clsa élcv
Canadian Longitudinal Study on Aging
Étude longitudinale canadienne sur le vieillissement

The Canadian Longitudinal Study on Aging: Genome-Wide DNA Methylation Profiling on 1,479 Participants using Illumina Infinium MethylationEPIC BeadChip Microarray Technology Data Support Document

David Tse Shen Lin, Julia L. MacIsaac, Katia E. Ramadori, Chris P. Verschoor, Lisa M. McEwen, and Michael S. Kobor

DNA methylation represents one of the most well-studied epigenetic marks, which has previously been shown to influence gene expression without altering the genomic sequence within an individual[1,2].

Investigating this mechanism within a population can provide insight into how the environment, including the process of aging, can influence cellular function and potentially an individuals' risk of adverse health outcomes. In this data release, we profiled genome-wide DNA methylation in peripheral blood mononuclear cells (PBMCs) isolated from 1,479 selected participants enrolled in the Canadian Longitudinal Study on Aging (CLSA) using the Illumina Infinium MethylationEPIC BeadChip microarrays (hereinafter referred to as *EPIC arrays*).

The EPIC arrays provide quantitative measurements of DNA methylation at 862,927 CpG sites and 2,932 CHH sites[3]. To generate these methylation raw data, genomic DNA was extracted from the PBMC samples and then bisulfite-converted to be used as input for the EPIC arrays. Once processed, the microarray chips were read using an Illumina *iScan* to yield raw methylation data in *.idat* file format. To facilitate ease of use for future analysts, these data were further preprocessed using GenomeStudio software (Illumina) which transformed the raw intensity values into beta values (*M/(M+U)*), representing a continuous value ranging from 0 to 1 which can be interpreted as the percentage of methylation at each CpG loci present in the sample[4]. Within GenomeStudio, colour-correction and background subtraction was performed by employing built-in control probes measured within each sample. Data was then processed in the *R* environment for further preprocessing and normalization procedures: Following initial removal of four samples with low bisulfite-conversion scores (<85%), inconsistently-performing and non-specific probes[5]were removed as part of the initial quality control process. We then employed rigorous procedures based on a combination of principal components analysis (PCA), sample beta-value correlations, and built-in outlier detection functions in the *wateRmelon* and *lumi* packages to identify 29 additional outlier samples which were excluded from downstream preprocessing procedures. Of the remaining 1,446 samples and 783,136 probes, missing beta values were imputed using a k-nearest neighbor-based (*imputeknn*) algorithm. The imputed data were subsequently normalized using inter-sample (Quantile) and intra-sample (Beta-Mixture quantile; "BMIQ") methods[6,7]. Variations due to technical batch effects were corrected using the *ComBat* function based on three array-related parameters (array row, sample bisulfite

conversion batch, and array BeadChip)[8]. Finally, we computationally estimated proportions of six major immune cell types in each PBMC sample using the Houseman method, and adjusted the beta-values to remove this variation across samples using a PCA-based regression method[9,10].

Importantly, 16 technical replicate PBMC samples from the same individual were included in these arrays, each incorporated into a separate array batch. These replicates were processed in different DNA extraction and bisulfite conversion batches to control for technical variation from these procedures. The final Pearson's correlation values for global DNA methylation between these technical replicates were determined to be $r > 0.99$ ($p < 2.2 \times 10^{-16}$), confirming the validity of our methods[3]. To further validate our data, we performed a preliminary epigenome-wide association study (EWAS) with the participants' chronological age as the main effect, sex as a covariate, and the preprocessed DNA methylation values at each CpG (Logit-transformed to M-values)[11] as the outcome in a linear regression model. We were able to replicate previous findings with known aging-related CpGs annotated to genes including ELOVL2[12] (*cg16867657*: Δβ: 0.13, FDR:1.48×10^{-219}; *cg24724428*: Δβ: 0.10, FDR:3.17×10^{-162}; *cg21572722*: Δβ: 0.08, FDR:1.41×10^{-142}), FHL2[12] (*cg17268658*: Δβ: 0.09, FDR:9.36×10^{-145}; *cg22454769*: Δβ: 0.09, FDR:1.45×10^{-135}; *cg06639320*: Δβ: 0.11, FDR:5.36×10^{-134}), CILP2[13] (*cg07544187*: Δβ: 0.10, FDR:4.54×10^{-123}), and CCDC102B[14] (*cg13552692*: Δβ: -0.11, FDR:4.70×10^{-118}; *cg19283806*: Δβ: -0.09, FDR:1.33×10^{-103}).

In addition, we also calculated epigenetic aging measures with DNA methylation data which was only colour-corrected and background-subtracted, as recommended, using the Horvath online DNA Methylation Age calculator software (https://dnamage.genetics.ucla.edu/home)[15,16]. These age-related variables include DNA methylation age estimates from the Horvath pan-tissue 353-CpG epigenetic clock[15], the Hannum blood-based 71-CpG epigenetic clock[17], as well as universal, intrinsic, and extrinsic age acceleration residuals[15,18,19] for each participant.

The following data are available for access by researchers that have followed the policies and guidelines set forth by the CLSA:

1. The raw DNA methylation data containing beta values from all 865,918 probes for all 1,479 participants on the EPIC arrays (provided as .idat files or as a beta-value matrix derived from the MethylumiSet object in the CSV file format)
2. The color-corrected/background-subtracted, probe-filtered, sample outlier-removed, normalized, batch and blood cell-type corrected beta values at 783,136 probes for 1,446 participants as a beta-value matrix in the CSV format
3. Epigenetic-age related measures for all 1,479 participants from the Horvath DNAmAge clock calculator, using the genome-studio colour-corrected and background-subtracted methylation data as input.

Appendix: Epigenetic Age-related Measures provided in this data release (also see references: 15-19 for more details).

1. **DNAmAge –** This is the absolute DNA methylation/epigenetic age estimates calculated based on the Horvath 353-CpG Pan-Tissue clock sites in the units of biological years. This value shows high correlation with a given individual's chronological age and is in theory unconfounded with cell type proportions[15].
2. **Age Acceleration Difference –** Absolute difference between chronological age and DNAmAge for an individual - ie. is calculated as (DNAmAge - Chronological Age)[15].
3. **Age Acceleration Residual –** Represents an epigenetic age acceleration measure defined as residual from regressing DNAmAge on chronological age - this is typically used as the universal measure of epigenetic age acceleration[15].
4. **Intrinsic Epigenetic Age Acceleration (IEAA) –** Represents an epigenetic age acceleration estimate measure that is attributed to *intrinsic changes within the cells* regardless of cell type proportions in a given sample[18]. IEAA is measured by accounting for both an individual's chronological age and blood cell type proportions.
5. **Extrinsic Epigenetic Age Acceleration (EEAA) –** Represents an epigenetic age acceleration estimate measure that is attributed to *age-related changes in blood cell type composition*[18]. EEAA is more related to immune system aging.
6. **Hannum Age –** Epigenetic Age calculated based on 71 CpG sites as defined by Hannum et al.[17] This clock was developed using whole blood samples.

References

1. Jones PA and Takai D. (2001) The role of DNA methylation in mammalian epigenetics. *Science*. **293**:1068-1070.
2. Zhu H, Wang G, and Qian J. (2016) Transcription factors as readers and effectors of DNA methylation. *Nat Rev Genet.* **17**:551-565.
3. Infinium MethylationEPIC BeadChip Data Sheet. *https://sciencedocs.illumina.com/documents/Microarray/infinium-methylation-epic-data-sheet- 1070-2015-008/Content/Source/Microarray/Infinium/MethylationEPIC/infinium-methylation-epic- data-sheet.html*
4. GenomeStudio Methylation Modul v1.8 User Guide. *https://support.illumina.com/content/dam/illumina-support/documents/documentation/software_documentation/genomestudio/genomestudio-2011-1/genomestudio-methylation-v1-8-user-guide-11319130-b.pdf*
5. Pidsley R, Zotenko E, Peters TJ, Lawrence MG, Risbridger GP, Molloy P, Van Diik S, Muhlhausler B, Stirzaker C, and Clark SJ. (2016) Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. *Genome Biol*. **17**:208.
6. Hicks SC and Irizarry RA. (2015) quantro: a data-driven approach to guide the choice of an appropriate normalization method. *Genome Biol.* **16**:117.
7. Teschendorff AE, Marabita F, Lechner M, Bartlett T, Tegner J, Gomez-Cabrero D, and Beck S. (2013) A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. *Bioinformatics*. **29**:189-196.
8. Johnson WE, Rabinovic A, and Li C. (2007) Adjusting batch effects in microarray expression data using Empirical Bayes methods. *Biostatistics*. **8**:118-127.
9. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, Wiencke JK, and Kelsey KT. (2012) DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics*. **13**:86.
10. Jones MJ, Islam SA, Edgar RD, and Kobor MS. (2017) Adjusting for Cell Type Composition in DNA Methylation Data Using a Regression-Based Approach. *Methods Mol Biol*. **1589**:99-106.
11. Du P, Zhang X, Huang CC, Jafari N, Kibbe WA, Hou L, and Lin SM. (2010) Comparison of Beta- value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics*. **11**:587.
12. Garagnani P, Bacalini MG, Pirazzini C, Gori D, Giuliani C, Mari D, Di Blasio AM, Gentilni D, Vitale G, Collino S, Rezzi S, Castellani G, Capri M, Salvioli S, and Franceschi C. (2012) Methylation of ELOVL2 gene as a new epigenetic marker of age. *Aging Cell*. **11**:1132-1134.
13. Marttila S, Kananen L, Häyrynen S, Jylhävä J, Nevalainen T, Hervonen A, Jylhä M, Nykter M, and Hurme M. (2015) Ageing-associated changes in human DNA methylome: genomic locations and effects on gene expression. *BMC Genomics*. **16**:179.
14. Florath I, Butterbach K, Müller H, Bewerunge-Hudler M, and Brenner H. (2013) Cross-sectional and longitudinal changes in DNA methylation with age: an epigenome-wide analysis revealing over 60 novel age-associated CpG sites. *Hum Mol Genet.* **23**:1186-1201.

15. Horvath S. (2013) DNA methylation age of human tissues and cell types. *Genome Biol*. **14**:R115.

16. McEwen LM, Jones MJ, Lin DTS, Edgar RD, Husquin LT, MacIsaac JL, Ramadori KE, Morin AM, Rider CF, Carlsten C, Quintana-Murci L, Horvath S, and Kobor MS. (2018) Systematic evaluation of DNA methylation age estimation with common preprocessing methods and the Infinium MethylationEPIC BeadChip array. *Clin Epigenetics*. **10**:123.

17. Hannum G, Guinney J, Zhao L, Zhang L, Hughes G, et al. (2013) Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol Cell*. **49**:359-367.

18. Chen BH, Marioni RE, Colicino E, Peters MJ, Ward-Caviness CK, Tsai P, Roetker NS, Just AC, Demerath EW, Guan W, Bressler J, Fornage M, Studenski S, Vandiver AR, Moore AZ, Tanaka T, Kiel DP, Liang L, Vokonas P, Schwartz J, Lunetta KL, Murabito JM, Bandinelli S, Hernandez DG, Melzer D, Nalls M, Pilling LC, Price TR, Singleton AB, Gieger C, Holle R, Kretschmer A, Kronenberg F, Kunze S, Linseisen J, Meisinger C, Rathmann W, Waldenberger M, Visscher PM, Shah S, Wray NR, McRae AF, Franco OH, Hofman A, Uitterlinden AG, Absher D, Assimes T, Levine ME, Lu AT, Tsao PS, Hou L, Manson JE, Carty CL, LaCroix AZ, Reiner AP, Spector TD, Feinberg AP, Levy D, Baccarelli A, van Meurs J, Bell JT, Peters A, Deary IJ, Pankow JS, Ferrucci L, and Horvath S. (2016) DNA methylation-based measures of biological age: meta-analysis predicting time to death. *Aging*. **8**:1844-1859.

19. Quach A, Levine ME, Tanaka T, Lu AT, Chen BH, Ferrucci L, Ritz B, Bandinelli S, Neuhouser ML, Beasley JM, Snetselaar L, Wallace RB, Tsao PS, Absher D, Assimes TL, Stewart JD, Li Y, Hou L, Baccarelli AA, Whitsel EA, and Horvath S. (2017) Epigenetic clock analysis of diet, exercise, education, and lifestyle factors. *Aging*. **9**:419-437.