Molecular Cell
Resource

Genome-wide Methylation Profiles Reveal Quantitative Views of Human Aging Rates

Gregory Hannum,[1,12] Justin Guinney,[5,12] Ling Zhao,[2,3,6] Li Zhang,[2,3,6,7] Guy Hughes,[2,3] SriniVas Sadda,[8] Brandy Klotzle,[9] Marina Bibikova,[9] Jian-Bing Fan,[9] Yuan Gao,[10] Rob Deconde,[1,4] Menzies Chen,[1] Indika Rajapakse,[11] Stephen Friend,[5] Trey Ideker,[1,2,4,*] and Kang Zhang[2,3,6,*]
[1]Department of Bioengineering
[2]Institute for Genomic Medicine
[3]Department of Ophthalmology
[4]Department of Medicine
University of California, San Diego, San Diego, CA 92093, USA
[5]Sage Bionetworks, Seattle, WA 98109, USA
[6]Molecular Medicine Research Center and Department of Ophthalmology, State Key Laboratory of Biotherapy, West China Hospital, Sichuan University, Chengdu 610041, China
[7]Guangzhou iGenomics, Guangzhou 510300, China
[8]Doheny Eye Institute, University of Southern California, Los Angeles, CA 90033, USA
[9]Illumina, San Diego, CA 92122, USA
[10]Lieber Institute, Johns Hopkins University, Baltimore, MD 21205, USA
[11]Division of Basic Sciences and Biostatistics, Fred Hutchinson Cancer Research Center, Seattle, WA 98110, USA
[12]These authors contributed equally to this work
*Correspondence: tideker@ucsd.edu (T.I.), kang.zhang@gmail.com (K.Z.)
http://dx.doi.org/10.1016/j.molcel.2012.10.016

SUMMARY

The ability to measure human aging from molecular profiles has practical implications in many fields, including disease prevention and treatment, forensics, and extension of life. Although chronological age has been linked to changes in DNA methylation, the methylome has not yet been used to measure and compare human aging rates. Here, we build a quantitative model of aging using measurements at more than 450,000 CpG markers from the whole blood of 656 human individuals, aged 19 to 101. This model measures the rate at which an individual's methylome ages, which we show is impacted by gender and genetic variants. We also show that differences in aging rates help explain epigenetic drift and are reflected in the transcriptome. Moreover, we show how our aging model is upheld in other human tissues and reveals an advanced aging rate in tumor tissue. Our model highlights specific components of the aging process and provides a quantitative readout for studying the role of methylation in age-related disease.

INTRODUCTION

Not everyone ages in the same manner. It is well known that women tend to live longer than men, and lifestyle choices such as smoking and physical fitness can hasten or delay the aging process (Austad, 2006; Blair et al., 1989). These observations have led to the search for molecular markers of age that can be used to predict, monitor, and provide insight into age-associated physiological decline and disease. One such marker is telomere length, a molecular trait strongly correlated with age (Harley et al., 1990) that has been shown to have an accelerated rate of decay under environmental stress (Epel et al., 2004; Valdes et al., 2005). Another marker is gene expression, especially for genes that function in metabolic and DNA repair pathways, which are predictive of age across a range of different tissue types and organisms (Fraser et al., 2005; Zahn et al., 2007; de Magalhães et al., 2009).

A growing body of research has reported associations between age and the state of the epigenome—the set of modifications to DNA other than changes in the primary nucleotide sequence (Fraga and Esteller, 2007). In particular, DNA methylation associates with chronological age over long time scales (Alisch et al., 2012; Christensen et al., 2009; Bollati et al., 2009; Boks et al., 2009; Rakyan et al., 2010; Bocklandt et al., 2011; Bell et al., 2012), and changes in methylation have been linked to complex age-associated diseases such as metabolic disease (Barres and Zierath, 2011) and cancer (Jones and Laird, 1999; Esteller, 2008). Studies have also observed a phenomenon dubbed "epigenetic drift," whereby the DNA methylation marks in identical twins increasingly differ as a function of age (Fraga et al., 2005; Boks et al., 2009). Thus, the idea of the epigenome as a fixed imprint is giving way to the model of the epigenome as a dynamic landscape that reflects a variety of chronological changes. The current challenge is to determine whether these changes can be systematically described and modeled to detect different rates of human aging, and to tie these rates to related clinical or environmental variables.

The mechanisms that drive changes in the aging methylome are not well understood, although they have been attributed to at least

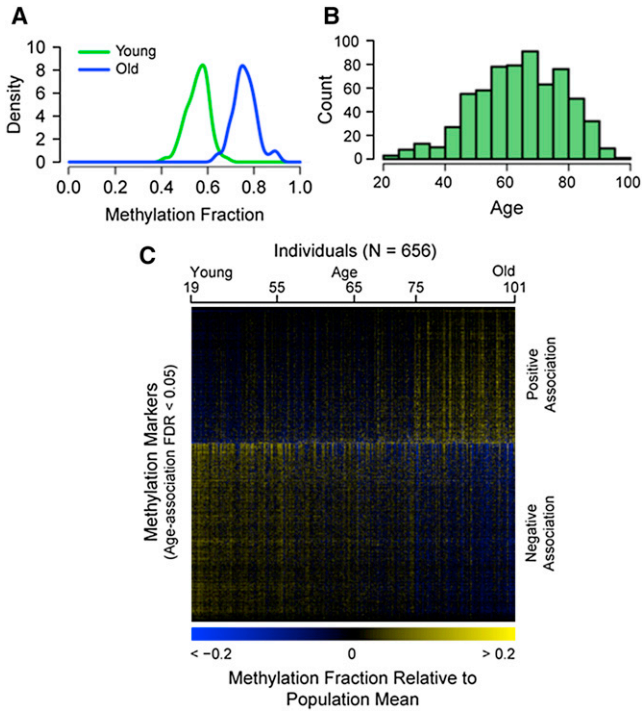

Figure 1. Global Data on the Aging Methylome

(A) A density plot of methylation fraction values for the marker cg16867657, separated by young (green) and old (blue) individuals.

(B) A histogram of the age distribution for all individuals.

(C) A heatmap of the age-associated methylation markers, sorted by the magnitude of association (regression coefficient). The individuals are ordered youngest to oldest.

See also Figure S1 and Tables S1 and S2 for a specific example of an age-associated region and for annotation coincidence tables, respectively.

two underlying factors (Vijg and Campisi, 2008; Fraga et al., 2005). First, it is possible that environmental exposure will over time activate cellular programs associated with consistent and predictable changes in the epigenome. For example, stress has been shown to alter gene expression patterns through specific changes in DNA methylation (Murgatroyd et al., 2009). Alternatively, spontaneous epigenetic changes may occur with or without environmental stress, leading to fundamentally unpredictable differences in the epigenome between aging individuals. Spontaneous changes may be caused by chemical agents that disrupt DNA methyl groups or through errors in copying methylation states during DNA replication. Both mechanisms lead to differences between the methylomes of aging individuals, suggesting that quantitative measurements of methylome states may identify factors involved with slowed or accelerated rates of aging.

To better understand how the methylome ages and to determine whether human aging rates can be quantified and compared, we initiated a project to perform genome-wide methylomic profiling of a large cohort of individuals spanning a wide age range. Based on these findings, we constructed a predictive model of aging rate which we show is influenced by gender and specific genetic variants. These data help explain epigenetic drift and suggest that age-associated changes in the methylome lead to changes in transcriptional patterns over time. These findings were replicated in a second large cohort.

RESULTS AND DISCUSSION

Global Methylomic Profiling over a Wide Age Range

We obtained methylome-wide profiles of two different cohorts (N_1 = 482, N_2 = 174) sampled from a mixed population of 426 Caucasian and 230 Hispanic individuals, aged 19 to 101. Samples were taken as whole blood and processed with the Illumina Infinium HumanMethylation450 BeadChip assay (Bibikova et al., 2011), which measures the methylation states of 485,577 CpG markers. Methylation was recorded as a fraction between zero and one, representing the frequency of methylation of a given CpG marker across the population of blood cells taken from a single individual. Conservative quality controls were applied to filter spurious markers and samples (Experimental Procedures). For simplicity, we discarded values for markers on sex chromosomes. Association tests revealed that 70,387 (15%) of the markers had significant associations between methylation fraction and age (Figure 1, false discovery rate [FDR] < 0.05 by F test, Experimental Procedures). We were able to verify at a $p < 0.05$ significance level 53,670 (76%) of these associations using 40 young and old samples recently published by Heyn et al. (2012). More detailed accounts of the individual aging markers and their genomic features are presented in the Supplemental Information (Figure S1 and Tables S1 and S2). The resulting data set represents the largest and highest-resolution collection of methylation data produced for the study of aging, providing an unprecedented opportunity to understand the role of epigenetics in the aging process. The complete methylation profiles are available at the Gene Expression Omnibus (GSE40279).

A Predictive Model for the Aging Methylome

We built a predictive model of aging on the primary cohort using a penalized multivariate regression method known as Elastic Net (Zou and Hastie, 2005), combined with bootstrap approaches (Experimental Procedures). The model included both methylomic and clinical parameters such as gender and body mass index (BMI) (Figure 2A). The optimal model selected a set of 71 methylation markers that were highly predictive of age (Figure 2A and Table S3). The accuracy of the model was high, with a correlation between age and predicted age of 96% and an error of 3.9 years (Figure 2B). Nearly all markers in the model lay within or near genes with known functions in aging-related conditions, including Alzheimer's disease, cancer, tissue degradation, DNA damage, and oxidative stress. By way of example, two markers lay within the gene somatostatin (*SST*), a key regulator of endocrine and nervous system function (Yacubova and Komuro, 2002). SST is known to decline with age and has been linked to Alzheimer's disease (Saito et al., 2005). As a second example, six model markers lay within the transcription factor *KLF14*, which has been called a "master regulator" of obesity and other metabolic traits (Small et al., 2011). Given the links between aging, longevity, and metabolic activity (Lane et al., 1996; Tatar et al., 2003), it is not surprising that several of our model markers are implicated in obesity and metabolism.

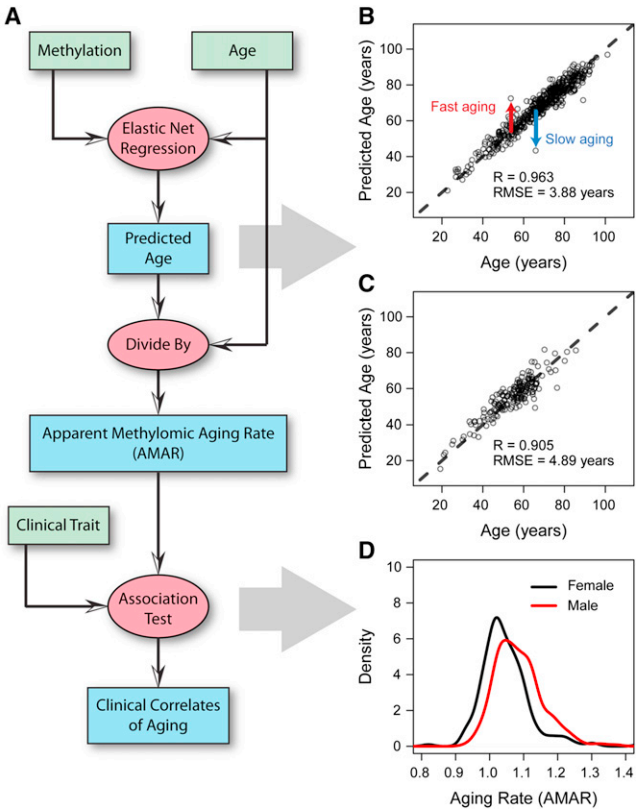

Figure 2. Model Predictions and Clinical Variables

(A) A flow chart of the data (green boxes) and analyses (red ovals) used to generate aging predictions (blue boxes).

(B) A comparison of predicted and actual ages for all individuals based on the aging model.

(C) Out-of-sample predictions for individuals in the validation cohort.

(D) Apparent methylomic aging rate (AMAR) for each individual, based on the aging model without clinical variables. The distribution of aging rates shows faster aging for men than women. A table of the markers used in the aging model is provided in Table S3.

See also Figures S2 and S3 and Table S3.

We validated this model on the secondary cohort, consisting of an additional 174 independent samples. These samples were processed in the same manner as the primary cohort and were then used to predict age based on the original model (i.e., as trained on the original cohort). The predictions were highly accurate, with a correlation between age and predicted age of 91% and an error of 4.9 years (Figure 2C). The significance of the aging model was also confirmed by the data set presented in Heyn et al., verifying the age association of 70 of the 71 markers (Heyn et al., 2012). Furthermore, the model was able to fully separate old and young individuals in the Heyn et al. study, even for profiles obtained via bisulfate sequencing rather than the bead-chip technology used in this study (Figure S2).

Methylome Aging Rate and Its Associations

While the aging model is able to predict the age of most individuals with high accuracy, it is equally valuable as a tool for identifying individual outliers who do not follow the expectation. For example, Figure 2B highlights two individuals whose age is vastly over- or underpredicted on the basis of their methylation data. To examine whether these differences reflect true biological differences in the state of the individual (i.e., versus measurement error or intrinsic variability), we used the aging model to quantify each individual's *apparent methylomic aging rate* (AMAR), defined as the ratio of the predicted age, based on methylation data, to the chronological age. We then tested for associations between AMAR and possibly relevant clinical factors, including gender and BMI. Analysis of ethnicity and diabetes status was not possible due to correlations with batch variables (Figure S3). We found that gender, but not BMI had significant contributions to aging rate (F test, $p = 6 \times 10^{-6}$, $p > 0.05$, Experimental Procedures). The methylome of men appeared to age approximately 4% faster than that of women (Figure 2D), even though the overall distributions of age were not significantly different between the men and women in the cohort ($p > 0.05$, KS test). Likewise, the validation cohort confirmed the increased aging rate for men ($p < 0.05$), but was inconclusive for BMI ($p > 0.05$). This complements a previous finding of an epigenetic signal for BMI that does not change with age (Feinberg et al., 2010).

As genetic associations have been previously reported with human longevity and aging phenotypes (Atzmon et al., 2006; Suh et al., 2008; Willcox et al., 2008; Wheeler et al., 2009), we examined whether the model could distinguish aging rates for individuals with different genetic variants. For this purpose, we obtained whole-exome sequences for 252 of the individuals in our methylome study at 15× coverage. After sequence processing and quality control, these sequences yielded 10,694 common single-nucleotide variants across the population (Experimental Procedures). As a negative control, we confirmed that none of the genetic variants were significant predictors of age itself, which is to be expected since the genome sequence is considered to be relatively static over the course of a lifetime. On the other hand, one might expect to find genetic variants that modulate the methylation of age-associated markers, i.e., methylation quantitative-trait loci or meQTLs (Bell et al., 2011). Testing each genetic variant for association with the top age-associated methylation markers, we identified 303 meQTLs (Experimental Procedures, FDR < 0.05, Figure 3A). For validation, we selected eight genetic variants (corresponding to 14 meQTLs) to test in a validation cohort of 322 individuals from our methylation study. This analysis found that seven of eight genetic variants (corresponding to seven meQTLs) remained highly significant in the validation cohort (FDR < 0.05, Table S4). While all of these variants acted in *cis* with their meQTLs (within 150 kbp), we confirmed that none directly modified the CpG site or associated probe sequence of the associated methylation marker.

The methylation marker cg27193080 was one of those found to be significantly associated with age ($p < 10^{-17}$), and its methylation fraction was found to be influenced by the single-nucleotide polymorphism (SNP) variant rs140692 ($p < 10^{-21}$) (Figure 3B). This meQTL was particularly interesting as both the SNP and the methylation marker mapped to the gene *methyl-CpG binding domain protein 4* (*MBD4*, with the SNP in an intron and the methylation marker just upstream of the coding region), one of the few known genes encoding a protein that can bind to methylated

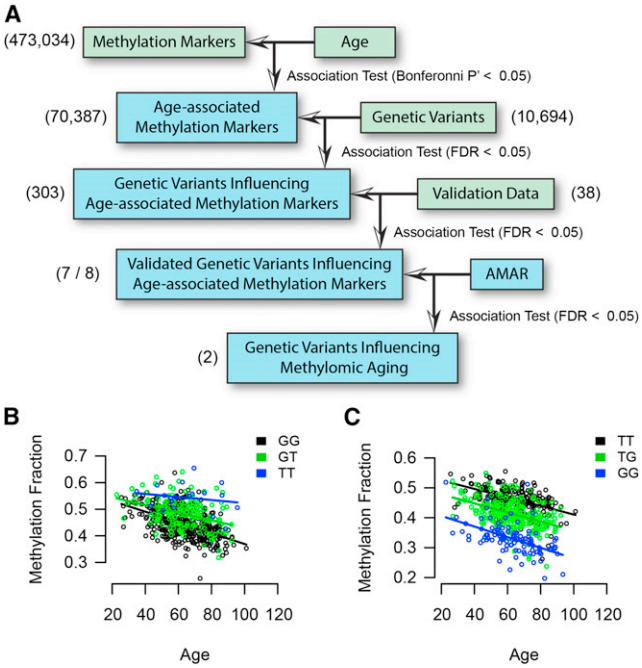

Figure 3. Genetic Effects on Methylomic Aging

(A) We surveyed genomic variants for an association with age-associated methylation markers. Eight genetic variants, corresponding to 14 meQTLs, were chosen for validation. Of these, seven were significant in the validation cohort and two showed an association with AMAR.

(B) A plot of the trend between the methylation marker cg27367526 (STEAP2) and age. The state of variant rs42663 (GTPBP10) causes an offset in this relationship.

(C) A second example for cg18404041 and rs2230534 (ITIH1, NEK4).

See also Table S4 for a table of confirmed genetic associations.

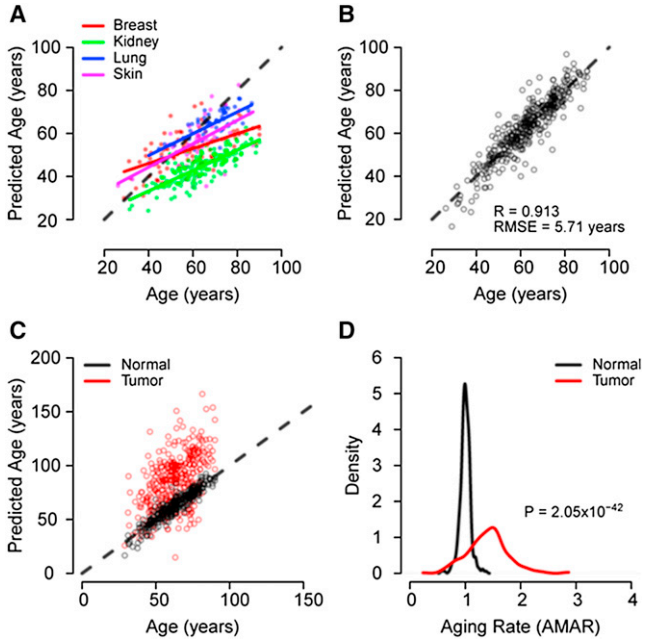

Figure 4. Multitissue Support

(A) Predictions of age made by the full aging model on the TCGA control samples. There is a high correlation between chronological and predicted age, but each tissue has a different linear intercept and slope.

(B) After adjusting the intercept and slope of each tissue, the error of the model is similar to that of the original whole-blood data. Age predictions made on cancer samples are presented in Figure S2.

(C) Age predictions made on matched normal and tumor samples from TCGA. Predictions are adjusted for the linear offset of the parent tissue (breast, kidney, lung, or skin).

(D) Tumor samples show a significant increase in AMAR.

See also Figure S4 and Table S5.

DNA. This meQTL thus captures a *cis* relationship in which rs140692 influences the methylation state of *MBD4*. That *MBD4* plays a role in human aging is supported by previous work linking *MBD4* to DNA repair, as well as work showing that mutations and knockdowns of *MBD4* lead to increased genomic instability (Bellacosa et al., 1999; Bertoni et al., 2009).

Of the seven validated meQTLs, three were identified that had a statistically significant association not only with age but also with aging rate (AMAR, FDR < 0.05, Figures 3B and 3C). One is the genetic marker rs2230534, which is a synonymous mutation in the gene *NEK4*, and has a *cis* association with the methylation marker cg18404041. The *NEK* family of kinases plays a key role in cell-cycle regulation and cancer (Moniz et al., 2011). The second variant is rs2818384, which is a synonymous mutation in the gene *JAKMIP3* and has a *cis* association with the methylation marker cg05652533. Copy-number variants in *JAKMIP3* have been previously associated with glioblastoma (Xiong et al., 2010). The final variant found to influence AMAR is rs42663, which is a missense mutation in the gene *GTPBP10* and associates with cg27367526 in the gene *STEAP2*. *STEAP2* is known to play a role in maintaining homeostasis of iron and copper—metals that serve as essential components of the mitochondrial respiratory chain (Ohgami et al., 2006). Studies have shown that perturbations of iron concentrations can induce DNA damage through oxidative stress in mammalian cells (Hartwig and Schlepegrell, 1995; Karthikeyan et al., 2002). These meQTLs represent genetic variants that appear to broadly influence the aging methylome and may be good candidates for further age-associated disease and longevity research.

A Multitissue Diagnostic

Our aging model was derived from whole blood, which is advantageous in the design of practical diagnostics and for testing samples collected from other studies. To investigate whether our aging model was representative of other tissues, we obtained DNA methylation profiles for 368 individuals in the control category of The Cancer Genome Atlas (TCGA) (Collins and Barker, 2007), including 83 breast, 183 kidney, 60 lung, and 42 skin samples. An aging model based on both our primary and validation cohorts demonstrated strong predictive power for chronological age in these samples (expected value R = 0.72), although each tissue had a clear linear offset (intercept and slope) from the expectation (Figure 4A). This offset was consistent within a tissue, even across different batches of the TCGA data. We adjusted for each tissue trend using a simple linear model, producing age predictions with an error comparable to that found in blood (Figure 4B). Furthermore, predicted AMARs

in each tissue supported the effect of men appearing to age more quickly than women ($p < 0.05$). Thus, computation of aging rate (AMAR) from blood samples reflects trends that are not specific to blood and may be common throughout many tissues of the human body. Furthermore, this analysis provides evidence that the observed methylomic changes are intrinsic to the methylome and not due primarily to cell heterogeneity, i.e., changing cell-type composition of whole blood with age. In this regard, this study is consistent with a prior analysis of purified CD4+ T cells and CD14+ monocytes, in which the age-associated epigenetic modifications were found to be similar to the changes observed in whole blood (Rakyan et al., 2010).

To investigate the similarities and differences between the tissues, we built age models de novo for breast, kidney, and lung tissues (Table S5; the skin cohort had too few samples to build a model). Most of the markers in the models differed, although all of these models and the primary model share the markers cg23606718 and cg16867657. These markers are both annotated to the gene *ELOVL2*, which has been linked to the photoaging response in human skin (Kim et al., 2010).

The TCGA data set also contains methylome profiles representing a total of 319 tumors and matched normal tissue samples (breast, kidney, lung, and skin). Interestingly, use of our aging model indicated that tumors appear to have aged 40% more than matched normal tissue from the same individual (Wilcox test, $p < 10^{-41}$, Figures 4C and 4D). Accelerated tumor aging was apparent regardless of the primary tissue type. We investigated whether this was the result of broad shifts in global methylation levels by examining all 70,387 age-associated markers, of which 44% tend to increase and 56% tend to decrease with age. Methylation fraction values in matched tumor and normal samples supported the finding that tumors coincide with older values for 74% of the markers regardless of the trending direction (binomial $p \sim 0$). Furthermore, separate aging models built in the matched normal and tumor samples confirm the apparent aging effect (Figure S4).

Different Aging Rates Lead to Divergent Methylomes

If individuals indeed age at different rates, it might be expected that their individual methylomes should diverge over time. This is based on the premise that the methylomes of the very young share certain similarities and that these similarities diminish as individuals accumulate changes over time. This effect, called epigenetic drift, has been observed in monozygotic twins (Fraga et al., 2005), but few specific hypothesis have been put forth to account for it. To examine epigenetic drift in our samples, we computed the deviance of each methylation marker value as its squared distance from the expected population mean (Figure 5A and the Experimental Procedures). Then, in addition to testing for markers whose methylation fraction changes with age (Figures 5B and 5C), we were able to test for markers whose deviance changes with age (Figures 5D and 5E) (Breusch and Pagan, 1979). Increasing deviance was a widespread phenomenon—we identified 27,800 markers for which the deviance was significantly associated with age (FDR < 0.05), of which 27,737 (99.8%) represented increased rather than decreased deviance (Figures 5E and S5). For any given individual, especially high or low methylome deviance was a strong predictor of aging rate ($R = 0.47$, $p \sim 0$), suggesting that differences in aging rates account for part of methylome heterogeneity and epigenetic drift.

Another way to examine epigenetic drift is in terms of Shannon entropy, or loss of information content in the methylome over time (Shannon and Weaver, 1963). An increase in entropy of a CpG marker means that its methylation state becomes less predictable across the population of cells, i.e., its methylation fraction tends toward 50% (Experimental Procedures). Indeed, over all markers associated with a change in methylation fraction in the sample cohort, 70% tended toward a methylation fraction of 50% (Figure 6A, binomial $p \sim 0$, Table S2). Consequently, we observed a highly significant increase in methylome entropy over the sample cohort ($R = 0.21$, $p < 10^{-7}$). Furthermore, extreme methylome entropy for an individual was highly correlated with accelerated aging rate based on AMAR ($R = 0.49$, $p \sim 0$, Figure 6B).

Aging Rates and the Transcriptome

As changes in methylation have been directly linked to changes in gene expression (Sun et al., 2011), we were interested in whether these changes in the aging methylome were mirrored on a functional level in the human transcriptome and reflected differences in aging rates. For this purpose, we obtained and analyzed publicly available gene expression profiles from the whole blood of 488 individuals spanning an age range of 20 to 75 (Emilsson et al., 2008). We found strong evidence for genes whose expression associates with age (326 genes, FDR < 0.05) and for genes with increasing expression deviance (binomial $p < 10^{-276}$, Experimental Procedures). Strikingly, we found that genes with age-associated expression profiles were more likely to have nearby age-associated methylation markers in our data ($p < 0.01$, Table S6). We used this information to build a model of aging based on the expression of genes that were associated with age in the methylome (Figure 7A, Table S7, and the Experimental Procedures). This model demonstrated a clear ability to measure aging rate using expression data, reproducing our finding of increased aging rates for men as compared to women (Figure 7B, 11% difference, $p < 10^{-4}$). The gender effect was not present in a model built using all available genes rather than those associated with age-related changes in the methylome ($p > 0.05$). Thus, age-associated changes to the methylome are indicative of functional changes in gene expression patterns.

Conclusions

In this study, we have shown that genome-wide methylation patterns represent a strong and reproducible biomarker of biological aging rate. These patterns enable a quantitative model of the aging methylome that demonstrates high accuracy and an ability to discriminate relevant factors in aging, including gender and genetic variants. Moreover, our ability to apply this model in multiple tissues suggests the possibility of a common molecular clock, regulated in part by changes in the methylome. It remains to be seen whether these changes occur on an intracellular level uniformly across a population of cells, or reflect consistent changes in tissue composition over time.

The ability to predict age from whole blood may permit a wider analysis in longitudinal studies such as the Framingham Study,

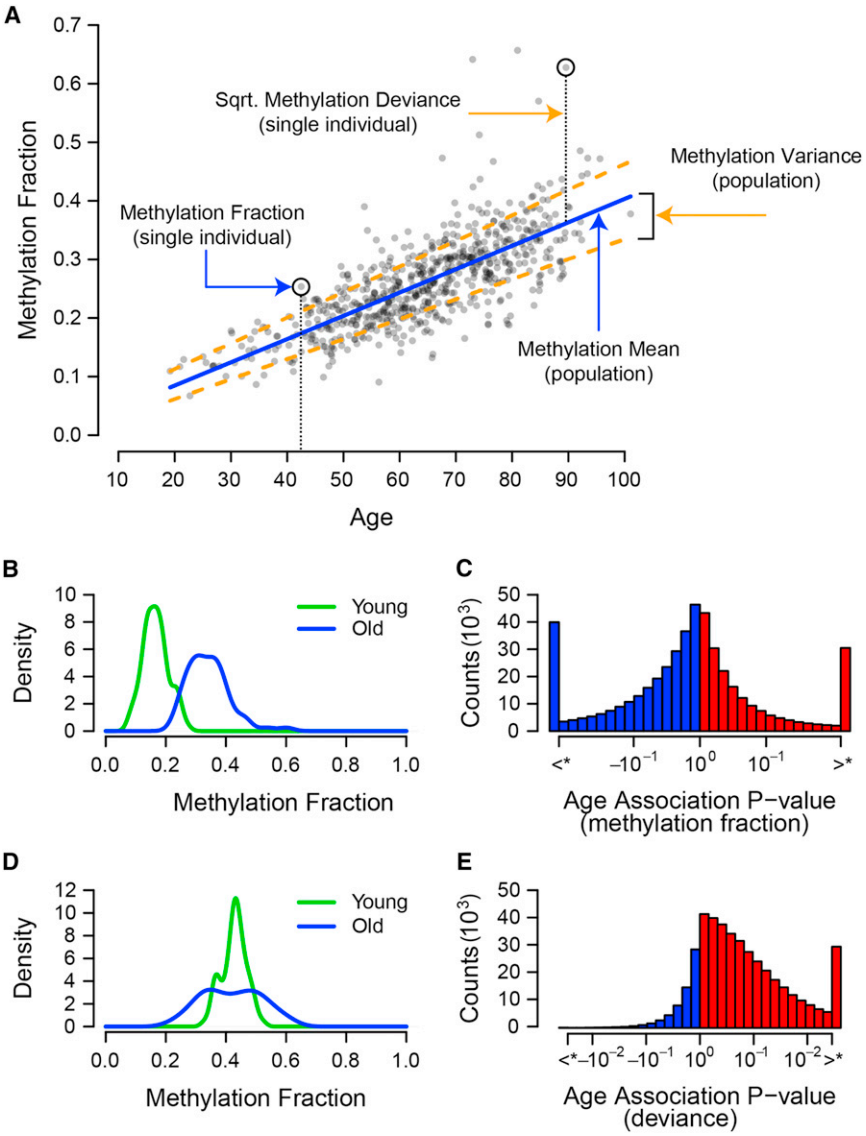

Figure 5. Age Associations for Methylation Fraction and Deviance

(A) Methylation fraction values for are shown for the marker cg24724428. Over any subset of the cohort, we consider two group methylation statistics: the mean and variance. Marker variance is a measure of the mean methylation deviance, which is defined as the squared difference between an individual's methylation fraction and their expected methylation fraction.

(B) A density plot showing the change in mean methylation with age for the marker cg24724428. Young and old groups are based on the top and bottom 10%.

(C) A histogram of the significance of association between the methylation fraction of all markers and age. p values are signed such that positive values represent an increase of methylation with age. Markers that exceeded the FDR < 0.05 threshold are grouped into the most extreme bins.

(D) A density plot showing the change in methylation deviance with age for the marker cg24724428.

(E) A histogram, in the same form as (D), of the significance of association between the methylation deviance of all markers and age. Aging trends are mapped for CpG islands in Figure S3.

See also Figure S5.

the Women's Health Initiative, blood samples collected on neonatal Guthrie cards, and other longitudinal studies with rich annotation of biometric and disease traits. Aging trends could emerge from such studies with many potential practical implications, from health assessment and prevention of disease to forensic analysis. Similar to the effect of gender in this study, the identification of additional biometric or environmental factors that influence AMAR, such as smoking, alcohol consumption, or diet, will permit quantitative assessments of their impacts on health and longevity. A useful example would be to periodically assess the rate of aging of an individual using AMAR and determine whether diet or environmental factors can accelerate or retard the aging process and diseases such as age related macular degeneration. As models of human aging improve, it is conceivable that biological age, as measured from molecular profiles, might one day supersede chronological age in the clinical evaluation and treatment of patients.

EXPERIMENTAL PROCEDURES

Sample Collection and Test Procedures

This study was approved by the institutional review boards of the University of California, San Diego; the University of Southern California; and West China Hospital. All participants signed informed consent statements prior to participation. Blood was drawn from a vein in the patient's arm into blood collection tubes containing the anticoagulant acid citrate dextrose. Genomic DNA was extracted from the whole blood with a QIAGEN FlexiGene DNA Kit and stored at −20°C. Methylation fraction values for the autosomal chromosomes were measured with the Illumina Infinium HumanMethylation450 BeadChip (Bibikova et al., 2011). This procedure uses bisulfate-treated DNA and two site-specific probes for each marker, which bind to the associated methylated and unmethylated sequences. The intensity of the methylated probe relative to the total probe intensity for each site represents the fractional level of methylation at that site in the sample. These values were adjusted for internal controls with Illumina's Genome Studio software. Methylation fraction values with a detection p value greater than 0.01 were set to "missing." One patient sample and 830 markers were removed as they had greater than 5% missing values. The remaining missing values were imputed with the KNN approach (ten nearest markers) using the R "impute" package (Troyanskaya et al., 2001). We performed exome sequencing on 258 of these samples, using a solution hybrid selection method to capture DNA followed by parallel sequencing on an Illumina HiSeq platform. Genotype calls were made with the SOAP program (Li et al., 2008). Calls with a quality score less than twenty were set as missing. Only variants that had fewer than 10% missing calls, were within Hardy-Weinberg equilibrium ($p \leq 10^{-4}$), and were of a common frequency (>5%) were retained (10,694). Individuals with less than 20% missing calls (252) were retained. Additional genotyping was done with multiplex PCR followed by MALDI-TOF mass spectrometry analysis with the iPLEX/MassARRAY/Typer platform.

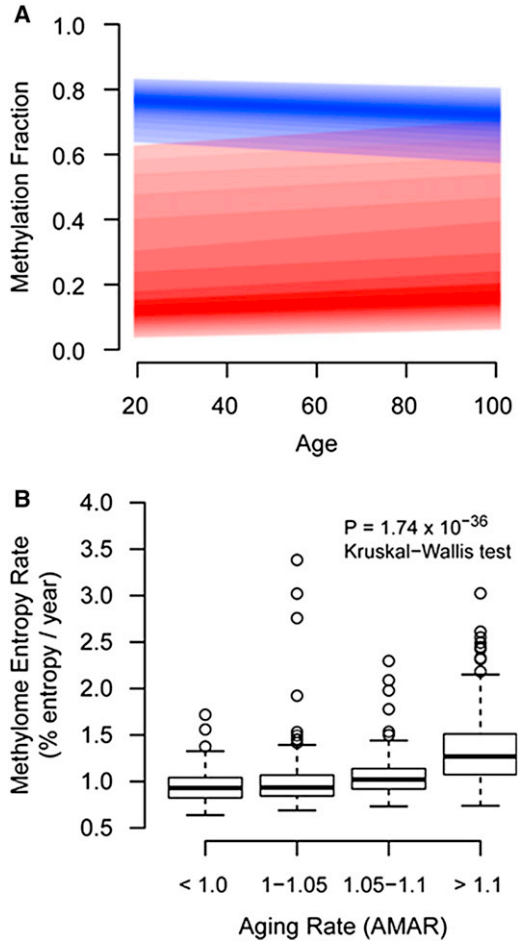

Figure 6. Methylome-wide Trends with Age

(A) Aggregate regression lines for all methylation markers that increased with age (red) and decreased with age (blue). The darkest color represents the median regression line and the bounds represent the 25% and 75% quantile. Both increasing and decreasing markers trend toward moderate methylation fraction values.

(B) An entropy aging rate was calculated as the mean Shannon entropy of age-associated methylation markers divided by chronological age. This was strongly associated with AMAR.

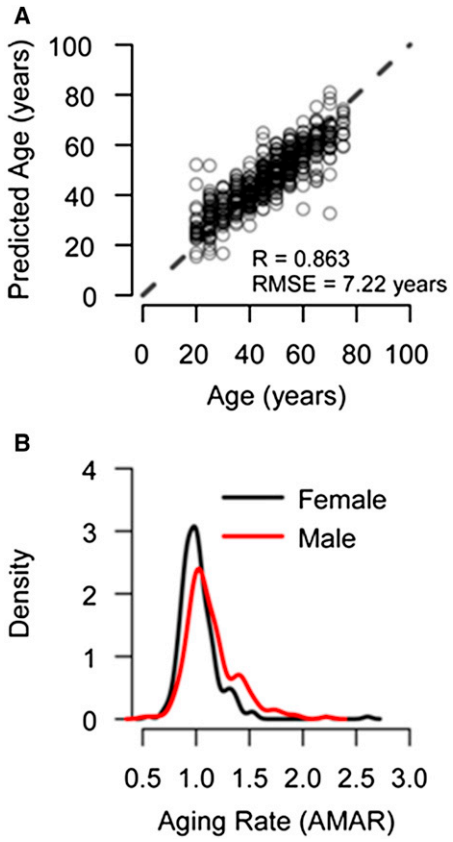

Figure 7. Transcription Aging Model

(A) We built an aging model using mRNA expression data for genes that showed an aging trend in the methylome. Its standard error (RMSE = 7.22 years) is increased due to the rounding of ages to the nearest 5 year interval in the data set.

(B) Similar to the methylome, the transcriptome shows an increased aging rate for men as compared to women ($p < 10^{-4}$).

See also Table S6 and Table S7.

Methylation Quality Control

We used principal component (PC) analysis to identify and remove outlier samples. We converted each sample into a z score statistic, based on the squared distance of its 1st PC from the population mean. The z statistic was converted to a false-discovery rate with the Gaussian cumulative distribution and the Benjamini-Hochberg procedure (Benjamini and Hochberg, 1995). Samples falling below an FDR of 0.2 were designated at outliers and removed. This filtering procedure was performed iteratively until no samples were determined to be an outlier. A total of 24 samples were removed in this manner.

Association Testing

Association tests for trends in methylation fraction and deviance were performed with nested linear models and the F test. As methylation levels may be sensitive to a number of factors, we included several covariates, including gender, BMI, diabetes status, ethnicity, and batch. Tests for whole-methylome changes in deviance were computed with the binomial test, based on the number of markers with a positive rather than negative coefficient. Markers were annotated as having support from the TCGA data when the coefficient of aging was the same sign and the significance was better than $p < 0.05$.

Annotation Enrichment

Methylation marker annotations for CpG islands and GO terms were obtained from the IlluminaHumanMethylation450k.db database from Bioconductor (Gentleman et al., 2004). Annotation enrichment tests were performed with the two-sided Fisher's exact test.

Aging Model

The diagnostic model of age was made with a multivariate linear model approach based on the Elastic Net algorithm implemented in the R package "glmnet" (Friedman et al., 2010). This approach is a combination of traditional Lasso and ridge regression methods, emphasizing model sparsity while appropriately balancing the contributions of correlated variables. It is ideal for building linear models in situations where the number of variables (markers) greatly outweighs the number of samples. Optimal regularization parameters were estimated via 10-fold crossvalidation. We employed bootstrap analysis, sampling the data set with replacement 500 times and building a model for each bootstrap cohort. We included in the final model only markers that were present in more than half of all bootstraps. The covariates gender, BMI, diabetes status, ethnicity, and batch were included in the model and were exempted from penalization (regularization). p values are based on

a least-squares model built with the same terms and drop-one F tests. As BMI was strongly associated with age, the term was first adjusted for age before computing significance in the model. AMAR was computed with the aging model, but without the variables of gender, BMI, and diabetes status. The coefficients were not changed. AMAR was then taken as an individual's predicted age divided by her or his actual age.

Genetic Variant Associations

Each genetic variant was tested for association in an additive model with the top aging-associated methylation markers with nested linear models and the F test. We included covariates for gender, BMI, diabetes status, ethnicity, and batch. Variant positions were based on the human reference build GRCh37 and gene annotations were based on chromosomal proximity within 20 kbp.

Computing Methylation Deviance

Methylation deviance was computed via the following approach: First, we removed the methylation trends due to all given variables, including age, gender, and BMI by fitting a linear model for each marker and acting only on the residuals. Next, we identified and removed highly nonnormal markers on the basis of the Shapiro-Wilk test ($p < 10^{-5}$). To allow for naturally occurring extreme deviations in the normality test, we first estimated the outliers of each marker based on a Grubb's statistic, choosing the inclusion threshold based on the Benjamini-Hochberg FDR (Benjamini and Hochberg, 1995). If any samples had an FDR less than 0.4, we ignored them and repeated the outlier detection until no outliers were detected. Finally, the deviance of each remaining marker was computed as the square of its adjusted methylation value.

Entropy Analysis

Entropy statistics were computed on methylation data adjusted for covariates and filtered for normality (see Computing Methylation Deviance). We computed the normalized Shannon entropy (Shannon and Weaver, 1963) of an individual's methylome according to the formula

$$Entropy = \frac{1}{N * \log\left(\frac{1}{2}\right)} \sum_i [MFi * \log(MFi) + (1 - MFi) * \log(1 - MFi)],$$

where *MFi* is the methylation fraction of the i^{th} methylation marker and N is the number of markers.

Mapping CpG Islands

Genomic positions and marker annotations for 27,176 CpG islands were obtained from the IlluminaHumanMethylation450k.db database from Bioconductor (Gentleman et al., 2004). We obtained the positions for markers within each island with at least four markers (25,028), as well as the nearest 100 markers upstream and downstream. These positions were then combined with the marker value of interest (i.e., methylation fraction, aging coefficient, or deviance) to produce a genomic map for each island and the surrounding region. After normalizing each map to the center of the island, we averaged the values at each relative genomic point across all islands to produce a common map.

ACCESSION NUMBERS

The complete methylation profiles have been deposited in NCBI's Gene Expression Omnibus under accession number GSE40279.

SUPPLEMENTAL INFORMATION

Supplemental Information includes five figures and seven tables and can be found with this article online at http://dx.doi.org/10.1016/j.molcel.2012.10.016.

ACKNOWLEDGMENTS

We thank Janusz Dutkowski, Kumar Sharma, and Mariano Alvarez for critical discussions and Daniel O'Conner for reviewing the manuscript. This work is supported by grants from the 973 Program (2013CB967504), NSFC (grant 81130017), and NIH (grants EY014428, EY018660, EY019270, EY021374, P50GM085764, and R01E5014811), a grant from the King Abdulaziz City for Science and Technology through the UC San Diego Center of Excellence in Nanomedicine center, and the Burroughs Wellcome Fund Clinical Scientist Award in Translational Research. This work is a product of the Sage Federation, a consortium of research labs whose goal is to encourage greater openness and sharing of biomedical data and analyses. L. Zhao, L. Zhang, G. Hughes, S.S, and Y.G. collected and processed samples with guidance from K.Z.; B.K., M.B., and J.F. performed the methylation assays; L. Zhao, L. Zhang, and K. Z. performed exome sequencing and genotyping; and G. Hannum and J.G. performed the principal statistical analyses with guidance from T.I., R.D., M.C., and S.F. I.R. discussed the entropy metric. G. Hannum, J.G., T.I., Y.G., and K.Z. wrote the manuscript. B.K, M.B, and J.F. work for Illumina.

Received: June 11, 2012
Revised: July 24, 2012
Accepted: October 10, 2012
Published online: November 21, 2012

REFERENCES

Alisch, R.S., Barwick, B.G., Chopra, P., Myrick, L.K., Satten, G.A., Conneely, K.N., and Warren, S.T. (2012). Age-associated DNA methylation in pediatric populations. Genome Res. *22*, 623–632.

Atzmon, G., Rincon, M., Schechter, C.B., Shuldiner, A.R., Lipton, R.B., Bergman, A., and Barzilai, N. (2006). Lipoprotein genotype and conserved pathway for exceptional longevity in humans. PLoS Biol. *4*, e113.

Austad, S.N. (2006). Why women live longer than men: sex differences in longevity. Gend. Med. *3*, 79–92.

Barres, R., and Zierath, J.R. (2011). DNA methylation in metabolic disorders. Am. J. Clin. Nutr. *93*, 897S–900.

Bell, J.T., Pai, A.A., Pickrell, J.K., Gaffney, D.J., Pique-Regi, R., Degner, J.F., Gilad, Y., and Pritchard, J.K. (2011). DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. Genome Biol. *12*, R10. http://www.ncbi.nlm.nih.gov/pubmed/21251332.

Bell, J.T., Tsai, P.-C., Yang, T.-P., Pidsley, R., Nisbet, J., Glass, D., Mangino, M., Zhai, G., Zhang, F., Valdes, A., et al.; MuTHER Consortium. (2012). Epigenome-wide scans identify differentially methylated regions for age and age-related phenotypes in a healthy ageing population. PLoS Genet. *8*, e1002629.

Bellacosa, A., Cicchillitti, L., Schepis, F., Riccio, A., Yeung, A.T., Matsumoto, Y., Golemis, E.A., Genuardi, M., and Neri, G. (1999). MED1, a novel human methyl-CpG-binding endonuclease, interacts with DNA mismatch repair protein MLH1. Proc. Natl. Acad. Sci. USA *96*, 3969–3974.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J. R. Stat. Soc. B *57*, 289–300.

Bertoni, C., Rustagi, A., and Rando, T.A. (2009). Enhanced gene repair mediated by methyl-CpG-modified single-stranded oligonucleotides. Nucleic Acids Res. *37*, 7468–7482.

Bibikova, M., Barnes, B., Tsan, C., Ho, V., Klotzle, B., Le, J.M., Delano, D., Zhang, L., Schroth, G.P., Gunderson, K.L., et al. (2011). High density DNA methylation array with single CpG site resolution. Genomics *98*, 288–295.

Blair, S.N., Kohl, H.W., 3rd, Paffenbarger, R.S.J., Jr., Clark, D.G., Cooper, K.H., and Gibbons, L.W. (1989). Physical fitness and all-cause mortality. A prospective study of healthy men and women. JAMA *262*, 2395–2401.

Bocklandt, S., Lin, W., Sehl, M.E., Sánchez, F.J., Sinsheimer, J.S., Horvath, S., and Vilain, E. (2011). Epigenetic predictor of age. PLoS ONE *6*, e14821.

Boks, M.P., Derks, E.M., Weisenberger, D.J., Strengman, E., Janson, E., Sommer, I.E., Kahn, R.S., and Ophoff, R.A. (2009). The relationship of DNA methylation with age, gender and genotype in twins and healthy controls. PLoS ONE *4*, e6767.

Bollati, V., Schwartz, J., Wright, R., Litonjua, A., Tarantini, L., Suh, H., Sparrow, D., Vokonas, P., and Baccarelli, A. (2009). Decline in genomic DNA methylation through aging in a cohort of elderly subjects. Mech. Ageing Dev. *130*, 234–239.

Breusch, T.S., and Pagan, A.R. (1979). A Simple Test for Heteroscedasticity and Random Coefficient Variation. Econometrica *47*, 1287.

Christensen, B.C., Houseman, E.A., Marsit, C.J., Zheng, S., Wrensch, M.R., Wiemels, J.L., Nelson, H.H., Karagas, M.R., Padbury, J.F., Bueno, R., et al. (2009). Aging and environmental exposures alter tissue-specific DNA methylation dependent upon CpG island context. PLoS Genet. *5*, e1000602.

Collins, F.S., and Barker, A.D. (2007). Mapping the cancer genome. Pinpointing the genes involved in cancer will help chart a new course across the complex landscape of human malignancies. Sci. Am. *296*, 50–57.

de Magalhães, J.P., Curado, J., and Church, G.M. (2009). Meta-analysis of age-related gene expression profiles identifies common signatures of aging. Bioinformatics *25*, 875–881.

Emilsson, V., Thorleifsson, G., Zhang, B., Leonardson, A.S., Zink, F., Zhu, J., Carlson, S., Helgason, A., Walters, G.B., Gunnarsdottir, S., et al. (2008). Genetics of gene expression and its effect on disease. Nature *452*, 423–428.

Epel, E.S., Blackburn, E.H., Lin, J., Dhabhar, F.S., Adler, N.E., Morrow, J.D., and Cawthon, R.M. (2004). Accelerated telomere shortening in response to life stress. Proc. Natl. Acad. Sci. USA *101*, 17312–17315.

Esteller, M. (2008). Epigenetics in cancer. N. Engl. J. Med. *358*, 1148–1159.

Feinberg, A.P., Irizarry, R.A., Fradin, D., Aryee, M.J., Murakami, P., Aspelund, T., Eiriksdottir, G., Harris, T.B., Launer, L., Gudnason, V., and Fallin, M.D. (2010). Personalized epigenomic signatures that are stable over time and covary with body mass index. Sci. Transl. Med. *2*, 49ra67.

Fraga, M.F., and Esteller, M. (2007). Epigenetics and aging: the targets and the marks. Trends Genet. *23*, 413–418.

Fraga, M.F., Ballestar, E., Paz, M.F., Ropero, S., Setien, F., Ballestar, M.L., Heine-Suñer, D., Cigudosa, J.C., Urioste, M., Benitez, J., et al. (2005). Epigenetic differences arise during the lifetime of monozygotic twins. Proc. Natl. Acad. Sci. USA *102*, 10604–10609.

Fraser, H.B., Khaitovich, P., Plotkin, J.B., Pääbo, S., and Eisen, M.B. (2005). Aging and gene expression in the primate brain. PLoS Biol. *3*, e274.

Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. J. Stat. Softw. *33*, 1–22.

Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., et al. (2004). Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. *5*, R80.

Harley, C.B., Futcher, A.B., and Greider, C.W. (1990). Telomeres shorten during ageing of human fibroblasts. Nature *345*, 458–460.

Hartwig, A., and Schlepegrell, R. (1995). Induction of oxidative DNA damage by ferric iron in mammalian cells. Carcinogenesis *16*, 3009–3013.

Heyn, H., Li, N., Ferreira, H.J., Moran, S., Pisano, D.G., Gomez, A., Diez, J., Sanchez-Mut, J.V., Setien, F., Carmona, F.J., et al. (2012). Distinct DNA methylomes of newborns and centenarians. Proc. Natl. Acad. Sci. USA *109*, 10522–10527.

Jones, P.A., and Laird, P.W. (1999). Cancer epigenetics comes of age. Nat. Genet. *21*, 163–167.

Karthikeyan, G., Lewis, L.K., and Resnick, M.A. (2002). The mitochondrial protein frataxin prevents nuclear damage. Hum. Mol. Genet. *11*, 1351–1362.

Kim, E.J., Kim, M.-K., Jin, X.-J., Oh, J.-H., Kim, J.E., and Chung, J.H. (2010). Skin aging and photoaging alter fatty acids composition, including 11,14,17-eicosatrienoic acid, in the epidermis of human skin. J. Korean Med. Sci. *25*, 980–983.

Lane, M.A., Baer, D.J., Rumpler, W.V., Weindruch, R., Ingram, D.K., Tilmont, E.M., Cutler, R.G., and Roth, G.S. (1996). Calorie restriction lowers body temperature in rhesus monkeys, consistent with a postulated anti-aging mechanism in rodents. Proc. Natl. Acad. Sci. USA *93*, 4159–4164.

Li, R., Li, Y., Kristiansen, K., and Wang, J. (2008). SOAP: short oligonucleotide alignment program. Bioinformatics *24*, 713–714.

Moniz, L., Dutt, P., Haider, N., and Stambolic, V. (2011). Nek family of kinases in cell cycle, checkpoint control and cancer. Cell Div. *6*, 18.

Murgatroyd, C., Patchev, A.V., Wu, Y., Micale, V., Bockmühl, Y., Fischer, D., Holsboer, F., Wotjak, C.T., Almeida, O.F.X., and Spengler, D. (2009). Dynamic DNA methylation programs persistent adverse effects of early-life stress. Nat. Neurosci. *12*, 1559–1566.

Ohgami, R.S., Campagna, D.R., McDonald, A., and Fleming, M.D. (2006). The Steap proteins are metalloreductases. Blood *108*, 1388–1394.

Rakyan, V.K., Down, T.A., Maslau, S., Andrew, T., Yang, T.-P., Beyan, H., Whittaker, P., McCann, O.T., Finer, S., Valdes, A.M., et al. (2010). Human aging-associated DNA hypermethylation occurs preferentially at bivalent chromatin domains. Genome Res. *20*, 434–439.

Saito, T., Iwata, N., Tsubuki, S., Takaki, Y., Takano, J., Huang, S.-M., Suemoto, T., Higuchi, M., and Saido, T.C. (2005). Somatostatin regulates brain amyloid beta peptide Abeta42 through modulation of proteolytic degradation. Nat. Med. *11*, 434–439.

Shannon, C.E., and Weaver, W. (1963). The Mathematical Theory of Communication (Champaign, IL: University of Illinois Press).

Small, K.S., Hedman, A.K., Grundberg, E., Nica, A.C., Thorleifsson, G., Kong, A., Thorsteindottir, U., Shin, S.-Y., Richards, H.B., Soranzo, N., et al.; GIANT Consortium; MAGIC Investigators; DIAGRAM Consortium; MuTHER Consortium. (2011). Identification of an imprinted master trans regulator at the KLF14 locus related to multiple metabolic phenotypes. Nat. Genet. *43*, 561–564.

Suh, Y., Atzmon, G., Cho, M.-O., Hwang, D., Liu, B., Leahy, D.J., Barzilai, N., and Cohen, P. (2008). Functionally significant insulin-like growth factor I receptor mutations in centenarians. Proc. Natl. Acad. Sci. USA *105*, 3438–3442.

Sun, Z., Asmann, Y.W., Kalari, K.R., Bot, B., Eckel-Passow, J.E., Baker, T.R., Carr, J.M., Khrebtukova, I., Luo, S., Zhang, L., et al. (2011). Integrated analysis of gene expression, CpG island methylation, and gene copy number in breast cancer cells by deep sequencing. PLoS ONE *6*, e17490.

Tatar, M., Bartke, A., and Antebi, A. (2003). The endocrine regulation of aging by insulin-like signals. Science *299*, 1346–1351.

Troyanskaya, O., Cantor, M., Sherlock, G., Brown, P., Hastie, T., Tibshirani, R., Botstein, D., and Altman, R.B. (2001). Missing value estimation methods for DNA microarrays. Bioinformatics *17*, 520–525.

Valdes, A.M., Andrew, T., Gardner, J.P., Kimura, M., Oelsner, E., Cherkas, L.F., Aviv, A., and Spector, T.D. (2005). Obesity, cigarette smoking, and telomere length in women. Lancet *366*, 662–664.

Vijg, J., and Campisi, J. (2008). Puzzles, promises and a cure for ageing. Nature *454*, 1065–1071.

Wheeler, H.E., Metter, E.J., Tanaka, T., Absher, D., Higgins, J., Zahn, J.M., Wilhelmy, J., Davis, R.W., Singleton, A., Myers, R.M., et al. (2009). Sequential use of transcriptional profiling, expression quantitative trait mapping, and gene association implicates MMP20 in human kidney aging. PLoS Genet. *5*, e1000685.

Willcox, B.J., Donlon, T.A., He, Q., Chen, R., Grove, J.S., Yano, K., Masaki, K.H., Willcox, D.C., Rodriguez, B., and Curb, J.D. (2008). FOXO3A genotype is strongly associated with human longevity. Proc. Natl. Acad. Sci. USA *105*, 13987–13992.

Xiong, M., Dong, H., Siu, H., Peng, G., Wang, Y., and Jin, L. (2010). Genome-Wide Association Studies of Copy Number Variation in Glioblastoma. Proceedings of the 4th International Conference on Bioinformatics and Biomedical Engineering (iCBBE), 1–4.

Yacubova, E., and Komuro, H. (2002). Stage-specific control of neuronal migration by somatostatin. Nature *415*, 77–81.

Zahn, J.M., Poosala, S., Owen, A.B., Ingram, D.K., Lustig, A., Carter, A., Weeraratna, A.T., Taub, D.D., Gorospe, M., Mazan-Mamczarz, K., et al. (2007). AGEMAP: a gene expression database for aging in mice. PLoS Genet. *3*, e201.

Zou, H., and Hastie, T. (2005). Regularization and variable selection via the elastic net. J. R. Stat. Soc. Series B Stat. Methodol. *67*, 301–320.