESA transcriptomic data. To characterize expression of genes associated with changing methylation, we selected all genes with DMRs located in their promoters (−10 kb; +3 kb) around the TSS (see description of the annotation procedure above). Raw counts were generated for our RNA-seq dataset by HTSeq as described above. Gene expression levels for all genes were log-transformed (log1(1 + raw count)) and quantile normalized. Same sets of genes were used to perform GSEA using ranked list obtained for MESA transcriptional dataset (see details above). Genes were ordered based on t-statistics generated by limma. 10,000 permutations were performed to evaluate significance via fgsea package (v1.10.0).

TF binding sites overrepresentation analysis. Binding sites of 485 transcription factors were obtained from ReMap 2018 (v1.2) website (http://pedagogy-tagc.univ-mrs.fr/remap/index.php?page=download). Hg19 merged peak files were downloaded. General workflow of the overrepresentation analysis is described above. To account for enrichment of up DMRs in CpG islands, we used only a fraction of the genome as described above. Gene expression levels for all genes were log-transformed (log1(1 + raw count)) and quantile normalized. Same sets of genes were used to perform GSEA using ranked list obtained for MESA transcriptional dataset (see details above). Genes were ordered based on t-statistics generated by limma. 10,000 permutations were performed to evaluate significance via fgsea package (v1.10.0).

Integration with public data. First, we investigated the behaviour of up and down DMRs in several DNA methylation array-based datasets. Mean methylation level of up or down DMRs was estimated for each sample in a public dataset as an average of all CpGs covered in the dataset that reside within the corresponding set of DMRs. Spearman correlation coefficients between DMR mean methylation and disease measure were calculated when applicable. We fit a methylation age–sex model using lm and took residuals to estimate age- and sex-adjusted DMR methylation. A two-sided Mann–Whitney U-test was used to calculate P values. We explored seven public datasets generated using DNA methylation arrays:

1. Twin study dataset⁶⁶ (GSE52114). Probe methylation values were used as provided in GEO.

2. MESA dataset⁷⁵ (GSE56046). See details of data processing above.

3. Bulk brain tissue dataset (GSE46351). Original study⁷ was focused on Alzheimer’s disease, so we pre-selected all affected individuals from the dataset and kept only a cohort of healthy controls. We used data from bulk brain tissue and removed one sample (GSM2808945) as an outlier based on PCA.
Second, we compared our data with RRBS-based datasets that focused on obesity and smoking. (1) Obesity DMRs were acquired from supplementary table 3 of the original study. 149 obesity DMRs had at least one CpG covered in our data. Mean methylation of up and down obesity DMRs was calculated for all donors as an average methylation of all CpGs from obesity up/down DMRs that were covered in our data (Fig. 6g). BMI-adjusted methylation of obesity DMRs was calculated by fitting a methylation – BMI model and taking residuals. P values were calculated using a two-sided Mann–Whitney U-test.

Data visualization. In all figures, the lower and upper hinges of all boxes represent the 25th and 75th percentiles. Horizontal bars show the median value. Whiskers extend to the values that are no further than 1.5×IQR from either the upper or lower hinge. IQR stands for interquartile range, which is the difference between the 75th and 25th percentiles. For all PCAs, percentage of variance explained by principal components is shown in brackets.

References

1. Marengoni, A. et al. Aging with multimorbidity: a systematic review of the literature. Ageing Res. Rev. 10, 430–439 (2011).
2. Bektas, A., Schurman, S. H., Sen, R. & Ferrucci, L. Aging, inflammation and the environment. Exp. Gerontol. 105, 10–18 (2018).
3. Cesari, M. et al. Antioxidants and physical performance in elderly persons: the Invecchiare in Chianti (InCHIANTI) study. Am. J. Clin. Nutr. 79, 289–294 (2004).
4. Reynolds, L. M. et al. Age-related variations in the methylene associated with gene expression in human monocytes and T cells. Nat. Commun. 5, 5366 (2014).
5. Patten, M. B. et al. Natural variation in the parameters of innate immune cells is preferentially driven by genetic factors. Nat. Immunol. 19, 302–314 (2018).
6. Reerman, I. et al. Proliferation-dependent alterations of the DNA methylation landscape under hematopoietic stem cell aging. Cell Stem Cell 12, 413–423 (2013).
7. Benyoun, B. A., Pollina, E. A. & Brunet, A. Epigenetic regulation of aging: linking environmental inputs to genomic stability. Nat. Rev. Mol. Cell Biol. 16, 593–610 (2015).
8. Horvath, S. & Raj, K. DNA methylation-based biomarkers and the epigenetic clock theory of ageing. Nat. Rev. Genet. 19, 371–384 (2018).
9. Johnson, A. A. et al. The role of DNA methylation in aging, rejuvenation, and age-related disease. Rejuvenation Res. 15, 483–494 (2012).
10. Zhang, W., Qu, J., Liu, G. H. & Belmonte, J. C. I. The ageing epigenome and its rejuvenation. Nat. Rev. Mol. Cell Biol. 21, 137–150 (2020).
11. BrindAmour, I. et al. An ultra-low-input native ChIP-seq protocol for genome-wide profiling of core aging populations. Nat. Commun. 6, 6033 (2015).
12. Brodin, P. et al. Variation in the human immune system is largely driven by non-heritable influences. Cell 160, 37–47 (2015).
13. Trabado, S. et al. The human plasma-metabolome: reference values in 800 French healthy volunteers; impact of cholesterol, gender and age. PLoS ONE 12, e0173615 (2017).
14. Rodriguez-Rodero, S. et al. Aging genetics and aging. Aging Dis. 2, 195–203 (2011).
15. Mitchell, S. J., Scheibe-Knudsen, M., Longo, D. L. & de Cabo, R. Animal models of aging research: implications for human aging and age-related diseases. Ann. Rev. Anim. Biosci. 3, 283–303 (2015).
16. Kim, H. O., Kim, H. S., Youn, J. C., Shin, E. C. & Park, S. Serum cytokine profiles in healthy young and elderly population assessed using multiplexed bead-based immunoassays. J. Transl. Med. 9, 113 (2011).
17. Clark, J. A. & Peterson, T. C. Cytokine production and aging: overproduction of IL-8 in elderly males in response to lipopolysaccharide. Mech. Ageing Dev. 77, 127–139 (1994).
18. Wolf, J. et al. The effect of chronological age on the inflammatory response of human fibroblasts. Exp. Gerontol. 47, 749–753 (2012).
19. Franceschi, C., Garagnani, P., Parini, P., Gismondi, C. & Santoro, A. Inflammaging: a new immune-metabolic viewpoint for age-related diseases. Nat. Rev. Endocrinol. 14, 576–590 (2018).

Received: 10 May 2020; Accepted: 6 October 2020; Published online: 23 November 2020
22. Mahlknecht, U. & Kaiser, S. Age-related changes in peripheral blood counts in humans. Exp. Ther. Med. 1, 1019–1025 (2010).
23. Geiger, H., de Haan, G. & Florian, M. C. The ageing haematopoietic stem cell compartment. Nat. Rev. Immunol. 13, 376–389 (2013).
24. Conte, M. et al. Human aging and longevity are characterized by high levels of mitokines. J. Gerontol. A 74, 600–607 (2018).
25. Tanaka, T. et al. Plasma proteomic signature of age in healthy humans. Aging Cell 17, e12799 (2018).
26. Labrie, F., Belanger, A., Cusan, L., Gomez, J. L. & Candus, B. Marked decline in serum concentrations of adrenal C19 sex steroid precursors and conjugated androgen metabolites during aging. J. Clin. Endocrinol. Metab. 82, 2396–2402 (1997).
27. Liu, Y. et al. Methylinetics of gene expression in human monocytes. Hum. Mol. Genet. 22, 5065–5074 (2013).
28. Reynolds, L. T. et al. Transcriptomic profiling of ages in purified human immune cells. BMC Genomics 16, 333 (2015).
29. Bocklandt, S. et al. Epigenetic predictor of age. PLoS ONE 6, e14821 (2011).
30. Florath, I., Butterbach, K., Müller, H., Bewerunge-Hudler, M. & Brenner, H. Cross-sectional and longitudinal changes in DNA methylation with age: an epigenome-wide analysis revealing over 60 novel age-associated CpG sites. Hum. Mol. Genet. 23, 1186–1201 (2014).
31. Hannum, G. et al. Genome-wide methylation profiles reveal quantitative views of human aging rates. Mol. Cell 49, 359–367 (2013).
32. Horvath, S. DNA methylation age of human tissues and cell lines. Genome Biol. 14, R115 (2013).
33. Horvath, S. et al. Aging effects on DNA methylation modules in human brain and blood tissue. Genome Biol. 13, R97 (2012).
34. Weidner, C. I. et al. Aging of blood can be tracked by DNA methylation changes at just three CpG sites. Genome Biol. 15, R24 (2014).
35. Garrett-Bakelman, F. E. et al. Enhanced reduced representation bisulfite sequencing for assessment of DNA methylation at base pair resolution. J. Vis. Exp. 96, e52246 (2015).
36. Heyn, H. et al. Distinct DNA methylomes of newborns and centenarians. Proc. Natl Acad. Sci. USA 109, 10522–10527 (2012).
37. Wilson, V. L., Smith, R. A., Ma, S. & Cutler, R. G. Genomic aging in human stem and differentiated cells. Nucleic Acids Res. 43, D267–D275 (2018).
38. Labrie, F., Belanger, A., Cusan, L., Gomez, J. L. & Candas, B. Marked increase of epigenetic variations with aging. J. Biol. Chem. 269, 1186–1200 (1994).
39. Slieker, R. C. et al. Age-related accrual of methylomic variability is linked to obesity identifies novel gene changes. Nat. Commun. 7, 157–168 (2016).
40. Mozhui, K. & Pandey, A. K. Conserved effect of aging on DNA methylation regions across multiple tissues in human and the link to complex diseases. Sci. Rep. 5, 15145 (2015).
41. Sershenhichek, A. A. An algorithm for fast preranked gene set enrichment analysis using cumulative statistic calculation. Preprint at bioRxiv https://doi.org/10.1101/060012 (2016).
42. Jha, A. K. et al. Network integration of parallel metabolic and transcriptional data reveals metabolic modules that regulate macrophage polarization. Immunity 42, 419–430 (2015).
43. Derr, A. et al. End sequence analysis toolkit (ESAT) expands the extractable information from single-cell RNA-seq data. Genome Res. 26, 1397–1410 (2016).
44. Storey, J. D. & Tibshirani, R. Statistical significance for genomewide studies. Proc. Natl Acad. Sci. USA 100, 9440–9445 (2003).
45. Ernst, J. & Kellis, M. Chromatin-state discovery and genome annotation with integrative and comparative epigenomics. Bioinformatics 31, D493–D496 (2014).
46. Legrand, C. et al. Statistically robust methylation calling for whole-transcriptome bisulfite sequencing reveals distinct methylation patterns for mouse RNAs. Genome Res. 27, 1589–1596 (2017).
47. Wu, G., Wang, L. G. & He, Q. Y. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. Bioinformatics 31, 2382–2383 (2015).
48. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842 (2010).
49. Boigny, A. R. BEDTools: the Swiss-army tool for genome feature analysis. Curr. Protoc. Bioinformatics 47, 1.12.1–1.12.34 (2014).
50. Thorvaldsdottir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief. Bioinform. 14, 178–192 (2013).
51. Robinson, J. T. et al. Integrative genomics viewer. Nat. Biotechnol. 29, 24–26 (2011).
52. Du, P., Kibbe, W. A. & Lin, S. M. lumi: a pipeline for processing Illumina microarray. Bioinformatics 24, 1547–1548 (2008).
53. Consortium, E. P. An integrated encyclopedia of DNA elements in the human genome. Nature 489, 57–74 (2012).
54. Stranska-Rasch, S., Kurzawa, N., Eils, R. & Herrmann, C. A comprehensive comparison of methods for differential ChIP-seq analysis. Brief. Bioinform. 17, 953–966 (2016).
55. Niedomus-Johnson, J. et al. DNA methylation in lung cells is associated with asthma endotypes and genetic risk. JCI Insight 1, e90151 (2016).
56. Lunnun, K. et al. Methylocific profiling implicates cortical deregulation of ANK1 in Alzheimer’s disease. Nat. Neurosci. 17, 1164–1170 (2014).
57. Perez-Riverol, Y. et al. The PRIDE database and related tools and resources in 2019: improving support for quantification data. Nucleic Acids Res. 47, D442–D450 (2019).

Acknowledgements
The study was supported by funding from the Aging Biology Foundation to the Artyomov laboratory. The Bagaitkar lab is partially supported by GM125504 and DE28296. The Dax lab is supported in part by NIH grants P01AG051459, AI105097, AG051459 and AR070811, the Glenon Foundation on Aging Research and Cure Alzheimer’s Fund. This publication is solely the responsibility of the authors and does not necessarily represent the official view of the National Centre for Research Resources (NCRR) or the National Institutes of Health (NIH). We thank the Genome Technology Access Centre in the Department of Genetics at Washington University School of Medicine for help with genomic analysis. The centre is partially supported by NCI Cancer Centre Support grant number P30 CA91842 to the Siteman Cancer Centre and the NIH Roadmap for Medical Research. We also thank the Epigenomic Core of the Artyomov laboratory. The Bagaitkar lab is partially supported by GM125504 and DE28296. The Dax lab is supported in part by NIH grants P01AG051459, AI105097, AG051459 and AR070811, the Glenon Foundation on Aging Research and Cure Alzheimer’s Fund. This publication is solely the responsibility of the authors and does not necessarily represent the official view of the National Centre for Research Resources (NCRR) or the National Institutes of Health (NIH). We thank the Genome Technology Access Centre in the Department of Genetics at Washington University School of Medicine for help with genomic analysis. The centre is partially supported by NCI Cancer Centre Support grant number P30 CA91842 to the Siteman Cancer Centre and by ICTS/CTSA grant number UL1TR000448 from the NCRR, a component of the NIH, and the NIH Roadmap for Medical Research. We also thank the Epigenomic Core of the Wall Cancer Medicine for the initial analysis of the methylation data (EBRBS and raw data pre-processing). We acknowledge the ENCODE consortium and the ENCODE production laboratories that generated the datasets used in the manuscript. We thank I. Miralda for the Fig. 1 schematic.
Author contributions
I.S., J.B. and O.S. contributed equally. I.S., J.B., O.S., S.D. and M.N.A. designed the project. J.B., D.A.M., E.L., S.P., L.A., A.S. and A.P. performed wet laboratory experiments. I.S., J.B, O.S., E.L., S.P, D.A.M., E.W., P.C., G.D., M.A., K.Z., S.S., C.C., M.B., L.A., A.S., A.P, A.D. E.K., P.T., R.C., V.D.D., M.J., S.D., S.A.S., M.I.D., E.M.O. and M.N.A. participated in data collection and data analysis. I.S., J.B., O.S., G.D, M.A., E.K., P.T. and R.C. made the figures and tables. Writing and editing were done by I.S., J.B., O.S. and M.N.A.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s43587-020-00002-6.

Correspondence and requests for materials should be addressed to M.N.A.

Peer review information Nature Aging thanks the anonymous reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2020
Extended Data Fig. 1 | Cell counts and cytokine profiles. a, Blood cytokine levels measured by bioplex assay and b, blood differentials obtained using Hemavet in young (n = 20) and old (n = 20) cohorts. Normal ranges for humans are shown below the boxplots. P-values for all comparisons were calculated using two-sided Mann-Whitney U test. In both panels, the lower and upper hinges of all boxplots represent the 25th and 75th percentiles. Horizontal bars show median value. Whiskers extend to the values that are no further than 1.5*IQR from either upper or lower hinge. IQR stands for inter-quartile range, which is the difference between the 75th and 25th percentiles.
Extended Data Fig. 2 | See next page for caption.
Extended Data Fig. 2 | Metabolomic profile of plasma. a, PCA of standardized levels of metabolites in plasma. Each dot represents a donor. b, Dendrogram produced by unsupervised hierarchical clustering of metabolic data. Clustering using average algorithm and Euclidian distance as the distance metric. c, List of significantly different metabolites. d, Selected differentially regulated metabolites from sex steroids synthesis pathway in young (n=20) and old (n=20) cohorts. Statistical analyses by two-sided Mann-Whitney U test with Benjamini-Hochberg correction for multiple testing. e, Schema of sex steroids synthesis. f, Pathway analysis of metabolic data. Each boxplot summarizes log2FC of all members of the corresponding pathway. Pathways with mean log2FC significantly different from zero are highlighted (two-sided Mann-Whitney U test and Benjamini-Hochberg correction for multiple testing). Dots represent outliers. N for each pathway is shown in brackets. g, PCA as in Extended Data 2a. Z-scores were calculated for all metabolites. For each sample, color of the dot represents averaged z-scores of all metabolites belonging to the pathway. In panels (D) and (F), the lower and upper hinges of all boxplots represent the 25th and 75th percentiles. Horizontal bars show median value. Whiskers extend to the values that are no further than 1.5*IQR from either upper or lower hinge. IQR stands for inter-quartile range, which is the difference between the 75th and 25th percentiles.
Extended Data Fig. 3 | See next page for caption.
Extended Data Fig. 3 | Proteomic profile of plasma (Somascan). a, PCA of standardized proteins levels in plasma (Somascan). Each dot represents a donor. b, Dendrogram produced by unsupervised hierarchical clustering of Somascan data. Clustering using average algorithm and Euclidian distance as the distance metric. c, Differential analysis results for plasma proteomic profile: volcano plot. Each dot represents a protein. P-values and logFC were calculated by two-sided test from Limma package and adjusted by Benjamini–Hochberg method. d, ELISA validation of selected proteins in young (n = 20) and old (n = 20) cohorts. Adjusted p-values for Somascan results (as in (C)) are shown in the top panel and ELISA validation of the same targets are shown below. ELISA p-values were calculated using two-sided Mann-Whitney U test. The lower and upper hinges of all boxplots represent the 25th and 75th percentiles. Horizontal bars show median value. Whiskers extend to the values that are no further than 1.5*IQR from either upper or lower hinge. IQR stands for inter-quartile range, which is the difference between the 75th and 25th percentiles. e, Heatmap representation of absolute values of Spearman’s correlation coefficients (rho) between plasma proteins (rows) and plasma metabolites (columns) calculated within old (left panel) and young (right panel) cohorts. Clustering of old cohort rows and columns was done using complete algorithm and Euclidian distance as a metric. Order of rows and columns in young cohort heatmap matches order established for the old cohort. f, Each point represents a donor. Smoothing was done by lm function separately in young and old groups, shaded error bands represent SE.
Extended Data Fig. 4 | Monocyte transcriptome and proteome. (a) Left panel: schematic representation of CD14+CD16− monocytes isolation using magnetic beads. Right panel: flow cytometry validation and estimation of purity. (b) Number of significant genes detected after downsampling MESA dataset (youngest and oldest 25%). Downsampling was repeated n = 50 times for each group size. The lower and upper hinges of all boxplots represent the 25th and 75th percentiles. Horizontal bars show median value. Whiskers extend to the values that are no further than 1.5*IQR from either upper or lower hinge. IQR stands for inter-quartile range, which is the difference between the 75th and 25th percentiles. (c) GSEA enrichment curves illustrate pathways that significantly change with age in MESA dataset. P-values are one-sided and corrected by Benjamini-Hochberg method. (d) Monocytes were differentiated into macrophages by one-week incubation with M-CSF. Both cell types were stimulated by LPS for 24 hours. (e) PCA of normalized expression levels estimated by RNA-seq for monocyte differentiation and activation experiment. (f) Each dot represents a monocyte protein significantly different between age groups. LogFC proteomics (x axis) as in Fig. 2f, LogFC transcriptomics as in Fig. 2c.
Extended Data Fig. 5 RRBS quality control and monocyte methylome. **a.** Library depth for each sample. **b.** Hannum and Horvath methylation clocks for old (n = 20) and young (n = 20) groups. Methylation levels of CpGs that were used in the model but were not covered in our eRRBS data were imputed using mean methylation of [−100 kb; +100 kb] region around the CpG. CpG methylation was set to zero if imputation was not possible. P-values were calculated using two-sided Mann-Whitney U test. The lower and upper hinges of all boxplots represent the 25th and 75th percentiles. Horizontal bars show median value. Whiskers extend to the values that are no further than 1.5*IQR from either upper or lower hinge. IQR stands for inter-quartile range, which is the difference between the 75th and 25th percentiles. **c.** Enrichment of DMRs in CpG islands. Histogram shows distribution of simulated intersection sizes (n = 100,000 random simulations). **d.** Comparison to Blueprint dataset. Each dot represents one DMR detected in our dataset. X axis – difference in Blueprint – difference in our dataset. Y axis – difference between old donors and cord blood from Blueprint. **e.** Plot as in right panel of Fig. 3g for a newborn vs centenarian WGBS dataset (GSE31263). **f.** PCA on MESA data as in Fig. 3i using all cytosines profiled by DNA methylation array. **g.** Number of methylated (methylation level > 0) cytosines in CpG, CHG and CHH context shared by one to 40 samples. **h.** PCA of CpGs methylation levels from old and young groups. Each dot represents a sample. **i.** Dendrogram produced by unsupervised hierarchical clustering of the samples. Each sample described as a vector of CpG methylation levels. Clustering using Ward algorithm and Manhattan distance as the distance metric. Outliers are labelled. **j.** Distribution of CpG methylation levels. For each segment fractions of CpGs with corresponding methylation level in each sample are shown by dots. Bar shows average fraction of CpG across all samples. Outlying samples are labelled. **k.** PCA of CpGs methylation levels. Outliers (OD11, OD17, YD9) were excluded.
Extended Data Fig. 6 | See next page for caption.
Extended Data Fig. 6 | ULI-ChIP-seq processing and quality control. a, Snapshot of the H3K4me1 tracks across all donors shows distinct signal for all samples with high variability in the signal-to-noise ratio. b, Number of peaks for each mark in each donor yielded by classical peak calling tools. c, Schematic representation of overlap metric used in panels (E) and (H). d, Number of peaks yielded by SPAN, MACS2 and SICER. e, Overlap between all pairs of samples for peaks generated by SPAN, MACS2 and SICER. SICER was used for wide modifications only. N for each bar is equal to a number of possible pairs between all samples that passed QC. f, Summary for panel (E). Mean and standard deviation (SD) of overlaps between samples are shown. g, Overlap characteristics as in (F) for SPAN runs with various annotation sizes. h, Directional overlap of SPAN peaks between all samples and all histone modifications. i, Two-way overlap with ENCODE CD14+ monocytes data for different peak calling approaches. In panels (B), (D), and (I) N for each bar is equal to a number of ChIP-seq samples that passed QC for each modification. See Fig. 4b for exact numbers. Error bars in all panels represent SD.
Extended Data Fig. 7 | SPAN test set errors. Test set errors for golden standard tools (MACS2, SICER) as well as SPAN trained with various numbers of labels. Each dot represents a sample, n = 40 for H3K4me3, H3K27me3, H3K27ac, n = 32 for H3K4me1, n = 39 for H3K36me3. In all panels, the lower and upper hinges of all boxplots represent the 25th and 75th percentiles. Horizontal bars show median value. Whiskers extend to the values that are no further than 1.5*IQR from either upper or lower hinge. IQR stands for inter-quartile range, which is the difference between the 75th and 25th percentiles.
Extended Data Fig. 8 | See next page for caption.
Extended Data Fig. 8 | Chromatin context of DMRs. 

**a**, Enrichment of DMRs in chromatin state segments from ENCODE ChromHMM partition. Upper: for each chromatin state dark grey bar represents number of DMRs intersecting at least one segment of the state. Light grey bar shows expected number of intersections estimated by \( n = 100,000 \) random simulations. Error bars show SD, their centers represent expected intersection. Results shown for all DMRs, hypermethylated DMRs only and hypomethylated DMRs only. Chromatin states that are significantly over- or under-represented among DMRs are marked by asterix (*). Bottom: heatmap shows general statistics for each chromatin state. Values are normalized within each row. 

**b**, PCA of standardized methylation levels in old and young groups for three chromatin states: bivalent Tss (14 TssBiv), bivalent enhancers (15 EnhBiv) and Polycomb-repressed regions (16 ReprPC). Each dot represents a sample. 

**c**, Intensity of H3K4me1, H3K27me3, H3K4me3, H3K27ac, and H3K36me3 signals was calculated for each DMR via Diffbind package, normalized with respect to the DMR length and averaged across the cohorts. Normalized signals were compared between hypo- (\( n = 423 \)) and hypermethylated (\( n = 737 \)) DMRs using two-sided Mann-Whitney U test. Dots represent outliers. The lower and upper hinges of all boxplots represent the 25th and 75th percentiles. Horizontal bars show median value. Whiskers extend to the values that are no further than 1.5*IQR from either upper or lower hinge. IQR stands for inter-quartile range, which is the difference between the 75th and 25th percentiles. 

**d**, Volcano plot as in Figs. 4i and 4k, colored in accordance with intersection with H3K4me3, H3K27ac and H3K36me3. 

**e**, Density plot protein coding gene expression with transcription factors of interest highlighted.
Extended Data Fig. 9 | Up and down DMRs in different physiological conditions. a, Comparison of age- and sex-adjusted DMR mean methylation between Alzheimer’s patients (n=15 for glia and neurons, n=48 for whole blood) and healthy controls (n=14 for glia and neurons, n=9 for whole blood). Each dot represents one donor. P-values calculated using two-sided Mann-Whitney U test. b, Plot as in (A) comparing data from lean (n=9) and obese (n=8) donors. c, Plot as in Fig. 6g. Mean methylation of smoking DMRs in our cohort: young lean (n=11), young overweight (n=8), old lean (n=8), and old overweight (n=10). P-values were calculated using two-sided Mann-Whitney U test. In all panels, the lower and upper hinges of all boxplots represent the 25th and 75th percentiles. Horizontal bars show median value. Whiskers extend to the values that are no further than 1.5*IQR from either upper or lower hinge. IQR stands for inter-quartile range, which is the difference between the 75th and 25th percentiles.
Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
  
  *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted
  
  *Give P values as exact values wherever suitable.*
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d, Pearson's r), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection

- No software was used

Data analysis

- ULJ-ChIP-seq processing pipeline is available on Github: https://github.com/etalBrains-Research/washu
- Custom tools (SPAN and JBR browser), models and tutorials are available online at https://artyomovlab.wustl.edu/aging/tools.html.
- The R statistical language with various Bioconductor packages as noted in Methods.
  
  STAMP (v2.6.1b)
  FastQC (v0.11.3)
  Picard tools (v2.18.4)
  HTSeq (v0.9.1)
  Methpipe (v3.4.3)
  Bedtools2 (v2.25.0)
  Bowtie (v1.1.3)
  Limma (v3.4.5)
  Pheatmap (1.0.12)
  DeSeq2 (v1.24.0)
  fgsa (v1.10.0)
  BisRNA (v0.2.2)
  ChiPseeker (v1.20.0)
  IGV (v2.3.72)
  lumi (v2.26.4)
  sva (v3.22.0)
  Macs2 (v2.1.1)
  SICER (v2.1)
  DiffBind (v2.4.8)
  differps (v1.55.6)
ChIPDiff (version from March 27, 2008)
Details and references can be found in the relevant Methods sections.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data
All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Raw human sequencing data (Figures 2, 3, and 4) are deposited to Synapse repository (https://www.synapse.org/#/Synapse:syn22020090/wiki/602603) and available for download upon request. Processed sequencing data, as well as raw and processed metabolic and proteomics data are available at the dedicated online portal at https://artyomovlab.wustl.edu/aging/.

Following publicly available datasets have been used in present study:
- GSE56045 MESA transcriptomic dataset -- expression table and annotation are provided. (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE56045)
- GSE56046 MESA DNA methylation dataset -- table of M values and annotation are provided. (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE56046)
- EGAD00001002523 Blueprint dataset -- bigwig files with methylation signal were downloaded. Experiment ids are EGAX00001097775, EGAX00001097772, EGAX000001086967, EGAX000001086968, EGAX000001086970, EGAX00001097774.
- GSE31263 WGBS dataset -- BED files with methylation signal were used. (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE31263)
- ENCODE ChIP-Seq samples were used: GSM1102782, GSM1102785, GSM1102788, GSM1102793, GSM1102797.
- ENCODE ChromHMM annotation ENCSR907LCD was used. (https://www.encodeproject.org/annotations/ENCSR907LCD/)
- ReMap 2018 database was downloaded for hg19. (http://pedagogix-tagc.univ-mrs.fr/remap/download/remap2018/hg19/MACS/remap2018_TF_archive_nr_macs2_hg19_v1_2.tar.gz)

For a list of datasets used in Figure 6 and Extended Data 9 please see “Integration with public data” section of Methods.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences
☐ Behavioural & social sciences
☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size
Sample size was not determined a priori. We did not focus on a specific effect-size and performed a discovery study, including all available samples that passed QC into analysis. To our knowledge, a comparative epigenetic human study of this scale has not been previously undertaken. We discuss potential limitations of our sample size in Results and Discussion sections. Specifically, we show that large cohorts are required for detection of low-degree changes in expression levels (Extended Data 4B).

Data exclusions
1 outlier was excluded from analysis of metabolic data based on its separate location on PCA plot. 3 outliers were excluded from DNA methylation analysis based on PCA plots and dendogram (Extended Data 5H and 5I) as well as irregular distribution of their methylation percentages (Extended Data 5J). Outlier detection based on exploratory data analysis (PCA etc.) was performed as a part of a planned data analysis process. Data from 191 out of 200 ULI-ChIP-seq experiments passed initial QC. See corresponding sections in Methods and Supplementary Information for details.

Replication
Selected targets from proteomics profile were confirmed by ELISA (Extended Data 3D). All proteins besides one were validated. OMD has not been validated by ELISA, possibly, due to high technical variation. All human samples were used in analysis, additional validation cohort was not recruited. Metabolomic profiling was not additionally validated due to literature support for the most major findings. RNA-seq, RRBS and ChIP-seq data was confirmed via comparison to publicly available datasets from GEO, Blueprint, ENCODE.

Randomization
For high-throughput data generation samples were randomized between batches to account for possible batch-effect. Stringent inclusion criteria were set to account for other possible confounding variables. Reproducibility of the identified trends was shown using publicly available data from cohorts of different demographics, showing independence of the identified signatures from potential confounding variables.

Blinding
All data acquisition was performed blinded to types of individual samples.

Reporting for specific materials, systems and methods
We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

| n/a | Involved in the study |
|-----|-----------------------|
| ☑   | Antibodies            |
| ☑   | Eukaryotic cell lines |
| ☑   | Palaeontology         |
| ☑   | Animals and other organisms |
| ☑   | Human research participants |
| ☑   | Clinical data         |

### Antibodies

**Antibodies used**

- H3K27me3 (Millipore 07-449, LOT 2736613, concentration used: 0.3 µg); H3K27Ac (Abcam ab4729, LOT GR312651-1, concentration used: 0.05 µg); H3K4me3 (Abcam ab8580, LOT GR2011420-1, concentration used: 0.03 µg); H3K4me1 (Abcam ab8895, LOT GR271478-1 and GR283603-1, concentration used: 0.2 µg); H3K36me3 (Abcam ab9050, LOT GR260274-1, concentration used: 0.1 µg);
- V450 CD16 (BD Biosciences #561310, Clone B73.1, LOT 6074532); FITC CD14 (BD Biosciences #347493, Clone MΦP9, LOT 6022603);

**Validation**

All antibodies are commercially available, and validation data are provided by the manufacturer.

We additionally validated ChIP-seq antibodies: using public data for CD14+CD16- monocytes in Roadmap Epigenomics Consortium and data pilot runs we developed a collection of positive and negative control primers. These primers were used to do an additional check of signal-to-noise ratio using real-time PCR (SYBR-green) before sequencing step (see Table S15 for sequences and locations).

Examples of previous publications validating the antibodies:

1. H3K27me3 – Prickaerts P, et al., 2016, Epigenetics & Chromatin, 9-46
2. H3K27Ac – Durbin A, et al., 2018, Nature Genetics 50, 1240-1246
3. H3K4me3 – Shah R, et al., 2018, Molecular Cell 72, 162-177
4. H3K4me1 – Shah R, et al., 2018, Molecular Cell 72, 162-177
5. H3K36me3 – Huang Y, et al., 2018, Journal of Biological Chemistry 293 (20), 7811-7823
6. CD14 – Badia R, et al., 2018, Nature Communications 9, 2739
7. CD16 – Maekawa T, et al., 2017, Leukemia 31, 2709-2716

### Human research participants

**Policy information about studies involving human research participants**

**Population characteristics**

Healthy, Caucasian, non-obese (BMI under 30) males were enrolled in the study in two groups (Figure 1A). Young donors between 24-30 years (n=20) and old non-frail donors between 57-70 years (n=20) of age were included. Additional cohort recruited for monocyte proteomic analysis exactly matched all criteria described for the main cohort.

**Recruitment**

Healthy volunteers matching inclusion criteria described above were recruited from St. Louis area, MO, USA. Using a screening questionnaire, subjects were asked about lifestyle and health issues. All subjects self-identified themselves as healthy. Subjects with any previous history of cancer, inflammatory conditions (rheumatoid arthritis, Crohn’s disease, colitis, dermatitis, fibromyalgia or lupus) or infections (HIV, hepatitis B, C) were excluded. Smokers were also excluded. Blood (~100ml) was collected by venous puncture in the morning (8 – 10 a.m.) after overnight fasting. Some of the older donors self-reported blood pressure alterations/medication, which was not considered as a reason for exclusion.

**Ethics oversight**

All human studies were approved by the Washington University in St. Louis School of Medicine Institutional Review Board (IRB-201604138). Written informed consent was obtained from all participants.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

### ChIP-seq

**Data deposition**

- ☑ Confirm that both raw and final processed data have been deposited in a public database such as GEO.
- ☑ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

**Data access links**

Raw human sequencing data are deposited to Synapse repository (https://www.synapse.org/#!Synapse:syn22020090/wiki/602603) and available for download upon request. Processed sequencing data are available at the dedicated online portal at https://artyomovlab.wustl.edu/aging/.
Files in database submission | Please see https://artyomovlab.wustl.edu/aging/download_data.html
---|---
Genome browser session (e.g. UCSC) | Files for UCSC can be uploaded from https://artyomovlab.wustl.edu/aging/explore_data.html

Methodology

| Replicates | 5 histone modifications (H3K4me3, H3K4me1, H3K27ac, H3K27me3, H3K36me3) were profiled in classical monocytes from 20 young and 20 old donors. Consistency between samples is illustrated in Extended Data 6. |
|---|---|
| Sequencing depth | See Supplementary Table 13 for details. 51 bp single-end reads were generated. Average sequencing depth was ~52 million reads. Average number of uniquely aligned reads was ~44 million reads. |
| Antibodies | H3K27me3 (Millipore 07-449, LOT 2736613, concentration used: 0.3 µg); H3K27Ac (Abcam ab4729, LOT GR312651-1, concentration used: 0.05 µg); H3K4me3 (Abcam ab8580, LOT GR3201240-1, concentration used: 0.03 µg); H3K4me1 (Abcam ab8895, LOT GR271478-1 and GR283603-1, concentration used: 0.2 µg); H3K36me3 (Abcam ab9050, LOT GR260274-1, concentration used: 0.1 µg). |
| Peak calling parameters | Peak calling parameters were optimized for each sample using a novel semi-supervised approach. Please see Methods and Supplementary Information for details. Optimized parameters are listed in Supplementary Table 16. |
| Data quality | Please see Methods and Supplementary Information for details. Peak calling summary and optimized parameters are listed in Supplementary Table 16. |
| Software | ULI-ChIP-seq processing pipeline is available on GitHub: https://github.com/JetBrains-Research/washu Custom tools (SPAN and JBR browser), models and tutorials are available online at https://artyomovlab.wustl.edu/aging/tools.html. Bowtie (v1.1.1) |