Distinct DNA methylomes of newborns and centenarians

Holger Heyn[a,1], Ning Li[b,c,1], Humberto J. Ferreira[a,d], Sebastian Moran[a], David G. Pisano[e], Antonio Gomez[a], Javier Diez[a], Jose V. Sanchez-Mut[a], Fernando Setien[a], F. Javier Carmona[a], Annibale A. Puca[f,g], Sergi Sayols[a], Miguel A. Pujana[h], Jordi Serra-Musach[h], Isabel Iglesias-Platas[i], Francesc Formiga[j], Agustin F. Fernandez[k], Mario F. Fraga[k,l], Simon C. Heath[m], Alfonso Valencia[e], Ivo G. Gut[m], Jun Wang[n,o,2], and Manel Esteller[a,p,q,2]

[a]Cancer Epigenetics and Biology Program; [h]Biomarkers and Susceptibility Unit, Spanish Biomedical Research Centre Network for Epidemiology and Public Health; Catalan Institute of Oncology, Bellvitge Biomedical Research Institute, L'Hospitalet, Barcelona, Catalonia 08908, Spain; [b]Beijing Genomics Institute, Shenzhen, Guangdong, China; [c]Beijing Genomics Institute-Europe, DK-2200 Copenhagen N, Denmark; [d]Programme in Experimental Biology and Biomedicine, Centre for Neurosciences and Cell Biology, University of Coimbra, Portugal; [e]Structural Biology and Biocomputing Programme, Spanish National Cancer Research Centre, Madrid, Spain; [f]Unit of Genetics, Cardiovascular Research Institute, Istituto Ricovero Cura Carattere Scientifico Multimedica, Sesto S. Giovanni, Italy; [g]Facolta' di medicina, Universita' di Salerno, Baronissi, Italy; [i]Neonatal Unit, Hospital Sant Joan de Déu, Fundació Sant Joan de Déu, Barcelona University, Barcelona, Catalonia, Spain; [j]Internal Medicine Service, University Hospital of Bellvitge, L'Hospitalet, Barcelona, Catalonia, Spain; [k]Instituto Universitario de Oncología del Principado de Asturias, Hospital Universitario Central de Asturias, Universidad de Oviedo, Oviedo, Spain; [l]Department of Immunology and Oncology, Centro Nacional de Biotecnología, Madrid, Spain; [m]Centre Nacional d'Anàlisi Genòmica, Barcelona, Catalonia, Spain; [n]The Novo Nordisk Foundation Center for Basic Metabolic Research, [o]Department of Biology, University of Copenhagen, Denmark; [p]Department of Physiological Sciences II, School of Medicine, University of Barcelona, Barcelona, Catalonia, Spain; and [q]Institucio Catalana de Recerca i Estudis Avançats, Barcelona, Catalonia, Spain

Edited by Peter A. Jones, University of Southern California, Los Angeles, CA, and accepted by the Editorial Board May 11, 2012 (received for review December 14, 2011)

Human aging cannot be fully understood in terms of the constrained genetic setting. Epigenetic drift is an alternative means of explaining age-associated alterations. To address this issue, we performed whole-genome bisulfite sequencing (WGBS) of newborn and centenarian genomes. The centenarian DNA had a lower DNA methylation content and a reduced correlation in the methylation status of neighboring cytosine—phosphate—guanine (CpGs) throughout the genome in comparison with the more homogeneously methylated newborn DNA. The more hypomethylated CpGs observed in the centenarian DNA compared with the neonate covered all genomic compartments, such as promoters, exonic, intronic, and intergenic regions. For regulatory regions, the most hypomethylated sequences in the centenarian DNA were present mainly at CpG-poor promoters and in tissue-specific genes, whereas a greater level of DNA methylation was observed in CpG island promoters. We extended the study to a larger cohort of newborn and nonagenarian samples using a 450,000 CpG-site DNA methylation microarray that reinforced the observation of more hypomethylated DNA sequences in the advanced age group. WGBS and 450,000 analyses of middle-age individuals demonstrated DNA methylomes in the crossroad between the newborn and the nonagenarian/centenarian groups. Our study constitutes a unique DNA methylation analysis of the extreme points of human life at a single-nucleotide resolution level.

epigenomics | longevity

During human aging, progressive impairment of organ and tissue functionality leads to an increasing probability of death. The molecular culprits behind this decline in physiological activities remain largely unknown. Studies of transcriptional and genomic associations in distinct tissues have identified several gene families and cellular pathways that might contribute to aging and alter lifespan. These families include the Sirtuins, DNA repair enzymes, insulin-signaling pathway/forkhead transcription factors, apolipoproteins, telomere biology, and oxidative damage/mitochondrial metabolism (1, 2). Aging-associated mechanisms apparently involve many networks within a given cell. Considering that epigenetic regulation has emerged as a critical driver of cell fate and survival that targets many pathways (3, 4), that epigenetic drift can occur even in genetically identical humans (5, 6), and that DNA methylation patterns are disrupted in a wide range of common human diseases (7–11), we wondered whether individuals at the most extreme points of their lifespan had different DNA methylomes. To address this issue, we used whole-genome bisulfite sequencing (WGBS) (12–16) and a 450,000 CpG DNA methylation microarray to examine the DNA methylation profiles of newborn and nonagenarian/centenarian samples.

Results and Discussion

WGBS of Newborn and Centenarian DNA. The initial data were generated from the cord blood of a newborn (male Caucasian; NB) and from a centenarian (103-y-old male Caucasian; Y103) using DNA extracted from $CD4^+$ T cells processed through an Illumina Genome Analyzer. The use of the Illumina HumanOmni5-Quad BeadChip, which covers 4,301,332 SNPs, demonstrated the absence of aneuploidy in the samples studied (Fig. S1). None of the identified differential SNPs was associated with DNA methyltransferases or methyl-group metabolic enzymes.

For WGBS, we generated 575M and 576M raw reads resulting in 51.79 (NB) and 51.85 (Y103) raw Gb of paired-end sequence data. Of these, 44.48 (NB) and 43.55 (Y103) Gb (86.68% and 85.27%, respectively) were successfully aligned to either strand of the reference genome (HG19), providing an average 13.4-fold (NB) and 13.06-fold (Y103) sequencing depth. Of all cytosines present in the reference genome sequence, 92.2% (NB) and 91.6% (Y103) of Cs and 93.9% (NB) and 93.5% (Y103) of CGs were covered. Based on alignment with in silico converted non-CpG cytosines, the bisulfite conversion rate was determined to be at

Author contributions: H.H. and M.E. designed research; H.H. performed research; H.H., N.L., H.J.F., S.M., D.G.P., A.G., J.D., J.V.S.-M., F.S., F.J.C., A.A.P., S.S., M.A.P., J.S.-M., I.I.-P., F.F., A.F.F., M.F.F., S.C.H., A.V., I.G.G., J.W., and M.E. analyzed data; and H.H. and M.E. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. P.A.J. is a guest editor invited by the Editorial Board.

Freely available online through the PNAS open access option.

Data deposition: The data reported in this paper have been deposited in the Gene Expression Omnibus (GEO) database, www.ncbi.nlm.nih.gov/geo [accession nos. GSE31438 (DNA SNP microarray data), GSE31263 (GSE31438DNA methylation sequencing data from NB, Y26 and Y103), GSE30870 (DNA methylation microarray data from newborns and nonagenarians), and GSE33233 (DNA methylation microarray data from middle-age group).

[1]H.H. and N.L. contributed equally to this work.

[2]To whom correspondence may be addressed. E-mail: wangj@genomics.org.cn or mesteller@idibell.cat.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1120658109/-/DCSupplemental.

least 99.34% (NB) and 99.24% (Y103), even when assuming that all non-CpG methylcytosines were derived from conversion failure; this ensured reliable ascertainment of CpG methylcytosines with a false-positive rate of < 0.5%. The full set of WGBS data from NB and Y103 is illustrated in Fig. 1*A*, using Circos (17).

We observed that the centenarian DNA had 494,595 less methylated CpG dinucleotides (mCpGs) than did the newborn DNA (16,775,090 vs. 16,280,495 on the Watson and Crick strand, respectively) (Fig. 1*B*). Moreover, the average methylation level of all covered CpG sites was lower in the centenarian sample (73.0%) than in the newborn one (80.5%) (Fig. 1*B*). This finding was observed for all chromosomes (Fig. 1*A*). Strikingly, when we developed WGBS (51.60 Gb; 574M raw reads; 82.7% mapped) of an additional sample of intermediate age (26-y-old Caucasian; Y26) (Fig. 1*A*), we detected intermediate CpG methylation levels (77.8%) and also an intermediate number of methylated CpGs (16,708,411) (Fig. 1*B*). Illustrative CpG methylation levels for single chromosomes are shown in Fig. 1*C*. The centenarian DNA (Y103) also showed a lower correlation in terms of the methylation status of nearby CpGs than the neonate sample (Wilcoxon test, $P < 2.2 \times 10^{-16}$) (Fig. 1*D*). The biological significance of this relative lack of neighbor site methylation concordance in the centenarian vs. the newborn might relate to a deficient targeting by the enzymes that add (DNA methyltransferases) or remove (tet methylcytosine dioxygenases) methy-groups, although its stochastic nature or occurrence across particular chromatin domains should be elucidated in future studies. Interestingly, the curve of the distance correlation between closest CpG sites of the WGBS of the intermediate age sample (Y26) appeared between the newborn and centenarian DNAs (Fig. 1*D*). From a functional genomics standpoint, the unmethylated CpGs observed in the centenarian DNA covered all genomic compartments, such as promoters (± 2 kb around the transcription start site), exonic, intronic, and intergenic regions (for each region: Wilcoxon test, $P < 2.2 \times 10^{-16}$) (Fig. 1*E*). When we subclassified the promoters according to their CpG content, we found that the centenarian DNA compared with the newborn sample exhibited more unmethylated CpGs at CpG-poor promoters (Wilcoxon test, $P < 2.2 \times 10^{-16}$) and more methylated CpGs at CpG island promoters (Wilcoxon test, $P = 1.721 \times 10^{-07}$) (Fig. 1*F*). CpG sites in CpG island "shores," extensively studied regions of comparatively low CpG density within 2 kb of CpG islands (15, 18, 19), also mostly had a more unmethylated status in the Y103 sample (Wilcoxon-test, $P < 2.2 \times 10^{-16}$) (Fig. S2).

Identification of Differentially Methylated Regions. The discrepancies between the newborn and centenarian DNA methylomes prompted us to search for particular differentially methylated regions (DMRs). We identified DMRs by considering all non-polymorphic CpG dinucleotides (those that presented an identical genotype at this position in both analyzed samples) and searching for maximal length sequences of consecutive CpGs with a consistent direction of methylation change between the two samples for all CpGs in the region, and where the first and last CpG dinucleotides of the region containing at least five CpGs had a χ^2 value of >3.84 for differential methylation, corresponding to a value of $P < 0.05$ for both residues. Using these criteria, we identified 17,930 DMRs between NB and Y103; these are illustrated for chromosome 2 in Fig. 2*A* using Circos (17) and are listed in detail in Dataset S1. The DMRs represent 617,338 CpG sites in the reference genome that are distributed across all human chromosomes (Fig. S3). DMRs were identified in promoter (10.2%), exonic (10.1%), intronic (45.3%), and intergenic (34.4%) regions (Fig. 2*B*). Among these latter sequences, the most common DMR-associated intergenic sequences were LINE 1 repeats (3,777), Alu repeats (3,242), mammalian interspersed repeats (MIRs) (1,890), mammalian apparent LTR-retrotransposons (1,448), and LINE 2 repeats (1,321) (Fig. S4*A*). These data are consistent with previous observations using candidate-sequence approaches that show an aging-progressive loss of DNA methylation in Alu and LINE-1 repetitive elements (20, 21). Quantitative real-time RT-PCR analyses for Alu (22) and MIR transcripts, showed an increase in the expression of these repetitive sequences in the Y103 sample compared with the newborn (Fig. S4*B*). Interestingly, 36% (6,482 of 17,930) of the identified DMRs significantly (Permutation test; P value $< 2.2 \times 10^{-16}$) occurred in lamina-associated domains of the genome (23), where cancer-type–specific DMRs have also been characterized (15, 16). Most importantly, only 0.37% ($n = 66$) of the 17,930 DMRs occurred in a region of copy number variation (CNV) between NB and Y103 detected by the 4,301,332 SNPs genomic microarray. Respectively, 80 and 71

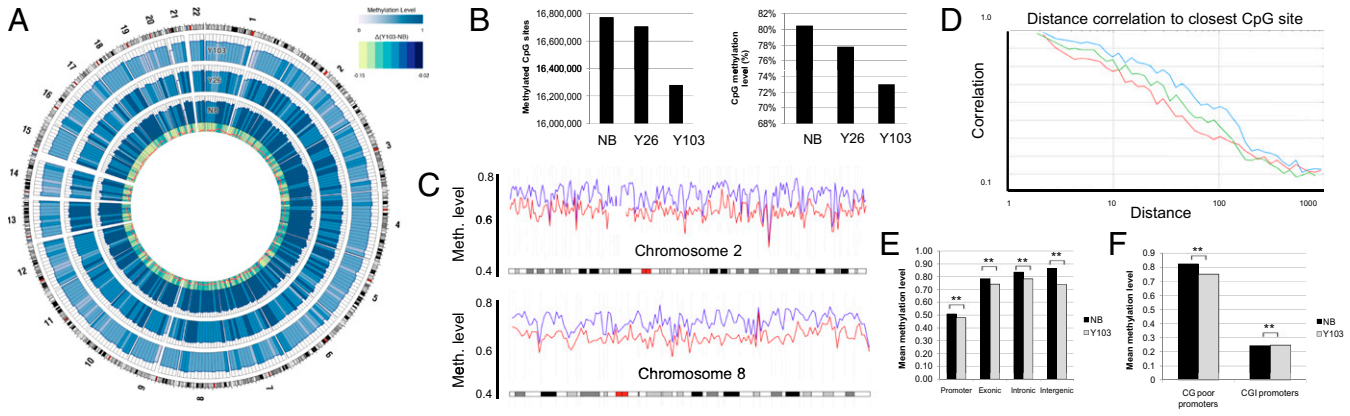

Fig. 1. WGBS of a newborn (NB) and a centenarian (Y103) individual. (*A*) Circos representation of genome-wide DNA methylation levels in the NB, Y26, and Y103 individuals. Average levels for all of the CGs in 297 10-Mbp-wide windows. Inner track indicates the magnitude of the difference between the Y103 and the NB individual for each window (color scale and red line). Average methylation levels in all of the regions are expressed as β-values (0–1) and are colored blue. (*B*) Total number of methylated CpG sites and the CpG methylation level (%) in the DNA from the newborn (NB), an intermediate 26-y-old sample (Y26), and the centenarian sample (Y103). (*C*) Illustrative CpG methylation levels for Y103 (red line) and NB (blue line) in chromosomes 2 and 8. (*D*) The curves of the distance correlation between the methylation status of neighboring CpG sites of NB (blue line), Y26 (green line), and Y103 (red line) samples determined by WGBS. A more pronounced declining curve indicates lower correlation in terms of the methylation status of nearby CpGs. (*E*) Mean CpG methylation levels among different genomic sequences in the centenarian and the newborn individuals. (*F*) Mean CpG methylation levels among promoters in the centenarian and the newborn individuals according to the presence or absence of a CpG island. $^{**}P < 0.01$ in the Fisher's exact *t* test. The mean methylation level is calculated by the number of total methylated reads divided by the numbers of total reads covering CpG sites located in the sum of the feature analyzed.

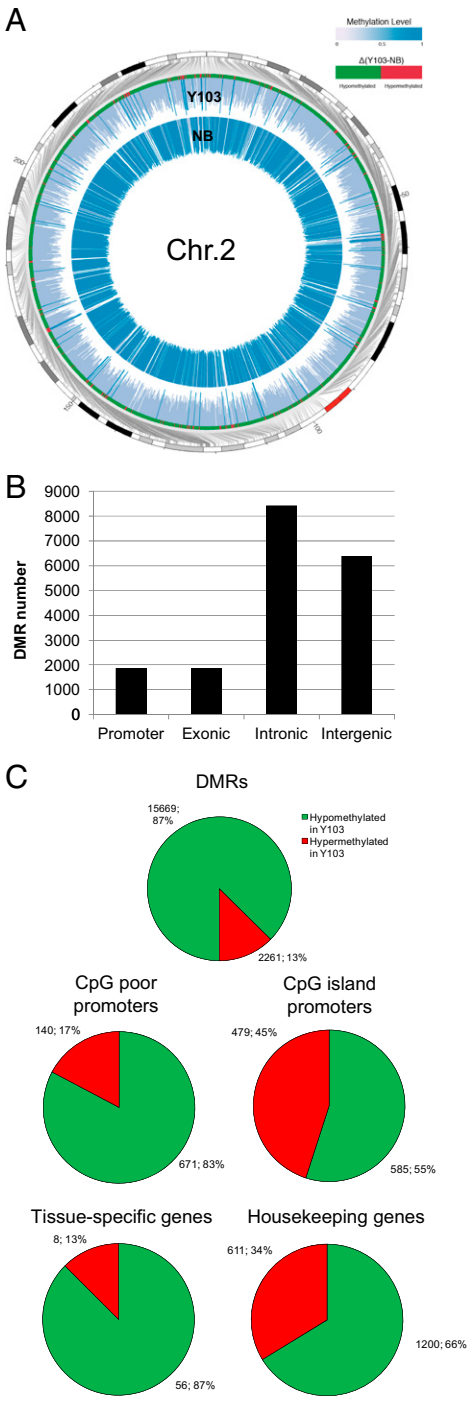

Fig. 2. DMRs of NB and Y103. (*A*) Circos representation of average methylation levels for all of the CGs in the DMRs in chromosome 2, indicating whether the region was methylated (red) or unmethylated (green) in the Y103 sample relative to the NB sample. Regions are equally spaced around the figure, but their original locations in the genome are indicated by gray lines. (*B*) DMR distribution among different genomic sequences. (*C*) Distribution of DMRs according to the direction of the DNA methylation change, the type of promoter (with or without a CpG island), and the reported expression pattern.

CNVs were observed in NB and Y103, which are values within the reported range for the general population (24).

As described above for the WGBS data, the most common DMR change was the presence of an unmethylated sequence in the centenarian that was methylated in the newborn: in Y103, 87% (15,669) of the DMRs corresponded to more unmethylated sequences, and the remaining 13% (2,261) of them represented more methylated sequences (Fig. 2*C*). Using the EpiGRAPH software (25), the hypomethylated DMRs in the centenarian compared with the neonate were found to be enriched not only for the previously mentioned DNA repetitive elements (Wilcoxon test, $P < 8.9 \times 10^{-16}$), but also for the repressive promoter-associated histone marks (K27H3me2, K9H3me2, and K9H3me3; $P < 2.2 \times 10^{-16}$), previously determined by chromatin immunoprecipitation-sequencing of $CD4^+$ T cells (26). The DMRs corresponding to more methylated CpGs in the centenarian were enriched for CpG islands (Wilcoxon test, $P < 2.2 \times 10^{-16}$). In this context, the differences in the DNA methylation patterns were particularly marked in the case of DMRs located in promoter sequences (1,875), with respect to their CpG content: CpG-poor promoters had more unmethylated sequences in 83% (671 of 811) of associated DMRs in the centenarian DNA, and reduced methylation was observed in 55% (585 of 1,064) of the DMRs located in promoter CpG islands (χ^2 test, $P < 2.2 \times 10^{-16}$) (Fig. 2*C*). We made the same observation even at the whole-genome level, using the National Center for Biotechnology Information reference sequences: CpG-poor promoters were enriched in more hypomethylated DMRs (671 of 16,631) in the centenarian vs. the newborn compared with the annotated CpG island promoters (585 of 23,139) (χ^2 test, $P < 3.1 \times 10^{-16}$). We also analyzed the DMR data with respect to the expression profile of the associated genes obtained from the tissue-specific gene expression and regulation (TiGER) database (27). Considering the expression profile, genes with restricted tissue-specific expression more commonly had DMR hypomethylated sequences in the Y103 sample vs. the NB sample (87%, 56 of 64) than did housekeeping genes (66%, 1,200 of 1,811) (χ^2 test, $P < 6.4 \times 10^{-4}$) (Fig. 2*C*). Quantitative real-time RT-PCR analyses for eight randomly selected genes with tissue-specific expression and a hypomethylated DMR in their gene promoter showed an increase in gene expression in the Y103 sample compared with the newborn (Fig. S4*C*). These results suggest that the centenarian $CD4^+$ T-cell sample had a lower DNA methylation content at CpG-poor promoters, which might be associated with an inappropriate expression of tissue-specific genes, but a subset of ubiquitously expressed genes had an enriched DNA methylation status at their CpG island promoters.

DNA Methylation Map of 450,000 CpG Sites in Newborns and Nonagenarians. We sought to extrapolate our WGBS data and their corresponding derived DMR analyses to a larger collection of newborn and centenarian samples. We used a newly developed DNA methylation microarray that assays the DNA methylation status of 450,000 CpG sites (28, 29). The platform is an extension of other DNA methylation microarrays, such as those for 1,500 CpG (30–33) and 27,000 CpG (34, 35) sites, which have been used to characterize age-related DNA methylation changes (32–36). These microarray platforms produce accurate DNA methylation data at a similar level to other approaches, such as methylated DNA immunoprecipitation sequencing, methylated DNA capture by affinity purification, or reduced-representation bisulfite sequencing (37).

The methylation levels obtained from all 450 K CpG sites included in the microarray significantly correlated with those values obtained from the WGBS technology (Pearson correlation, $r^2 = 0.94$, $P < 0.01$) (Fig. S5*A*). Among the 17,930 WGBS-derived DMRs, 2,337 DMRs (corresponding to 3,205 CpG sites) were also represented by at least one CpG site using the 450 K platform. These 450 K-DMRs are illustrated for NB and for Y103 in Fig. S5*B* using Circos (17) and are listed in detail in Dataset S2. For the WGBS-derived DMRs from the NB and the Y103 samples, we also observed a strong correlation with their methylation levels detected at the corresponding CpG sites using

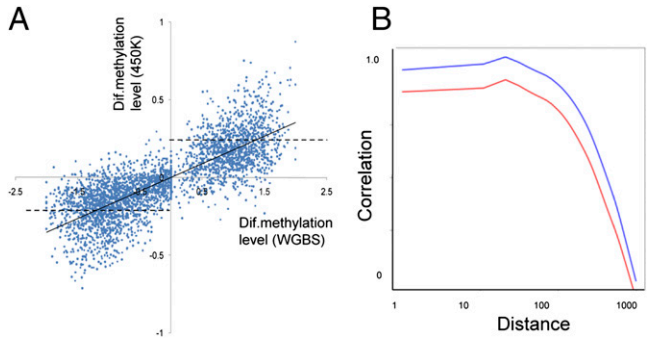

Fig. 3. WGBS data validation with a 450-K CpG site DNA methylation microarray. (*A*) Comparison of CpG methylation differences between the NB and Y103 obtained from WGBS (*x*-axis) and 450-K array (*y*-axis) technology. Displayed are all differentially methylated CpG sites derived from the WGBS approach that were also present on the 450-K array (3,205). The threshold of 0.20 change in β-values is indicated (broken line). (*B*) The curves of the distance correlation between the methylation status of the neighboring CpG sites of NB (blue line) and Y103 (red line) samples determined by the 450-K microarray. A more pronounced declining curve indicates lower correlation in terms of the methylation status of nearby CpGs.

the 450 K DNA methylation microarray (Kendall rank correlation, $\tau = 0.61$, $P < 2.2 \times 10^{-16}$) (Fig. 3*A*). In detail, from the differentially methylated CpG sites present on the 450 K array, 87.3% (2,797 of 3,205) revealed concordant directional changes between the platforms. For subsequent analysis only double-validated CpG sites that passed a threshold of a 0.20 change in β-values were used (1,149 CpG sites), representing 878 WGBS-DMRs. This was not an independent validation set of the WGBS-derived methylation signature, but instead was a technical validation of the results obtained from both platforms. All of these 1,149 CpG sites were concordant between both platforms and were associated with 615 genes. The 450 K DNA methylation microarray also confirmed the lower correlation in the methylation status of nearby CpGs within the Y103 sample in comparison with NB that we observed with the WGBS methodology (Fig. 3*B*). We then analyzed in more detail five randomly selected candidate promoters that were associated with double-validated (WGBS+450 K) DMRs. Targeted bisulfite genomic sequencing of multiple clones and pyrosequencing were used to determine the DNA methylation status; quantitative real-time RT-PCR was used for the expression analyses. Having previously observed that the most frequent DNA methylation change in the centenarian was the aforementioned lower level of methylation at CpG-poor promoters, we observed the same trend in the *AIM2* (absent in melanoma-2) and *TNFRSF9* coding genes and in the miR-21 locus (Fig. S5*C*). Importantly, in the centenarian DNA these unmethylated sequences were associated with higher levels of expression of each gene than in the newborn sample (Fig. S5*D*). For other, less frequently observed epigenetic drifts in the centenarian DNA, such as an increase (e.g., IGSF9B) or a decrease (e.g., PTPRE) in DNA methylation of CpG islands promoters, the single-locus DNA methylation analyses also confirmed the WGBS and 450 K results (Fig. 3*C*) and corroborated the association with the corresponding transcript levels (Fig. S5*D*).

Once we had characterized the DNA methylomes of NB and Y103 by WGBS, identified their DMRs and validated them in the 450 K platform, we extended the study to a larger set of individuals at either extreme of age: very young and very old. Thus, we generated 450 K DNA methylation microarray data using DNA extracted from 19 newborns (cord blood, Caucasians) and 19 nonagenarians [peripheral blood mononuclear cells (PBMCs), Caucasians, mean age: 92.6-y-old, range: 89–100 y old]. Unsupervised clustering analysis of 4,301,332 SNPs failed to distinguish the neonate and nonagenarian groups (Fig. S5*E*). An average of 88 and 92 CNVs were observed for the newborn and nonagenarian groups, respectively; these values are within the reported range for the general population (24). Aneuploidy was not observed in any case. To avoid any incorrect DNA methylation call in the 450 K microarray because of SNPs that were within the probe site, we excluded all probes containing SNPs with annotation in the University of California at Santa Cruz browser (190,490 CpG sites) (Dataset S3) from further analysis. We also excluded all CpG sites present in the X and Y chromosomes (11,135 and 460 CpG sites, respectively) to avoid sex-related DNA methylation differences.

The DNA methylation microarray analyses of each group confirmed the lower correlation in the methylation status of nearby CpGs in the nonagenarian group in comparison with the newborn group (Wilcoxon test, $P < 2.2 \times 10^{-16}$) (Fig. 4*A*), as we had previously found in the Y103 centenarian sample (Figs. 1*D* and 3*B*). The different shape of the curve is attributable to the absence of a large number of CpGs located in repeats and intergenic regions that are detectable by WGBS, but which are not printed in the 450 K DNA methylation array. Using unsupervised hierarchical clustering of the 283,579 CpGs present in the 450 K methylation array (after exclusion of the SNP-associated and X and Y chromosome CpG sites), the newborn and nonagenarian groups were found to cluster separately (Fig. 4*B*). Using the 1,149 technically double-validated CpG sites located in WGBS-derived DMRs, we also encountered significant differential DNA methylation patterns between the newborn and nonagenarian groups (multiscale bootstrap resampling, $n = 1{,}000$, $P < 0.0001$), which enabled them to be distinguished by the hierarchical clustering approach (Fig. 4 *C* and *D*). In particular, from those technically double-validated CpG sites derived from WGBS (that were not obtained from any independent validation set), we obtained 214 CpG dinucleotides (165 DMRs) that distinguished the two groups (Fig. 4 *C* and *E*). These "differentially methylated CpG sites" (dmCpGs) were identified using a threshold of 0.20 change in average β-values and a false-discovery rate (FDR) below 0.01 in the ANOVA test adjusted for multiple hypothesis testing when comparing both sets of cases. We further validated the differential DNA methylation status of the identified CpG sites by bisulfite genomic sequencing of multiple clones and pyrosequencing of the described WGBS-derived candidate genes (*AIM2*, *TNFRSF9*, *IGSF9B*, and *PTPRE*) in five newborn and five nonagenarian samples (Fig. S6). We also confirmed the differential DNA methylation status of two additional dmCpG-associated genes: *ZRP*-1 (Zyxin-related protein-1) and *FHL2* (Four-and-a-half LIM Domains 2) (Fig. S6).

When we included all of the studied 283,579 CpG sites of the microarray, we found 5,774 additional CpG sites (450K-dmCpGs) that were differentially methylated between the newborn and nonagenarian groups (0.20 change in average β-values, ANOVA test FDR < 0.01), which also enabled them to be distinguished by the hierarchical clustering approach (Fig. 4*F*). The 450K-dmCpGs were identified in promoter (24%), exonic (17%), intronic (56%), and intergenic regions (3%) (Fig. 4*G*). As we also previously observed for NB and Y103, quantitative real-time RT-PCR analyses for Alu and MIR transcripts showed an increase in the expression of these hypomethylated repetitive sequences in nonagenarian samples compared with a newborn (Fig. S7). We confirmed the differential DNA methylation status of four 450K-dmCpG associated genes: *PLCH1* (Phospholipase C eta 1), *EXTL1* (Exostoses multiple-like 1), *SLAMF7* (SLAM family member 7), and *IL1R2* (Interleukin 1 Receptor type II) (Fig. S8).

The 450K-dmCpGs identified in newborn and nonagenarian groups mimicked the CpG content-DNA methylation changes observed in the WGBS-DMRs. The most common 450K-dmCpG change was the presence of an unmethylated CpG in the

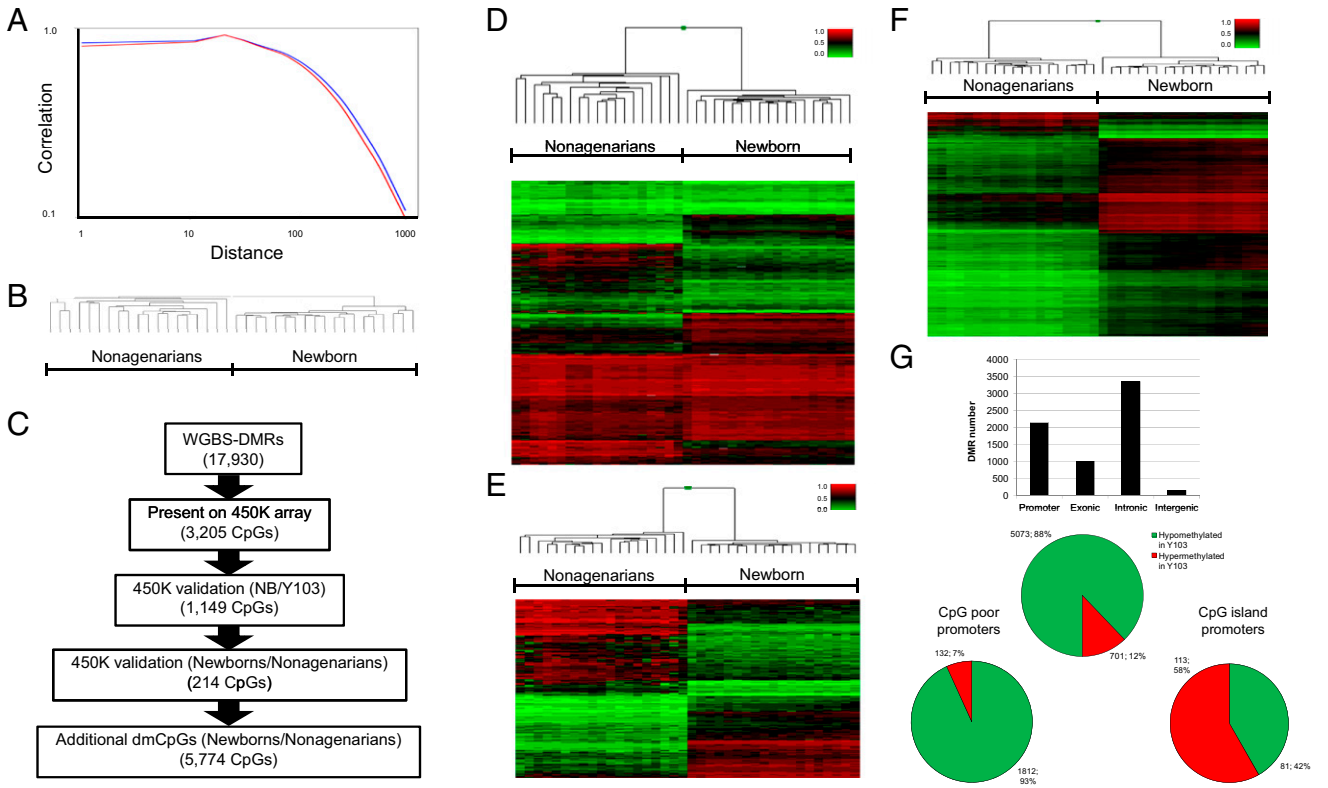

Fig. 4. Extension of the WGBS data to a comprehensive set of newborns and nonagenarians and the observed overlap in premature aging disorders. (*A*) The curves of the distance correlation between the methylation status of the neighboring CpG sites of newborns (blue line) and nonagenarians (red line) samples determined by the 450 K microarray. A more pronounced declining curve indicates lower correlation in terms of the methylation status of nearby CpGs. (*B*) Unsupervised hierarchical clustering of the 283,579 CpGs present in the 450 K methylation array (after exclusion of SNP-associated and X and Y chromosome CpG sites) in the newborn and nonagenarian groups. (*C*) Schematic overview of the assessment of differentially methylated CpG sites between the initial newborn and 103-y-old sample and the cohort of 19 newborns and 19 nonagenarians. (*D*) Hierarchical clustering approach using the 1,149 WGBS-derived dmCpGs in the newborn and nonagenarian groups. (*E*) Supervised clustering analysis with the 165 DMRs (214 dmCpG dinucleotides) that distinguished the two groups. (*F*) Hierarchical clustering approach using the additional 5,774 450K-dmCpG sites in the newborn and nonagenarian groups. (*G*) (*Upper*) The 450K-dmCpG distribution among different genomic sequences; (*Lower*) Distribution of the 450K-dmCpGs according to the direction of the DNA methylation change and the type of promoter (with or without a CpG island).

nonagenarian group that was methylated in the newborn set: in the aged group 88% (5,073 of 5,774) of the 450K-dmCpG corresponded to more unmethylated CpGs (Fig. 4*G*). Considering only the 450K-dmCpG in promoters (2,138 CpG sites), the lower level of DNA methylation was especially evident in CpG-poor promoters (χ^2 test, $P < 2.2 \times 10^{-16}$). In the nonagenarian DNAs, an unmethylated CpG dinucleotide in these loci was observed in 93% (1,812 of 1,944) of cases, whereas a methylated CpG was seen in 58% (113 of 194) of CpG island promoters (Fig. 4*G*).

We performed an ANOVA for each of the 5,988 age-related CpG site (214 dmCpGs from WGBS + 5,774 from the 450 K array) in the newborn and nonagenarian groups to evaluate the ratio of intergroup and intragroup variability. CpG sites showed little variation within the groups but differed between the newborn and nonagenarian groups (ANOVA, FDR-corrected $P < 8.8 \times 10^{-4}$). Furthermore, we randomly split our two original sample groups into training and test subsets. The cluster was derived from each training subset using a hierarchical cluster with a complete link type, calculating Euclidean distances between samples. This cluster model was then successfully applied to each test subset.

CpG-methylated sequences in nonagenarians, in comparison with the newborn group, were significantly more likely to occur in gene promoters with enriched Polycomb occupancy (χ^2 test, $P = 2.2 \times 10^{-16}$) (38), and a bivalent histone domain (K4H3me3 + K27H3me3) present in hematopoietic stem cells (χ^2 test, $P = 0.0029$) (39), as has recently been suggested (33–36). We were also encouraged by the fact that one-fifth of the age-related DNA methylation changes previously identified using the 1,505 CpG (32, 33) or 27 K CpG (34–36) DNA methylation platforms were also observed in our promoter dmCpGs. Among these, we may note *GLRA1* (Glycine receptor α-1 subunit) and *ASCL2* (Achaete-Scute homolog 2) for the hypermethylated genes in the nonagenarians vs. the neonates. Representative examples of the hypomethylated genes in the nonagenarians vs. the newborns include HLA-DPA1 (Major Histocompatibility Complex Class II DP α-1), *PXN* (Paxillin), and Complexin 2 (*CPLX2*). With respect to suspected aging targets and pathways derived from gene-association and functional studies (1, 2), we identified age-associated DMRs in the promoter regions of Sirtuin 5 (SIRT5), Sirtuin 7 (SIRT7), and insulin-like growth factor binding protein 4 (IGFBP4). The roles of the sirtuins and insulin/insulin-like growth factor signaling in the modulation of lifespan are well established, and their genes and proteins are part of interconnected pathways (1, 2). Gene-set enrichment analysis using ranked methylation differences between the newborn and the nonagenarian group in the 5,988 age-related CpG sites revealed significant enrichment only in one dataset: genes with differential expression in aging hematopoietic progenitor cells (FDR, q = 0.036) (40). The list of dmCpGs in promoters is provided in Dataset S4.

Although our study focused on the most extreme ages of the human lifespan, we also wondered about the natural history of the observed DNA methylation changes, particularly if a progressive gradient of epigenetic drift occurred during the aging process (30–34). The 450 K DNA methylation microarray analysis of an

additional cohort of 19 PBMCs from healthy donors with an average age of 59.8 y showed that, in this middle-aged group, 70% (4,179 of 5,988) of the CpGs found to differ between the newborn and the nonagenarian had intermediate levels of CpG methylation (ANOVA, $P < 0.05$) (Fig. S9), but only 2% (133) and 28% (1,692) were closer to the newborn or nonagenarian subsets, respectively. Using a lower CpG coverage platform (27,578 CpG site microarray), 1,030 CpG sites were previously identified as undergoing a major DNA methylation change in the first year of life, possibly related to maturation of the immune system (41). Of those 1,030 CpG sites, 987 (96%) are also included in the 450 K microarray used herein. Most importantly, only 5.9% (58 of 987) of those CpG sites were present in our observed newborn/nonagenarian-discriminating CpG sites. An evaluation of the frequency of biological process term annotations from the Gene Ontology (GO) database (*SI Materials and Methods*) at the GO level 5 demonstrated that our described age-related hyper- and hypomethylated CpG sites in the gene promoters occurred throughout a wide spectrum of biological processes (Dataset S5). Thus, our results support a model of small cumulative DNA methylation changes during a lifetime.

In summary, we are unique in reporting the complete DNA methylomes of $CD4^+$ T cells from newborn and centenarian individuals. The unmasked DNA methylation landscape shows that DNA obtained from a 103-y-old donor was more unmethylated overall than DNA from the same cell type obtained from a neonate. Furthermore, the centenarian samples showed a lower correlation in terms of the methylation status of neighboring CpGs in comparison with the newborn sample, which was more homogenously methylated in nearby located CpGs. The hypomethylated CpGs observed in the centenarian DNAs compared with the neonates covered all genomic compartments, such as promoters, exonic, intronic, and intergenic regions. From a gene-regulatory point of view, one of the main epigenetic features of the centenarian sample is its low DNA methylation density at CpG-poor promoters and tissue-specific genes. These findings have been validated in a larger cohort of newborn and nonagenarian individuals using a 450 K CpG DNA methylation microarray. Our results demonstrate that the DNA methylomes at the two extremes of the human lifespan are distinct.

Materials and Methods

Peripheral blood was obtained from healthy donors or from umbilical cord blood from newborns. Written informed consent was obtained from the individuals or their parents, respectively. To separate $CD4^+$ cells $CD4^+$ T-cell Isolation kit II (Miltenyi Biotec) was applied following the manufacturer's instructions. RNA and DNA were extracted using TRIzol (Invitrogen) and Phenol:Chloroform:Isoamylalcohol (Sigma), respectively. Bisulfite modification of genomic DNA was carried out with the EZ DNA Methylation Kit (Zymo) following the manufacturer's protocol. DNA methylation status was established by WGBS (13), 450 K CpG site microarray analysis (28), and bisulfite genomic sequencing (33) and pyrosequencing (33) of candidate genes. Expression of mRNAs was determined by real-time RT-PCR. Additional experimental details are provided in *SI Materials and Methods*. Bisulfite sequencing primers are listed in Table S1.

ACKNOWLEDGMENTS. We thank Dr. Assumpta Ferrer for her support. This study was supported by European Research Council Advanced Grant Epigenetic Disruption of Non-Coding RNAs in Human Cancer (EPINORC), SAF2009-07319, CSD2006-49, SAF2011-22803, Portuguese Foundation for Science and Technology, and Fundacio Cellex.

1. Wheeler HE, Kim SK (2011) Genetics and genomics of human ageing. *Philos Trans R Soc Lond B Biol Sci* 366:43–50.
2. Fadini GP, Ceolotto G, Pagnin E, de Kreutzenberg S, Avogaro A (2011) At the crossroads of longevity and metabolism: The metabolic syndrome and lifespan determinant pathways. *Aging Cell* 10:10–17.
3. De Carvalho DD, You JS, Jones PA (2010) DNA methylation and cellular reprogramming. *Trends Cell Biol* 20:609–617.
4. Berdasco M, Esteller M (2010) Aberrant epigenetic landscape in cancer: How cellular identity goes awry. *Dev Cell* 19:698–711.
5. Fraga MF, et al. (2005) Epigenetic differences arise during the lifetime of monozygotic twins. *Proc Natl Acad Sci USA* 102:10604–10609.
6. Kaminsky ZA, et al. (2009) DNA methylation profiles in monozygotic and dizygotic twins. *Nat Genet* 41:240–245.
7. Feinberg AP (2007) Phenotypic plasticity and the epigenetics of human disease. *Nature* 447:433–440.
8. Jones PA, Baylin SB (2007) The epigenomics of cancer. *Cell* 128:683–692.
9. Urdinguio RG, Sanchez-Mut JV, Esteller M (2009) Epigenetic mechanisms in neurological diseases: genes, syndromes, and therapies. *Lancet Neurol* 8:1056–1072.
10. Rodríguez-Paredes M, Esteller M (2011) Cancer epigenetics reaches mainstream oncology. *Nat Med* 17:330–339.
11. Rakyan VK, Down TA, Balding DJ, Beck S (2011) Epigenome-wide association studies for common human diseases. *Nat Rev Genet* 12:529–541.
12. Lister R, et al. (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* 462:315–322.
13. Li Y, et al. (2010) The DNA methylome of human peripheral blood mononuclear cells. *PLoS Biol* 8:e1000533.
14. Lister R, et al. (2011) Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. *Nature* 471:68–73.
15. Hansen KD, et al. (2011) Increased methylation variation in epigenetic domains across cancer types. *Nat Genet* 43:768–775.
16. Berman BP, et al. (2011) Regions of focal DNA hypermethylation and long-range hypomethylation in colorectal cancer coincide with nuclear lamina-associated domains. *Nat Genet* 10:969.
17. Krzywinski M, et al. (2009) Circos: An information aesthetic for comparative genomics. *Genome Res* 19:1639–1645.
18. Irizarry RA, et al. (2009) The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nat Genet* 41:178–186.
19. Doi A, et al. (2009) Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. *Nat Genet* 41:1350–1353.
20. Bollati V, et al. (2009) Decline in genomic DNA methylation through aging in a cohort of elderly subjects. *Mech Ageing Dev* 130:234–239.
21. Jintaridth P, Mutirangura A (2010) Distinctive patterns of age-dependent hypomethylation in interspersed repetitive sequences. *Physiol Genomics* 41:194–200.
22. Witt N, et al. (2009) An assessment of air as a source of DNA contamination encountered when performing PCR. *J Biomol Tech* 20:236–240.
23. Guelen L, et al. (2008) Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* 453:948–951.
24. Redon R, et al. (2006) Global variation in copy number in the human genome. *Nature* 444:444–454.
25. Bock C, Walter J, Paulsen M, Lengauer T (2007) CpG island mapping by epigenome prediction. *PLOS Comput Biol* 3:e110.
26. Barski A, et al. (2007) High-resolution profiling of histone methylations in the human genome. *Cell* 129:823–837.
27. Liu X, Yu X, Zack DJ, Zhu H, Qian J (2008) TiGER: A database for tissue-specific gene expression and regulation. *BMC Bioinformatics* 9:271.
28. Sandoval J, et al. (2011) Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. *Epigenetics* 6:692–702.
29. Bibikova M, et al. (2011) High density DNA methylation array with single CpG site resolution. *Genomics* 98:288–295.
30. Bibikova M, et al. (2006) High-throughput DNA methylation profiling using universal bead arrays. *Genome Res* 16:383–393.
31. Byun HM, et al. (2009) Epigenetic profiling of somatic tissues from human autopsy specimens identifies tissue- and individual-specific DNA methylation patterns. *Hum Mol Genet* 18:4808–4817.
32. Christensen BC, et al. (2009) Aging and environmental exposures alter tissue-specific DNA methylation dependent upon CpG island context. *PLoS Genet* 5:e1000602.
33. Fernandez AF, et al. (2012) A DNA methylation fingerprint of 1,628 human samples. *Genome Res* 22:407–419.
34. Rakyan VK, et al. (2010) Human aging-associated DNA hypermethylation occurs preferentially at bivalent chromatin domains. *Genome Res* 20:434–439.
35. Teschendorff AE, et al. (2010) Age-dependent DNA methylation of genes that are suppressed in stem cells is a hallmark of cancer. *Genome Res* 20:440–446.
36. Bocker MT, et al. (2011) Genome-wide promoter DNA methylation dynamics of human hematopoietic progenitor cells during differentiation and aging. *Blood* 117:e182–e189.
37. Bock C, et al. (2010) Quantitative comparison of genome-wide DNA methylation mapping technologies. *Nat Biotechnol* 28:1106–1114.
38. Lee TI, et al. (2006) Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell* 125:301–313.
39. Cui K, et al. (2009) Chromatin signatures in multipotent human hematopoietic stem cells indicate the fate of bivalent genes during differentiation. *Cell Stem Cell* 4:80–93.
40. Wagner W, et al. (2009) Aging and replicative senescence have related effects on human stem and progenitor cells. *PLoS ONE* 4:e5846.
41. Martino DJ, et al. (2011) Evidence for age-related and individual-specific changes in DNA methylation profile of mononuclear cells during early immune development in humans. *Epigenetics* 6:1085–1094.