

Cross-Platform Evaluation of Long-Read Sequencing for DNA Methylation and Portability of Methylation Predictors

Zongli Xu

National Institute of Environmental Health Sciences

John S. House

National Institute of Environmental Health Sciences

Joseph Breeyear

National Institute of Environmental Health Sciences

David C. Fargo

National Institute of Environmental Health Sciences

Charles P. Schmitt

National Institute of Environmental Health Sciences

Lawrence S. Kirschner

National Institute of Environmental Health Sciences

Trevor K. Archer

National Institute of Environmental Health Sciences

Joshua C. Denny

National Institutes of Health

Geoffrey S. Ginsburg

National Institutes of Health

Richard P. Woychik

National Institutes of Health

Alison A. Motsinger-Reif

`alison.motsinger-reif@nih.gov`

National Institute of Environmental Health Sciences

Research Article

Keywords: DNA methylation, long-read sequencing, PacBio HiFi, Oxford Nanopore, epigenetic clocks

Posted Date: September 30th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11079067/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Background

DNA methylation is a central epigenetic mark with broad applications in aging research, epigenome-wide association studies (EWAS), and biomarker development. While the Illumina EPIC array remains the standard platform for large-scale methylation profiling, long-read sequencing technologies, including Oxford Nanopore Technology (ONT) and PacBio HiFi, enable native and bisulfite-free methylation detection at near-genome-wide scale and uniquely offer simultaneous characterization of genetic variation and DNA methylation from a single experiment. Despite their promise, systematic evaluations of measurement reproducibility and cross-platform agreement, particularly for methylation-based prediction models, remain limited.

Methods

We present a cross-platform benchmarking study integrating two population-based datasets: peripheral blood samples from the Personalized Environment and Genes Study (PEGS), assayed using ONT, PacBio HiFi, and the Illumina EPIC array; and 1,787 unique participants from the *All of Us* Research Program (AoU), represented by 1,815 platform-specific long-read datasets, including 28 participants assayed on two platforms. We assessed CpG coverage, cross-platform agreement, the contribution of sequencing depth in simulation, and the performance of 173 methylation-based predictors.

Results

Both ONT and PacBio measured approximately 28 million CpGs at read depth ≥ 10 , including 0.78–0.82 million EPIC-array sites per PEGS sample. The ordinary Pearson correlations were 0.85–0.87 between paired ONT-PacBio measures, whereas mean-centered correlations were much lower, indicating limited agreement for methylation measures at individual CpGs. In simulations calibrated to PEGS EPIC data, depths of approximately 50–60 $\times$ approached a reproducibility comparable to published Illumina array raw data for technical duplicates. Methylation predictor portability increased with model size: 19 of 21 predictors containing $> 1,000$ CpGs had $r > 0.8$ between ONT-PacBio duplicate pairs, whereas only four of 137 predictors containing < 500 CpGs reached the same threshold. We further evaluated pooled sample sequencing as a cost-reduction strategy but found that the resulting read depths were insufficient for accurate individual CpG methylation estimates.

Conclusions

Long-read sequencing provides broad methylome coverage and supports methylation predictors built on larger CpG sets, but current cross-platform agreement does not support direct substitution at individual

CpG without additional depth, calibration, and platform-specific validation.

Background

DNA methylation at CpG dinucleotides is one of the most extensively studied epigenetic modifications in humans, with well-established roles in gene regulation, cellular differentiation, and aging. Over the past decade, epigenome-wide association studies (EWAS) have identified thousands of CpG loci associated with environmental exposures, disease, and aging phenotypes. Central to this progress has been the Illumina Methylation microarray, which simultaneously measures methylation levels at approximately 27K to 850K CpG sites throughout the human genome (1). The Illumina array offers satisfactory data quality and reproducibility at relatively low per-sample cost, making it the dominant platform for population-based methylation research, although measurement noise and technical variability remain non-trivial concerns in large-scale studies (2).

Despite its widespread adoption, the Illumina array covers approximately 3% of the roughly 28 million CpG sites in the human genome and relies on bisulfite conversion, a chemically harsh procedure that degrades input DNA and introduces technical biases requiring computational correction (3, 4). Long-read sequencing platforms, particularly Oxford Nanopore Technologies (ONT) and Pacific Biosciences High-Fidelity (PacBio HiFi) sequencing, offer a fundamentally different approach: methylation is inferred natively from polymerase kinetics (PacBio) (5) or ionic-current signatures (ONT) (6), without bisulfite conversion and with near-genome-wide CpG coverage. Long-read sequencing can also characterize single-nucleotide and structural variation together with DNA methylation (7, 8), making it attractive for epidemiological studies that seek to jointly model genetic and epigenetic variation.

However, several critical questions must be addressed before long-read methylation data can be broadly adopted in population-based epigenomic research. First, how well do methylation measurements from ONT and PacBio agree with each other and with EPIC array data? Second, are long-read methylation measures sufficiently reproducible for use as quantitative phenotypes in EWAS or biomarker studies? Third, can the large library of methylation-based prediction models, including epigenetic clocks for biological age (9–12), disease risk scores, and lifestyle predictors, developed and validated on Illumina array data be reliably applied to long-read data? Fourth, is pooled sample sequencing a viable cost-reduction strategy for epidemiological studies?

To address these questions, we used two population-based datasets. The first comprised peripheral blood samples from the Personalized Environment and Genes Study (PEGS), with paired ONT, PacBio HiFi, and Illumina EPIC-array measurements, together with pooled sequencing experiments. The second was drawn from the *All of Us* (AoU) Research Program (13) and comprised 1,787 unique participants represented by 1,815 samples profiled by ONT, PacBio Revio, or PacBio Sequel IIe sequencing. Twenty-eight participants had measurements from two platforms.

Methods

Study Cohort and Datasets

Personalized Environment and Genes Study (PEGS). Peripheral blood samples were obtained from participants in PEGS (14), a North Carolina population-based epidemiological cohort. High-quality genomic DNA was extracted from whole blood and processed at the National Institute of Environmental Health Sciences (NIEHS) Genomics Core Facility. Genome-wide methylation profiling with the Illumina EPIC array was available for 4,253 samples. To evaluate the concordance between long-read sequencing platforms and the EPIC array, a separate long-read sequencing study was conducted using both individual and pooled sample strategies. For individual sample resequencing, a total of 10 samples were selected and assayed on both Oxford Nanopore Technology (ONT) and PacBio HiFi (Revio) sequencing platforms; one sample was excluded due to poor sample quality, leaving nine samples available for cross-platform analyses. For pooled sample resequencing, two ONT sequencing runs were performed on an identical pool of eight samples, with methylation data aggregated across both runs prior to methylation calling; sequencing was targeted to candidate regions including CpG islands, shores, and shelves. In parallel, two PacBio sequencing runs were conducted, each comprising a distinct pool of four samples (samples 1–4 and samples 5–8, respectively), with genome-wide sequencing measurements obtained for each pool.

All of Us Research Program. The National Institutes of Health *All of Us* Research Program (13) is a large-scale, population-based biomedical cohort collecting broad health and genomic data from participants across the United States. We used 1,815 platform-specific long-read datasets from 1,787 unique participants: 324 ONT, 1,032 PacBio Revio, and 459 PacBio Sequel IIe datasets. Twenty-eight participants were assayed on two platforms, yielding 15 ONT-Revio pairs, 12 ONT-Sequel IIe pairs, and one Revio-Sequel IIe pair for direct cross-platform comparison.

Sequencing and Array Processing

Oxford Nanopore Technologies (ONT) Sequencing. Raw ONT electrical signal data were base-called using the Dorado software tool (Oxford Nanopore Technologies) with specific DNA base modification models to simultaneously identify 5-methylcytosine (5mC) and 5-hydroxymethylcytosine (5hmC). Because the Illumina EPIC array does not distinguish 5mC from 5hmC, the two ONT modification calls were combined for cross-platform comparison. Reads were aligned to GRCh38 using minimap2 (15), and genome-wide CpG methylation summaries were generated with modkit (Oxford Nanopore Technologies). The pipeline was implemented using wf-human-variation (<https://github.com/epi2me-labs/wf-human-variation>). For targeted pooled-sample sequencing, the same pipeline was used with a target BED file (provided by Oxford Nanopore Technologies) to restrict analysis to candidate CpG-rich regions (CpG islands, shores, and shelves).

PacBio HiFi Sequencing. Reads were aligned to GRCh38 using pbmm2 (Pacific Biosciences, <https://github.com/PacificBiosciences/pbmm2>), a minimap2-based (15) alignment tool optimized for PacBio data, with the HiFi preset. CpG methylation scores were extracted using

aligned_bam_to_cpg_scores from the PacBio pb-CpG-tools suite (<https://github.com/PacificBiosciences/pb-CpG-tools>), which estimates 5mC probability at each CpG from polymerase-kinetic signatures in HiFi reads (5).

For both ONT and PacBio, CpG sites with total read counts below 10 were excluded from all analyses to ensure reliable methylation estimates. Sites with read counts exceeding 80 (as recommended by ONT and PacBio technical support; comprising approximately 0.4–0.8% of sites per sample) were additionally excluded, as excessively high coverage may reflect alignment artifacts or reads originating from repetitive genomic regions.

EPIC Array. Whole blood DNA methylation was measured using the Illumina Infinium MethylationEPIC v1.0 BeadChip, interrogating over 850K CpG sites genome-wide. Raw intensity data were processed using the ENmix R package (16). Preprocessing included background correction, dye-bias correction (RELIC) (17), separate quantile normalization of methylated and unmethylated intensities for type I and type II probes, and RCP probe-type bias adjustment (18). Samples were excluded if more than 5% of CpG sites failed detection (detection $p > 1 \times 10^{-6}$ or bead count < 3) or if bisulfite intensity was less than 4,350. CpG probes were excluded if more than 5% of samples failed the quality thresholds.

Correlation and Replicability Analysis

Between-platform agreement was assessed using ordinary Pearson correlation and mean-centered Pearson correlation. For mean centering, the mean methylation value at each CpG was subtracted before calculating correlation between duplicates; this reduces the influence of baseline methylation level differences across CpGs and emphasizes sample-to-sample variation at each CpG site (2). For the simulation analysis, site-specific intraclass correlation coefficients (ICCs) were calculated across duplicate measurements, and $\text{ICC} \geq 0.5$ was treated as an acceptable-reliability threshold.

Simulation Studies

To evaluate the contribution of read depth to reproducibility, simulations were calibrated using CpG-specific methylation means and standard deviations estimated from 4,253 PEGS EPIC-array profiles. One hundred samples with duplicate measurements were simulated at fixed depths from 10× to 100×. At each depth, we summarized the mean-centered correlation between duplicates and the proportion of CpGs with $\text{ICC} \geq 0.5$.

Methylation-Based Prediction Models

A panel of 173 methylation-based predictors was calculated in all datasets using the methscore (19) function in the ENmix R package. Predictors included nine for chronological age, 13 for biological age, five for gestational age, four for stem-cell division, 19 for health-related traits, and 123 for plasma protein

levels. Cross-platform correlations were calculated among participants with paired measurements. We treated $r > 0.8$ as a pragmatic threshold for high agreement and summarized results by predictor model size ($> 1,000$, $500-1,000$, and < 500 required CpGs).

Results

CpG Coverage

Both ONT and PacBio achieved broad genome-wide CpG coverage. After excluding CpG sites with read depth < 10 or > 80 , 27.5–28.1 million CpGs were retained per PEGS sample (Supplementary Table 1), with 28.0–28.7 million per AoU sample (Supplementary Table 2). These counts approach the expected number of autosomal CpGs in the reference genome (20). In PEGS, ONT measurements overlapped 0.80–0.82 million EPIC-array sites per sample, and PacBio measurements overlapped 0.78–0.79 million sites per sample, indicating that both platforms captured most ($\sim 95\%$) array-targeted sites (Supplementary Table 1).

Between-Platform Correlation of Methylation Values

In the PEGS dataset, Pearson correlations between PacBio and ONT methylation values for duplicated samples across all CpGs ranged from 0.847 to 0.869, improving modestly to 0.911–0.922 when restricted to EPIC-overlapping CpGs (Fig. 1A). However, Pearson correlation in this context is largely driven by the wide range of baseline methylation values across CpGs and the relatively small within-CpG variance, making it difficult to distinguish duplicate pairs from non-duplicate pairs based on the ordinary Pearson correlation (Supplementary Fig. 3A). Scatter plots of individual duplicates illustrated that concordance was particularly poor for CpGs with intermediate methylation values (Supplementary Figs. 1–2).

Mean-centered correlations were substantially lower than ordinary Pearson correlations but showed much greater contrast between duplicate and nonduplicate pairs (Fig. 1B; Supplementary Fig. 3B), with correlations for nonduplicate pairs approaching zero. This supports mean-centered correlation as a more sensitive and informative metric of cross-platform agreement in individual CpG variation; it should not, however, be interpreted as a site-specific reliability estimate.

Read Depth and Reproducibility

In the nine PEGS paired samples, ONT showed higher agreement with EPIC than PacBio (Fig. 1). This pattern was consistent with the higher ONT read depth (mean $36\times$, SD 7.0) relative to PacBio (mean $26\times$, SD 1.9) (Supplementary Fig. 4) although the design cannot separate effects of depth from platform chemistry and calling algorithms. AoU read-depth distributions showed a similar pattern but with less difference between the two resequencing platforms (Fig. 2).

To isolate the contribution of read depth, we simulated 100 samples with duplicate measurements at fixed depths from 10× to 100× using CpG site-specific distributions calibrated to 4,253 PEGS EPIC-array profiles (Table 1). Mean-centered correlation and the proportion of CpGs with ICC ≥ 0.5 increased monotonically with depth. At 50×, the simulated centered correlation was 0.61, and 21.6% of CpGs had ICC ≥ 0.5 ; at 60×, the corresponding values were 0.66 and 24.9%. These values bracket a published benchmark for raw 450K-array technical duplicates (centered correlation of approximately 0.64; 22.5% of CpGs with ICC > 0.5) (2). Thus, under the simulation assumptions, depths of approximately 50–60× approached the raw 450K benchmark. This comparison does not demonstrate equivalence to preprocessed EPIC data; in the same reference, preprocessing increased the proportion of CpGs with ICC > 0.5 to 38.5% for Illumina array data. Observed ONT-PacBio-centered correlations were lower than simulation predictions at the corresponding depths, consistent with additional platform-specific error beyond read-depth variability.

Table 1
Simulation study of mean-centered correlation and the percentage of CpGs with ICC ≥ 0.5 at fixed sequencing depths. CpG-specific distributions were calibrated to PEGS EPIC-array data, and 100 samples with duplicate measurements were simulated at each depth.

Read depth (×)	Mean-centered correlation	% of CpGs with ICC ≥ 0.5
10	0.24	3.07
20	0.39	8.97
30	0.49	13.67
40	0.56	17.81
50	0.61	21.57
60	0.66	24.90
70	0.69	27.94
80	0.72	30.59
90	0.74	33.12
100	0.76	35.39

Portability of Methylation Predictors

We next assessed the portability of 173 predictors developed for Illumina array methylation data. Cross-platform correlations between technologies were computed for each predictor, and the proportion of predictors with good reproducibility ($r > 0.8$) was directly proportional to the number of CpGs in each prediction model (Table 2, Supplementary Table 3). Specifically, among predictors with more than 1,000 CpGs, the majority achieved $r > 0.8$ between ONT and PacBio in the *All of Us* dataset (19/21), whereas this proportion dropped sharply for predictors with 500–1,000 CpGs (4/15) and was very low for

predictors with fewer than 500 CpGs (4/137). These descriptive results show an association with model size but do not establish that CpG count alone determines portability.

Table 2

Number of methylation predictors with cross-platform correlation $r > 0.8$, stratified by the number of required CpGs in each prediction model.

	Model size > 1,000 CpGs		Model size 500-1,000 CpGs		Model size < 500 CpGs	
	$r \leq 0.8$	$r > 0.8$	$r \leq 0.8$	$r > 0.8$	$r \leq 0.8$	$r > 0.8$
ONT vs. PacBio (AoU)	2	19	11	4	133	4
ONT vs. PacBio (PEGS)	12	9	15	0	135	2
ONT vs. EPIC	11	10	13	2	135	2
PacBio vs. EPIC	10	11	15	0	130	7

We evaluated eight chronological-age predictors across AoU participants (Fig. 3). In platform-stratified analyses, predictors using more CpGs, particularly principal component (PC) clocks, generally had higher correlations with chronological age. However, PC-clock predictions from long-read data showed compressed age ranges relative to EPIC-based predictions in PEGS (Supplementary Fig. 5). ONT correlations were higher, and prediction ranges were slightly less compressed than PacBio in these unmatched platform groups (Fig. 3).

Methylation Measures in Pooled Samples.

We evaluated two PEGS pooling strategies: two pools of eight identical samples with ONT targeting CpG islands, shores, and shelves, and two pools of four different samples with genome-wide PacBio sequencing. After filtering, 3.8–9.1 million CpGs were retained, depending on sample and platform (Supplementary Table 1). Targeted ONT data overlapped 277,032–421,050 EPIC sites, whereas pooled PacBio data overlapped 79,671–213,983 sites, reflecting the CpG-rich targeting of the ONT design. Effective depth at retained CpGs averaged only 15–20 reads per site. Consistent with the simulations, these depths were insufficient for precise CpG-level estimation. Accordingly, the pooled designs as implemented here are not suitable for analyses requiring individual-level site-specific methylation; their utility for group-level summaries would require separate evaluation.

Discussion

This study presents a comprehensive cross-platform evaluation of PacBio HiFi and ONT long-read sequencing for DNA methylation measurement, assessed against the Illumina EPIC array across two independent population-based cohorts. Our findings address key practical questions surrounding the adoption of long-read sequencing in epigenetic epidemiology.

Coverage. Both long-read platforms achieved near-genome-wide CpG coverage (approximately 28 million sites with read depth ≥ 10), representing a more than 30-fold expansion over the Illumina EPIC array (1). This coverage creates opportunities to study regulatory regions, transposable elements, imprinted loci, and other contexts underrepresented on commercial arrays (21).

Cross-platform agreement and sequencing depth. Ordinary Pearson correlation is dominated by between-CpG methylation differences rather than within-CpG agreement and is thus insufficient by itself for evaluating platform agreement. Mean-centered correlation more directly summarizes within-CpG agreement across samples (2), but it remains an aggregate metric rather than a site-specific reliability measure. Under the simulation assumptions, approximately 50–60× depth approached a published raw 450K technical-duplicate benchmark. The lower observed ONT-PacBio agreement indicates that platform chemistry, base-modification models, alignment, and methylation callers contribute error beyond read depth. Compared to EPIC array data, the higher concordance of ONT relative to PacBio in the PEGS dataset appears to be driven primarily by the higher read depths achieved by ONT (mean 36×) compared to PacBio (mean 26×), rather than a fundamental difference between the two sequencing chemistries, consistent with simulation predictions.

Pooled sample sequencing. The pooled designs provided fewer retained CpGs and lower effective depth than individual sequencing. The greater EPIC overlap for targeted ONT reflected the intended enrichment of CpG-rich regions, not necessarily better measurement quality. As implemented here, neither pooling strategy supported precise individual-level CpG estimates, so these designs are not appropriate for conventional EWAS. This conclusion is specific to the tested designs; pooling could still be evaluated for group-level summaries or with methods that preserve sample identity and adequate per-sample depth.

Model transferability. Larger and PC-based predictors generally transfer better across platforms, plausibly because they can aggregate information across thousands of CpGs and are thus robust to measurement noise. Model size is nevertheless entangled with architecture, training, normalization, and missing-CpG handling, so it should not be interpreted as the sole determinant of portability. Sparse clocks such as HorvathAge (353 CpGs) (22) and HannumAge (71 CpGs) (23) showed unstable cross-platform correlations. Both PC and non-PC predictors require platform-specific validation of calibration before they can be reliably applied to long-read data.

ONT versus PacBio. ONT performed modestly better than PacBio in most comparisons, with higher individual read depths, stronger predictor correlations with chronological age, and less compression in predicted age ranges. This advantage likely reflects higher per-sample sequencing depth rather than a fundamental chemistry difference, consistent with prior work (8). However, platform, depth, chemistry, caller, and participant composition were not independently varied here, so intrinsic platform superiority cannot be established from these data. The *All of Us* dataset also allowed comparison between two PacBio instrument generations, Revio and Sequel IIe. Despite Revio's higher throughput and improved on-instrument methylation calling, the two performed comparably, with Sequel IIe achieving slightly higher depth. However, this comparison rests on a single duplicate pair and is not sufficient to draw firm

conclusions about instrument-generation effects. Together, these results suggest read depth is likely the primary driver of reproducibility across platforms and generations, but confirming this will require future studies with balanced within-platform technical replicates and depth-matched cross-platform samples.

Limitations. The PEGS replicability analysis was based on nine samples, limiting the precision of ICC estimates. The *All of Us* data do not include matched EPIC array measurements, precluding a direct three-way comparison in that cohort. Additionally, ONT data captures both 5mC and 5hmC, which were combined for cross-platform comparisons since PacBio and the EPIC array do not distinguish these modifications; while 5hmC levels in blood are generally low, this may introduce minor systematic differences. Finally, PEGS duplicates represent cross-platform replicates (ONT, PacBio, and EPIC array) rather than within-platform technical replicates, precluding a formal assessment of within-platform reproducibility and potentially contributing to the lower-than-expected centered correlations observed in that cohort.

Conclusions

ONT and PacBio HiFi both achieve near-genome-wide CpG coverage in peripheral blood, with methylation values broadly concordant across platforms. At current sequencing depths, however, cross-platform agreement at the individual-CpG level remains modest, falling short of array-based standards. Pooled sequencing strategies, while economically appealing, do not provide adequate depth for reliable individual-level site-specific estimation and are not recommended for epidemiological studies requiring precise CpG-level measurements. Among methylation predictors, large models, particularly PC-based clocks, were more portable across platforms than sparse models although all require platform-specific validation of calibration. Long-read sequencing's unique capacity to jointly characterize genetic variation and DNA methylation from a single experiment positions it as a promising tool for future epigenetic epidemiology, provided that sequencing depth, assay-specific error, and model validation are carefully addressed.

Declarations

Ethics approval and consent to participate

All PEGS and *All of Us* participants provided written informed consent for use of their data.

Consent for publication

Not applicable

Availability of data and materials

This study used data from PEGS and the NIH *All of Us* Research Program. PEGS data are available to qualified researchers upon request via (<https://www.niehs.nih.gov/research/atniehs/labs/crb/studies/pegs>). *All of Us* data are available

through controlled access via the Researcher Workbench (<https://www.researchallofus.org/>), requiring registration and required training; individual-level data cannot be downloaded and must be analyzed within the secure Workbench environment. Analyses used the Controlled Tier Dataset v9.

Competing interests

The authors have no competing interests to declare.

Funding

This research was supported by the Intramural Research Program of the NIH, National Institute of Environmental Health Sciences. The contributions of the NIH author(s) are considered works of the United States Government. The findings and conclusions presented in this paper are those of the authors and do not necessarily reflect the views of the NIH or the U.S. Department of Health and Human Services.

Acknowledgements

We thank the participants in the PEGS cohort and the *All of Us* Research Program who made this work possible.

References

1. Pidsley R, Zotenko E, Peters TJ, Lawrence MG, Risbridger GP, Molloy P, et al. Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. *Genome Biol.* 2016;17(1):208.
2. Xu Z, Taylor JA. Reliability of DNA methylation measures using Illumina methylation BeadChip. *Epigenetics.* 2021;16(5):495–502.
3. Tanaka K, Okamoto A. Degradation of DNA by bisulfite treatment. *Bioorg Med Chem Lett.* 2007;17(7):1912–5.
4. Olova N, Krueger F, Andrews S, Oxley D, Berrens RV, Branco MR et al. Comparison of whole-genome bisulfite sequencing library preparation strategies identifies sources of biases affecting DNA methylation data. *Genome Biol.* 2018;19.
5. Ni P, Nie F, Zhong ZY, Xu JR, Huang N, Zhang J et al. DNA 5-methylcytosine detection and methylation phasing using PacBio circular consensus sequencing. *Nat Commun.* 2023;14(1).
6. Simpson JT, Workman RE, Zuzarte PC, David M, Dursi LJ, Timp W. Detecting DNA cytosine methylation using nanopore sequencing. *Nat Methods.* 2017;14(4):407–10.
7. Liu Y, Rosikiewicz W, Pan ZW, Jillette N, Wang P, Taghbalout A et al. DNA methylation-calling tools for Oxford Nanopore sequencing: a survey and human epigenome-wide evaluation. *Genome Biol.* 2021;22(1).

8. Sigurpalsdottir BD, Stefansson OA, Holley G, Beyter D, Zink F, Hardarson MP et al. A comparison of methods for detecting DNA methylation from long-read sequencing of human genomes. *Genome Biol.* 2024;25(1).
9. Horvath S. DNA methylation age of human tissues and cell types. *Genome Biol.* 2013;14(10).
10. Hannum G, Guinney J, Zhao L, Zhang L, Hughes G, Sada S, et al. Genome-wide Methylation Profiles Reveal Quantitative Views of Human Aging Rates. *Mol Cell.* 2013;49(2):359–67.
11. Levine ME, Lu AT, Quach A, Chen BH, Assimes TL, Bandinelli S, et al. An epigenetic biomarker of aging for lifespan and healthspan. *Aging-Us.* 2018;10(4):573–91.
12. Lu AT, Quach A, Wilson JG, Reiner AP, Aviv A, Raj K, et al. DNA methylation GrimAge strongly predicts lifespan and healthspan. *Aging-Us.* 2019;11(2):303–27.
13. All of Us Research, Program I, Denny JC, Rutter JL, Goldstein DB, Philippakis A, Smoller JW, et al. The All of Us Research Program. *N Engl J Med.* 2019;381(7):668–76.
14. Akhtari FS, Madenspacher J, D'Agostin D, Shuptrine S, MacNell N, Ritter R et al. Personalized Environment and Genes Study (PEGS) Dataset-a resource for genomic, exposomic, and geospatial data. *Sci Data.* 2026;13(1).
15. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics.* 2018;34(18):3094–100.
16. Xu Z, Niu L, Li L, Taylor JA. ENmix: a novel background correction method for Illumina HumanMethylation450 BeadChip. *Nucleic Acids Res.* 2016;44(3):e20.
17. Xu ZL, Langie SAS, De Boever P, Taylor JA, Niu L. RELIC: a novel dye-bias correction method for Illumina Methylation BeadChip. *BMC Genomics.* 2017;18.
18. Niu L, Xu ZL, Taylor JA. RCP: a novel probe design bias correction method for Illumina Methylation BeadChip. *Bioinformatics.* 2016;32(17):2659–63.
19. Xu Z, Niu L, Kresovich JK, Taylor JA. methscore: a comprehensive R function for DNA methylation-based health predictors. *Bioinformatics.* 2024;40(5).
20. Raine A, Liljedahl U, Nordlund J. Data quality of whole genome bisulfite sequencing on Illumina platforms. *PLoS ONE.* 2018;13(4):e0195972.
21. Katsman E, Orlanski S, Martignano F, Fox-Fisher I, Shemer R, Dor Y et al. Detecting cell-of-origin and cancer-specific methylation features of cell-free DNA from Nanopore sequencing. *Genome Biol.* 2022;23(1).
22. Horvath S. DNA methylation age of human tissues and cell types. *Genome Biol.* 2013;14(10):R115.
23. Hannum G, Guinney J, Zhao L, Zhang L, Hughes G, Sada S, et al. Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol Cell.* 2013;49(2):359–67.

Figures

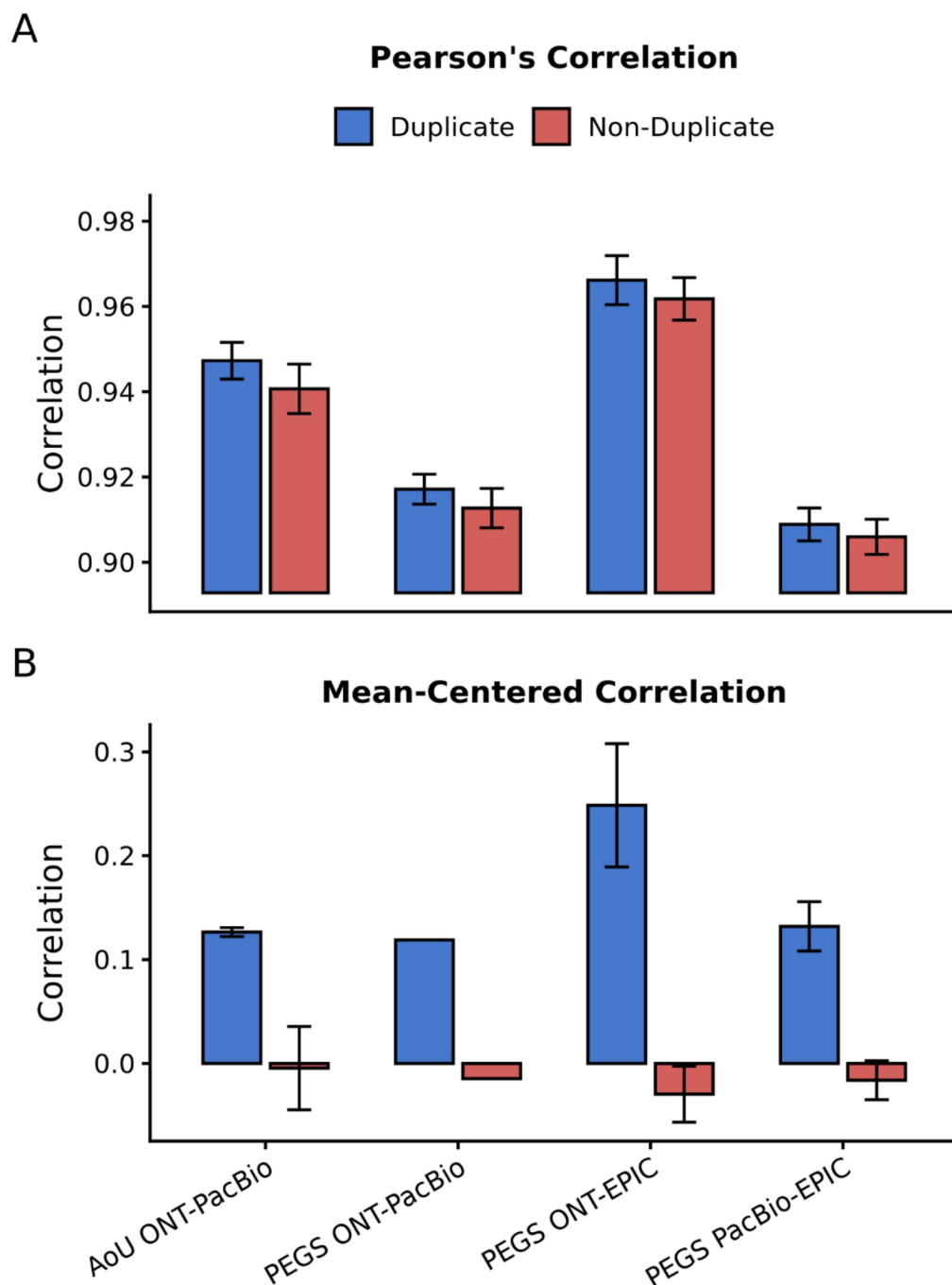


Figure 1

Cross-platform methylation agreement for duplicate and non-duplicate pairs. (A) Ordinary Pearson correlation across CpGs. (B) Mean-centered Pearson correlation. Bars show mean correlation; error bars indicate standard deviation.

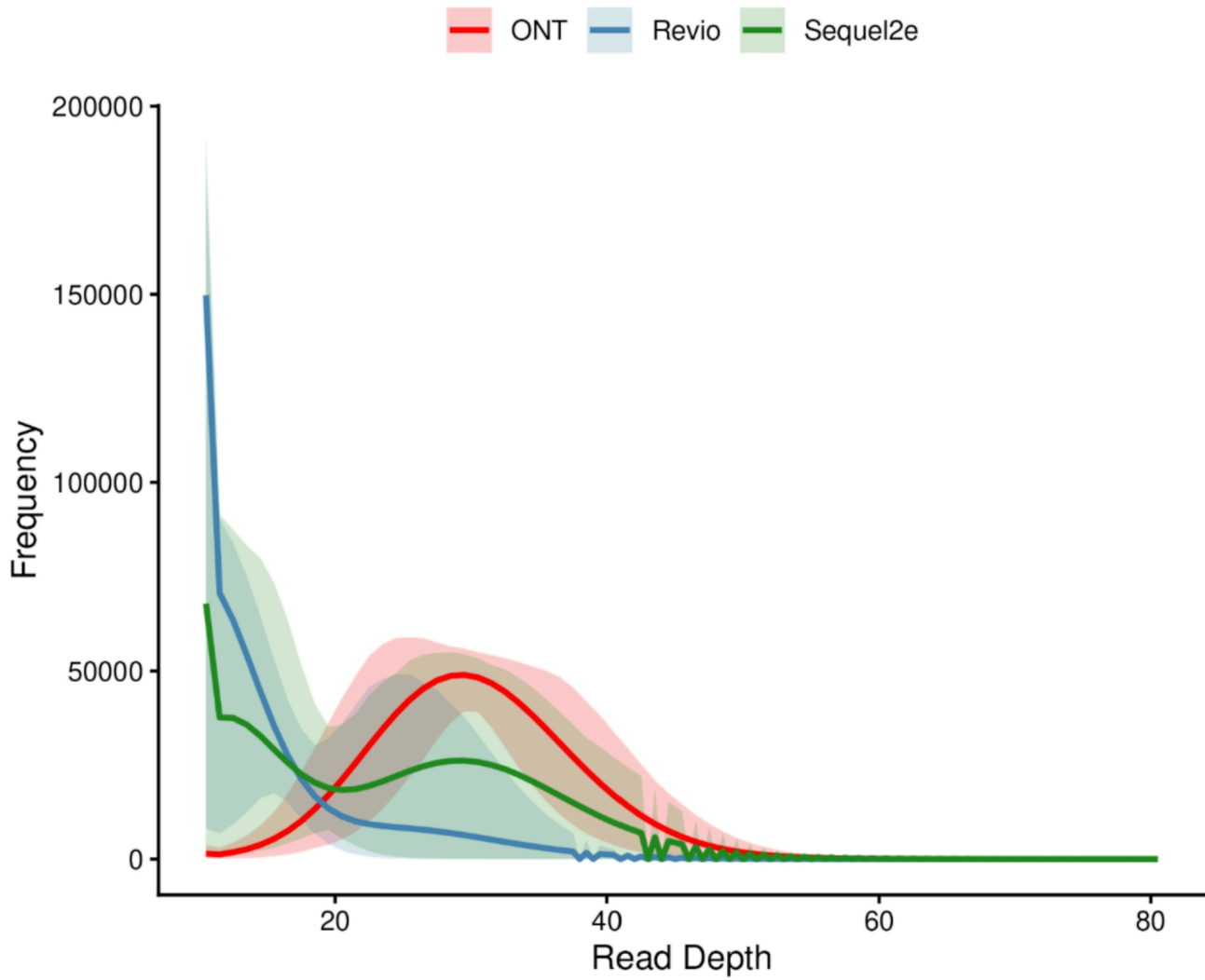


Figure 2

CpG-level read-depth distributions for ONT, PacBio Revio, and PacBio Sequel IIe datasets in *All of Us*. Lines show the mean read-depth frequency across samples for each platform; shaded ribbons denote the interquartile range (10th–90th percentile) across samples.

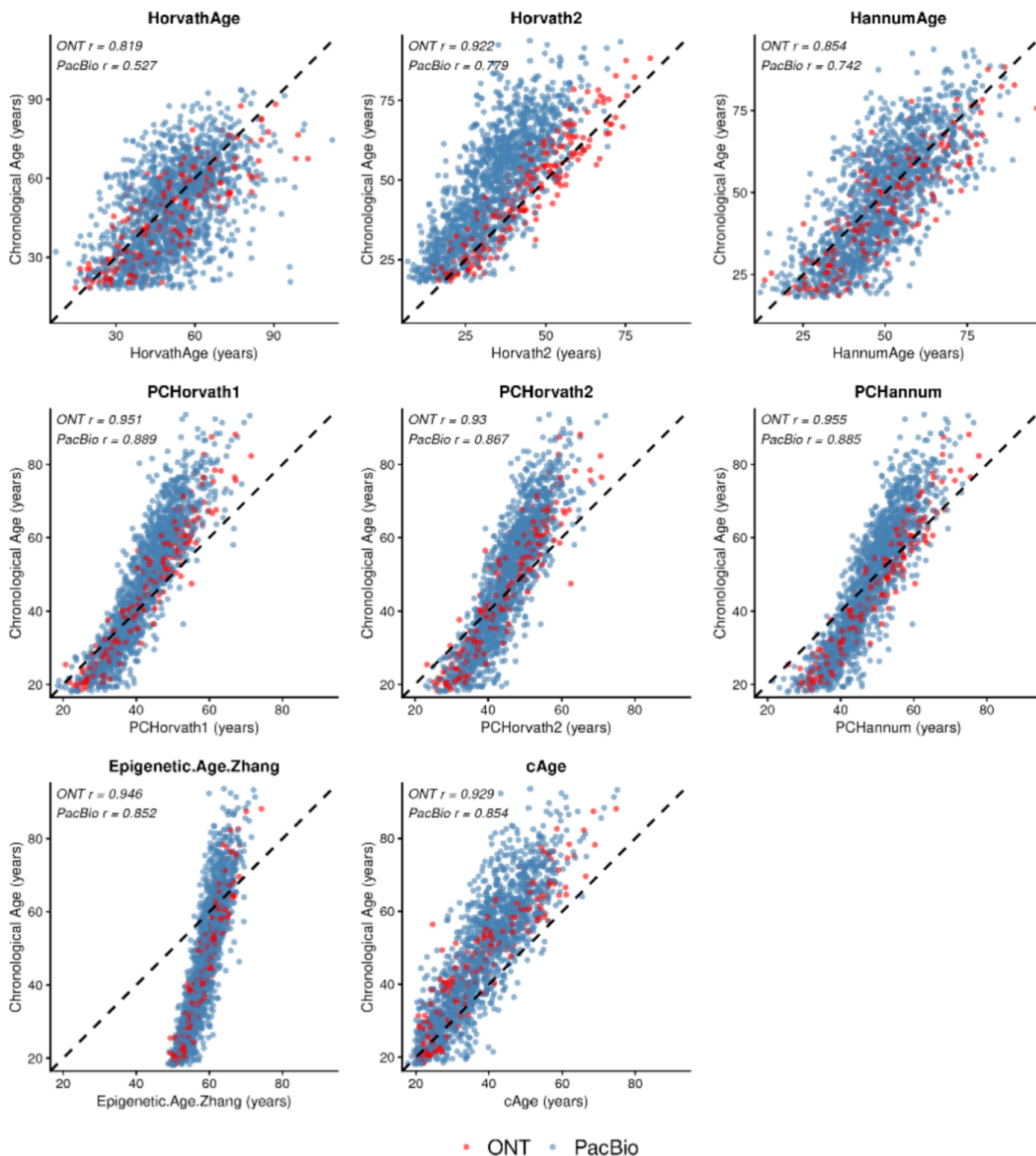


Figure 3

Chronological age versus methylation-predicted age for eight predictors across AoU participants with ONT ($n=324$, red) or PacBio ($n=1491$, blue) data. The black dashed line denotes identity ($y = x$), and r denotes the Pearson correlation for each platform.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [supplementaryResults09162026.docx](#)