Accounting for differences between Infinium MethylationEPIC v2 and v1 in DNA methylation-based tools

Beryl C. Zhuang†[1,2,3], Marcia Smiti Jude†[1,2,3], Chaini Konwar[1,2,3], Natan Yusupov[4,5], Calen P. Ryan[6], Hannah-Ruth Engelbrecht[1,2,3], Joanne Whitehead[1,2,3], Alexandra A. Halberstam[4,7], Julia L. MacIsaac[1,2,3], Kristy Dever[1,2,3], Toan Khanh Tran[8], Kim Korinek[9], Zachary Zimmer[10,11], Nanette R. Lee[12], Thomas W. McDade[13,14], Christopher W. Kuzawa[15], Kim M. Huffman[16], Daniel W. Belsky[6,17], Elisabeth B. Binder[4], Darina Czamara[4], Keegan Korthauer[3,18], Michael S. Kobor*[1,2,3,14]

† These authors contributed equally and should be considered as co-first authors

* Please address all correspondence concerning this article to Dr. Michael S. Kobor at michael.kobor@ubc.ca

1. Edwin S. H. Leong Centre for Healthy Aging, Faculty of Medicine, University of British Columbia , Vancouver, British Columbia, Canada
2. Centre for Molecular Medicine and Therapeutics and Department of Medical Genetics, University of British Columbia, Vancouver, British Columbia, Canada
3. British Columbia Children's Hospital Research Institute, Vancouver, British Columbia, Canada
4. Department Genes and Environment, Max Planck Institute of Psychiatry, Munich, Germany
5. International Max Planck Research School for Translational Psychiatry, Munich, Germany
6. Robert N. Butler Columbia Aging Center, Mailman School of Public Health, Columbia University, New York, United States
7. Harvard Medical School/MIT Institute of Technology MD-PhD program, Massachusetts, United States
8. Family Medicine Department, Hanoi Medical University, Hanoi, Vietnam
9. Department of Sociology, University of Utah, Salt Lake City, Utah, United States
10. Department of Family Studies and Gerontology, Mount Saint Vincent University, Halifax, Canada

11. Canada Research Chair, Global Aging and Community Initiative, Canada
12. USC-Office of Population Studies Foundation, Inc., University of San Carlos, Cebu City, Philippines
13. Department of Anthropology, Northwestern University, Evanston, Illinois, United States
14. Program in Child and Brain Development, Canadian Institute for Advanced Research, Toronto, Ontario, Canada
15. Department of Anthropology and Institute for Policy Research, Northwestern University, Evanston, Illinois, United States
16. Duke University School of Medicine, Durham, United States
17. Department of Epidemiology, Columbia University Mailman School of Public Health, New York, United States
18. Department of Statistics, Faculty of Science, University of British Columbia (Vancouver, British Columbia, Canada

Summary blurb

Our work illustrates the importance of accounting for EPIC version in DNA methylation-based tools, specifically in research scenarios when samples profiled across different versions are harmonized.

Keywords

Epigenetics, DNA methylation, Illumina EPIC array, epigenetic clocks, inflammation and lifestyle biomarkers, cell type deconvolution

Abstract

The recently launched Illumina Infinium MethylationEPIC v2.0 (EPICv2), successor of MethylationEPIC v1.0 (EPICv1), retains a majority of probes in EPICv1, while expanding coverage of regulatory elements.

The concordance between the two EPIC versions in DNA methylation-based tools has not yet been investigated. To address this, DNA methylation was profiled on both versions using matched blood samples across four cohorts spanning early to late adulthood. High concordance between versions at the array level but variable agreement at the individual probe level was noted. A significant contribution of EPIC version to DNA methylation variation was observed, though it was to a smaller extent compared to sample relatedness and cell type composition. Modest but significant differences in DNA methylation-based estimates between versions were observed, irrespective of the data preprocessing method used. Adjustments for EPIC version or calculation of estimates separately for each version largely mitigated these version-specific discordances. This work emphasizes the importance of accounting for EPIC version differences in research scenarios, especially in meta-analyses and longitudinal studies that require data harmonization across versions.

Introduction

Infinium Methylation BeadChip microarrays have been widely used to cost-effectively measure the human DNA methylome in large scale and population-wide studies(Battram et al. 2022; Kuzawa et al. 2022; Li et al. 2019). The recently developed Illumina MethylationEPIC BeadChip Infinium microarray v2.0 (900K, EPICv2) features a total of 936,866 probes, encompassing ~77% of the probes in the previous version, the MethylationEPIC BeadChip Infinium microarray v1.0 B5 (850K, EPICv1), and over 200,000 new probes designed for increased coverage of enhancers, open chromatin regions, and CTCF-binding domains(Infinium MethylationEPIC v2.0 BeadChip 2023). EPICv2 also differs from its predecessor in the overall probe content and utility, with annotation to the most recent GRCh38/h38 human genome build, differences in probe design type and strand switches, and incorporation of new “nv” probes targeting recurrent somatic cancer mutations. Unlike EPICv1 where each probe is unique, EPICv2 includes ~5100 probes that each have between 2-10 replicates, differentiated based on their probe names and

sequences(Peters et al. 2024). Approximately 143,000 poorly performing probes on the EPICv1 have been removed from the EPICv2, ~73% of which are likely to be influenced by underlying sequence polymorphisms(Kaur et al. 2023; Noguera-Castells et al. 2023). Overall, these modifications in EPICv2 intend to provide wider coverage of the DNA methylome, with optimized performance across primary tissues and cancer cell lines, and extended reliability across diverse human populations(Kaur et al. 2023; Noguera-Castells et al. 2023).

Previous iterations of Illumina Infinium microarrays (27K, 450K and EPICv1) have been widely used to develop DNA methylation-based bioinformatic tools including cell type deconvolution algorithms(Houseman et al. 2014; Salas et al. 2018, 2022), a rapidly increasing and diverse set of epigenetic clocks(Horvath 2013; Hannum et al. 2013; Horvath et al. 2018; Levine et al. 2018; Lu et al. 2019a, 2019b; Belsky et al. 2022; Yang et al. 2016), interleukin 6 (IL-6) and C-reactive protein (CRP) inflammation markers(Stevenson et al. 2021; Wielscher et al. 2022), and lifestyle biomarker predictors such as smoking and alcohol use(McCartney et al. 2018). Simplistically speaking, these tools are based on the strong correlations of DNA methylation levels at specific cytosine-guanine dinucleotides (CpGs) with measured cell types, chronological age, and biomarker measures, respectively. The currently available tools have been exclusively trained on one or more of the previous generation of microarrays(Horvath 2013; Hannum et al. 2013; Horvath et al. 2018; Levine et al. 2018; Lu et al. 2019b; Belsky et al. 2022), and many, but not all, of predictive CpGs employed by these tools are retained on EPICv2(Kaur et al. 2023). Illustrating the relevance of array iteration, estimates from some of these DNA methylation-based tools, while highly correlated, are significantly different between 450K and EPICv1 using matched samples(Solomon et al. 2018; McEwen et al. 2018; Dhingra et al. 2019) and between EPICv1 and EPICv2(Lussier et al. 2024; Tay et al. 2025). At a more basic level, DNA methylation profiles derived from a limited set of human cell lines suggest a high agreement between EPICv1 and EPICv2. Given the extensive use of Illumina DNA methylation arrays in primary human samples, and importantly, across large population studies, it is imperative to determine the concordance of EPICv1 and EPICv2 in these more complex yet highly relevant research settings.

101 Here, we used an unprecedented set of 67 primary human population samples and five technical
102 replicates to systematically assess the concordance of inferred estimates derived from a broad set of DNA
103 methylation-based tools between EPICv1 and EPICv2. We also tested whether EPIC version differences
104 might affect meta-analyses, which harness statistical power that comes from combining multiple cohorts,
105 and longitudinal studies, which often include samples profiled on different arrays/versions. Specifically,
106 we created sets of matched venous blood samples from 67 individuals across a diverse collection of four
107 populations spanning early to late adulthood, and profiled the DNA methylome using the EPICv1 and
108 EPICv2 arrays. Using this unique dataset, we tested the concordance of the two EPIC versions at both array
109 and probe levels and illustrated the potential contribution of EPIC version to overall DNA methylation
110 variation. To explore if these version differences are reflected in DNA methylation-based tools, including
111 immune cell type deconvolution algorithms, epigenetic clocks, inflammation and lifestyle biomarkers, we
112 compared their estimates between the two EPIC versions, and confirmed that these differences persisted
113 irrespective of data preprocessing methods. We also demonstrated different remediation methods to account
114 for these EPIC version discrepancies in epigenetic investigations. Collectively, this work encourages
115 careful consideration while harmonizing data profiled on the two EPIC versions and comparing epigenome-
116 wide association studies' (EWAS) findings from one version to another.

117 **Results**

118 EPIC version significantly influenced hierarchical clustering patterns and DNA methylation
119 variance, alongside sample relatedness and estimated cell type proportions
120
121 To compare the two most recent generations of Illumina MethylationEPIC BeadChip Infinium
122 microarrays, we measured the DNA methylomes of a subset of venous whole blood samples on both
123 EPICv1 and EPICv2 across three cohorts in the Kobor Lab, Vancouver, Canada (in-house facility): (i)
124 Vietnam Health and Aging Study (VHAS)(Korinek et al. 2019), (ii) Cebu Longitudinal Health and
125 Nutrition Survey (CLHNS)(Adair et al. 2011), and (iii) Comprehensive Assessment of Long-term Effects
126 of Reducing Intake of Energy (CALERIE)(Rochon et al. 2011), and an external cohort processed in a

127 different facility (Max Planck Institute of Psychiatry in Munich, Germany): Biological Classification of
128 Mental Disorders study (BeCOME)(Brückl et al. 2020) (Table 1, Figure 1). These four cohorts represented
129 distinct demographic and biological characteristics such as sex and age range. When we performed
130 clustering using the 57 SNP probes available on both EPIC versions, as expected, we noted that matched
131 samples from the same individual assessed on EPICv1 and EPICv2 grouped together (Supplementary Fig.
132 S1).

133 We first employed sample-to-sample Pearson correlations in pair-wise manner using the 721,378
134 probes shared between EPICv1 and EPICv2 using the functional normalized data. Array level Pearson
135 correlations were then hierarchically clustered in an unsupervised manner using complete linkage with
136 Euclidean distance (Figure 2A and 2B). Examining these clustering patterns in a cohort-wise manner, we
137 expectedly noted that samples clustered first by sex, a known biological driver of DNA methylation,
138 followed by EPIC version (Figure 2B). When clustering was performed on all four cohorts combined, we
139 observed a similar pattern (Figure 2A). Perhaps not surprisingly, samples from the BeCOME cohort, which
140 were processed at an external facility, clustered relatively separately from the samples of the other three
141 cohorts, highlighting the influence of the processing facility on some of the observed clustering patterns.
142 When smaller subsets of predictive CpGs employed by DNA methylation-based tools (Table 2) were used
143 for Pearson correlation calculation followed by clustering, we observed partial separation of samples based
144 on EPIC version, but not by sex (Supplementary Fig. S2), which may be due to these CpG subsets being
145 specifically enriched for predicting the targeted biological variable, thereby potentially reducing sensitivity
146 to sex-specific differences. In addition to Pearson correlations, to capture systematic, potentially non-linear
147 shifts in rank correlation, we calculated sample-to-sample Spearman rank correlations and subsequently
148 performed unsupervised hierarchical clustering using complete linkage with Euclidean distance. We
149 observed a clear demarcation of samples first by EPIC version within each cohort (Supplementary Fig. S3),
150 and sample clustering first by the facility they were processed and then by version when all four cohorts
151 were combined using functional normalized data (Supplementary Fig. S4).

152 Next, we performed principal component analysis (PCA) to capture the linear association between
153 top DNA methylation PCs and variables of interest, and measured the proportion of linear variance in PCs
154 that can be explained by each variable. The PC1 loadings, corresponding to >98% of DNA methylation
155 variation, were significantly associated with donor, measuring relatedness of matched samples, estimated
156 cell type proportions, and EPIC version, although EPIC version had a relatively smaller percentage
157 contribution, as assessed by adjusted R^2, compared to donor and a subset of estimated cell type
158 proportions (Supplementary Fig. S5A and S5C). When we similarly performed this analysis on an
159 individual cohort-basis, we noted that EPIC version, estimated cell types, and donor were associated with
160 PC1 with varying percentage contributions depending on the cohort investigated, suggesting that sample
161 sizes are a likely determinant of the relative contribution of each variable (Figure 2C, Supplementary Fig.
162 S5B).

163 **Table 1.** Demographics of the three in-house cohorts and external validation cohort.

164

	VHAS	CLHNS	CALERIE	External validation cohort: BeCOME
Samples (n)	48 (24 matched venous and capillary samples)	16	24	8
Country of origin	Vietnam	Philippines	United States of America	Germany
Biological sex (% Female)	36.4%	100%	54.2%	62.5%
Average age in years ± standard deviation	74.6 ± 6.7	48.4 ± 6.1	36.7 ± 8.1	40.0 ± 16.9

165

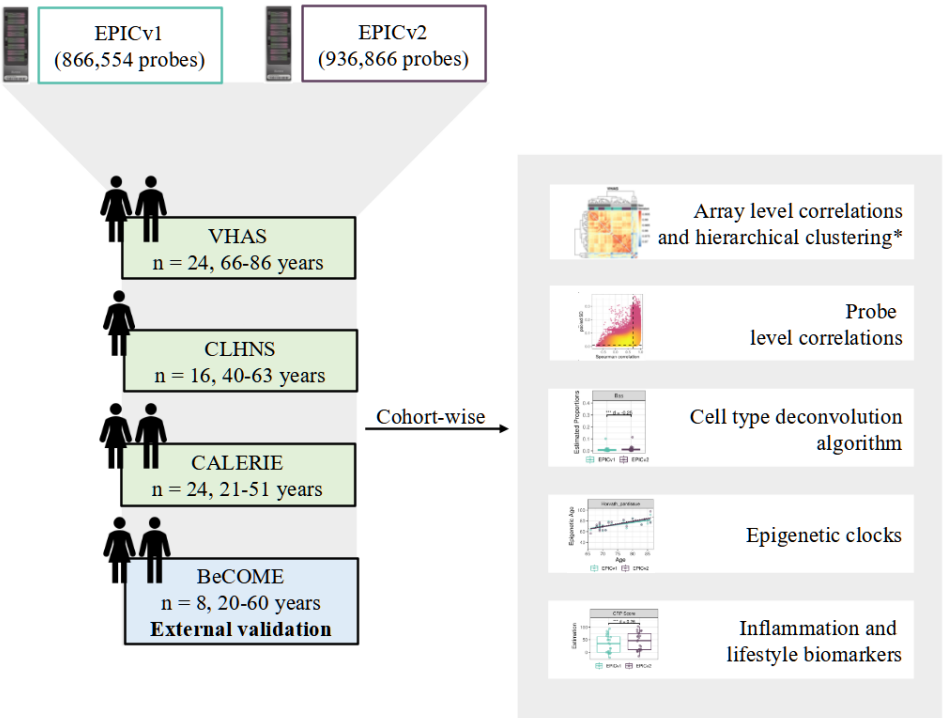
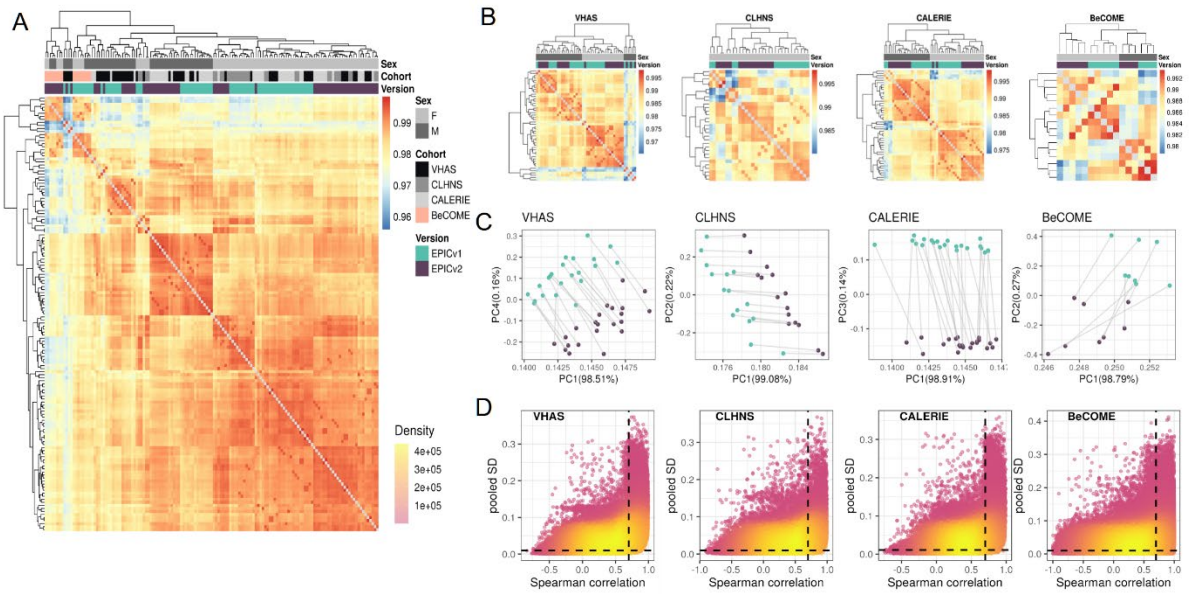


Figure 1. Overview of study design and analyses.

*Array level correlations and hierarchical clustering were also performed by combining cohorts.



170 **Figure 2. (A)** Unsupervised hierarchical clustering using complete linkage with Euclidean distance on
171 sample-to-sample Pearson correlations, calculated using the 721,378 probes shared between EPICv1 and
172 EPICv2 with functional normalized data across all four cohorts; blue to red color range denotes Pearson
173 correlation from 0.95-1.00. **(B)** Cohort-wise unsupervised hierarchical clustering on sample-to-sample
174 Pearson correlations. **(C)** Cohort-wise principal component analysis (PCA). Accounted variance of PCs are
175 shown in brackets in the x- and y-axis label. Grey lines indicate matched samples from the same donor
176 profiled on both EPICv1 and EPICv2. **(D)**. Probe-level Spearman correlation and pooled standard deviation
177 (SD) of common probes. The X-axis represents the Spearman correlation and Y-axis represents the pooled
178 SD of probes common to both EPIC versions. Dashed horizontal line indicates the pooled SD threshold set
179 at lower quartile pooled SD for each cohort, and the vertical dashed line indicates the correlation threshold
180 set at 0.70. The colors indicate the density of points, such that pink is low density and yellow is high density.
181 Probes unique to either version are not shown.

182

183 **EPICv1 and EPICv2 probes shared consistently high correlation at array level but not at the**
184 **individual probe level**

185 Within three cohorts, technical replicate samples derived from the same individual after bisulfite
186 conversion were quantified on both EPIC versions at least twice, allowing us to examine the technical
187 variation within each version. Given that all samples were matched on both versions, we were also able to
188 compare these technical replicate samples between versions. We first used Spearman correlation to assess
189 array level concordance in technical replicates, calculated by averaging DNA methylation (β) values across
190 shared probes on per sample basis. We showed high concordance between both between-versions and
191 within-version technical replicates, though the latter was relatively more correlated (between-versions
192 correlations: 0.9737-0.9774; within-version: 0.9858-0.9916). Next, we determined the reliability of
193 technical replicates both between- and within-version using intra-class correlation coefficients (ICC), and
194 found slightly lower agreement in between-versions technical replicates (0.9947-0.9958) than within-

195 version (0.9972-0.9983) (Supplementary Table S1). Finally, we assessed the error in technical replicates
196 between and within EPIC versions using root mean square root error (RMSE). As expected, the mean
197 RMSE between-versions (0.0367-0.0416) was slightly higher than the within-version (0.0214-0.0264)
198 (Supplementary Table S1).

199 Extending beyond the technical replicates to all the matched samples profiled on each version, we
200 noted high array level Spearman correlation ranging from 0.968 to 0.981 between EPICv1 and EPICv2,
201 though this was at the array level (Supplementary Table S2). In contrast to the strong array level
202 concordance, Spearman correlations at the probe level, calculated as the concordance of β values of
203 individual probes across samples between versions, were not as congruent. Specifically, only ~25% the
204 probes per cohort had Spearman correlations greater than 0.70 while the remaining ~75% probes showed a
205 low mean correlation of <0.30 (Supplementary Table S3, Figure 2D). These findings remained consistent
206 regardless of the level of data processing such as background correction or normalization, as similar patterns
207 were observed even when using raw data (Supplementary Table S4). Furthermore, on comparing these low-
208 concordance probes to previously identified low-concordant probes across 450K and EPICv1 (Logue et al.
209 2017; Sugden et al. 2020; Olstad et al. 2022), we observed that 73-93% probes in these reference lists
210 overlapped with our set of probes (Supplementary Table S5). One plausible explanation for low probe level
211 correlations of the majority of probes may be low inter-sample variability. It has been previously reported
212 that probes with a narrow range of DNA methylation β values across samples tend to have poor
213 correlations(Solomon et al. 2018). To test this, we calculated pooled standard deviation (SD) across samples
214 on a cohort basis, and defined the first lower quartile (≈ 0.01 for each cohort) as the pooled SD threshold
215 to denote low variability (Supplementary Table S2). Supporting our hypothesis, we found that probes with
216 Spearman correlations ≤ 0.70 had significantly lower pooled SD or low variability compared to probes with
217 higher correlations (Supplementary Table S3). Specifically, of the probes with Spearman correlation ≤ 0.70,
218 ~30% exhibited low inter-sample variability; in contrast, of probes with Spearman correlation > 0.70, only
219 0.069-5.577% exhibited low variability (Supplementary Table S6). We further evaluated probe variability
220 and Spearman correlation on the 2,169 probes selected by both probe selection methods implemented in

221 the *FlowSorted.Blood.EPIC* R package for cell type deconvolution (see Methods). As these probes are
222 specifically selected for their high variability to distinguish between cell types, as expected, they
223 demonstrated significantly greater variability and higher Spearman correlations across all cohorts compared
224 to the remaining shared probes (Supplementary Table S7). We also cross-referenced these poorly-
225 concordant probes (Supplementary Table S8), which we defined as probes with low correlation despite
226 their high variability, with low quality and unreliable probes identified in previous annotations of the
227 Illumina arrays(Price et al. 2013; Pidsley et al. 2016; McCartney et al. 2016; Peters et al. 2024). Specifically,
228 0.11-2.39% of these poorly-concordant probes overlapped with previously identified as cross hybridizing
229 to multiple genomic locations or mapped to genetic variant sites across all cohorts(Price et al. 2013; Pidsley
230 et al. 2016; McCartney et al. 2016; Peters et al. 2024) (Supplementary Table S9). Of the 82 probes that
231 have undergone design type switches in EPICv2(Kaur et al. 2023), 27-40 probes also overlapped with our
232 poorly-concordant probes in the four cohorts. Overall, only a small fraction (~10%) of these poorly-
233 concordant probes overlapped with probes of bad quality, design type switch(Kaur et al. 2023), and cross
234 hybridization to the genome(Price et al. 2013; Pidsley et al. 2016; McCartney et al. 2016; Peters et al. 2024)
235 (Supplementary Table S9), while the rest still remained unexplained. In contrast, of the highly correlated
236 or low variability probes, ~80% overlapped with platform-bias free and high-confidence mapping probes
237 recently identified in cell lines(Chen and Zhou 2024), confirming the agreement of the readouts at these
238 probes (Supplementary Table S9).

239 We used the first lower quartile as the pooled SD threshold and Spearman correlation ≤ 0.70 to
240 denote probes with low concordance and high variability, however we acknowledge that there is a
241 continuum in the relationship between variability and correlation (Figure 2D). To examine if a more lenient
242 threshold of variability and correlation would result in fewer unreliable probes shared between versions,
243 we additionally employed a pooled SD threshold of 0.05 and Spearman correlation threshold of 0.50. When
244 we relaxed the pooled SD threshold to 0.05, while maintaining the Spearman correlation threshold of 0.70,
245 the percentage of unreliable probes reduced to 1.7-6.5% of the probes shared between versions; a further
246 decrease to less than 0.6% of shared probes was noted when the Spearman correlation threshold was

247 reduced to 0.50 (Supplementary Table S6). In summary, we observed high concordance at the array level
248 between versions; however, on the probe level, there were a subset of probes with low concordance between
249 versions regardless of high inter-sample variability.
250

251 **Immune cell type proportions inferred by IDOL and auto probe selection methods were significantly**
252 **different between EPIC versions**

253 Cellular composition is a key contributor to whole blood DNA methylation variation and has been
254 associated with disease phenotypes and often included as a covariate in statistical models to account for
255 cell type heterogeneity(Merrill et al. 2023; Ryan et al. 2022; Kong et al. 2019). Most studies do not measure
256 actual cell counts and rely on predicted values from DNA methylation-based algorithms. Given that these
257 algorithms use references profiled on previous arrays/versions, we tested whether there were differences in
258 cell type proportions estimated from the two EPIC versions using one of the most commonly used cell type
259 deconvolution methods for whole blood(Salas et al. 2022). We compared the proportions of twelve immune
260 cell types using the *FlowSorted.Blood.EPIC* R package with two commonly used probe selection methods:
261 IDOL and auto. The 1200 pre-selected IDOL probes have comparable coverage on both the EPICv2 (>99%)
262 and the Infinium Methylation Screening Array (MSA) (98%), focusing on trait- and cell identity-associated
263 CpGs(Goldberg et al. 2024) (Table 2 and Supplementary material). On comparing the probes selected for
264 each EPIC version using the auto method, we identified a large overlap of over 90% probes between the
265 two EPIC versions, with only 10% these auto-selected probes overlapping with the pre-selected IDOL
266 probes (Supplementary Fig. S6). When we conducted PCA on the functional normalized data from the
267 combined three in-house cohorts or all four cohorts, we observed that the proportions of memory B cell
268 (Bmem), memory CD4+ T cell (CD4mem), memory CD8+ T cell (CD8mem), and neutrophil (Neu) were
269 significantly associated with PC1 (Supplementary Fig. S5A and S5C). However, when PCA was performed
270 on individual cohort, only CD8mem showed a consistent significant association with PC1 in VHAS,
271 CLHNS, and CALERIE (Supplementary Fig. S5B).

272 We found high Spearman correlations between cell type proportions inferred on EPICv1 and
273 EPICv2, with an average correlation of 0.883 by the IDOL method and 0.890 by the auto method
274 (Supplementary Tables S10-11). Despite their high correlation, we identified significant differences in
275 estimated proportions of five and nine cell types between EPICv1 and EPICv2 using the IDOL and auto
276 methods, respectively, across the three in-house cohorts, with four cell types commonly identified in both
277 methods: basophils (Bas), CD8mem, monocytes (Mono), natural killer cells (NK) (Figure 3 and
278 Supplementary Fig. S7). Consistent findings were observed in BeCOME with the auto method; two
279 additional cell types (CD8mem and NK) showed significant proportional differences between versions
280 when the IDOL method was used (Supplementary Fig. S8A, Supplementary Tables S10-11). Overall,
281 estimated proportions were significantly different by 1-2% between the EPIC versions across multiple cell
282 types, and these findings remained consistent irrespective of the probe selection method, albeit with varying
283 effect sizes depending on cell type and cohort (Figure 3, Supplementary Fig. S7-8, and Supplementary
284 Tables S10-11).

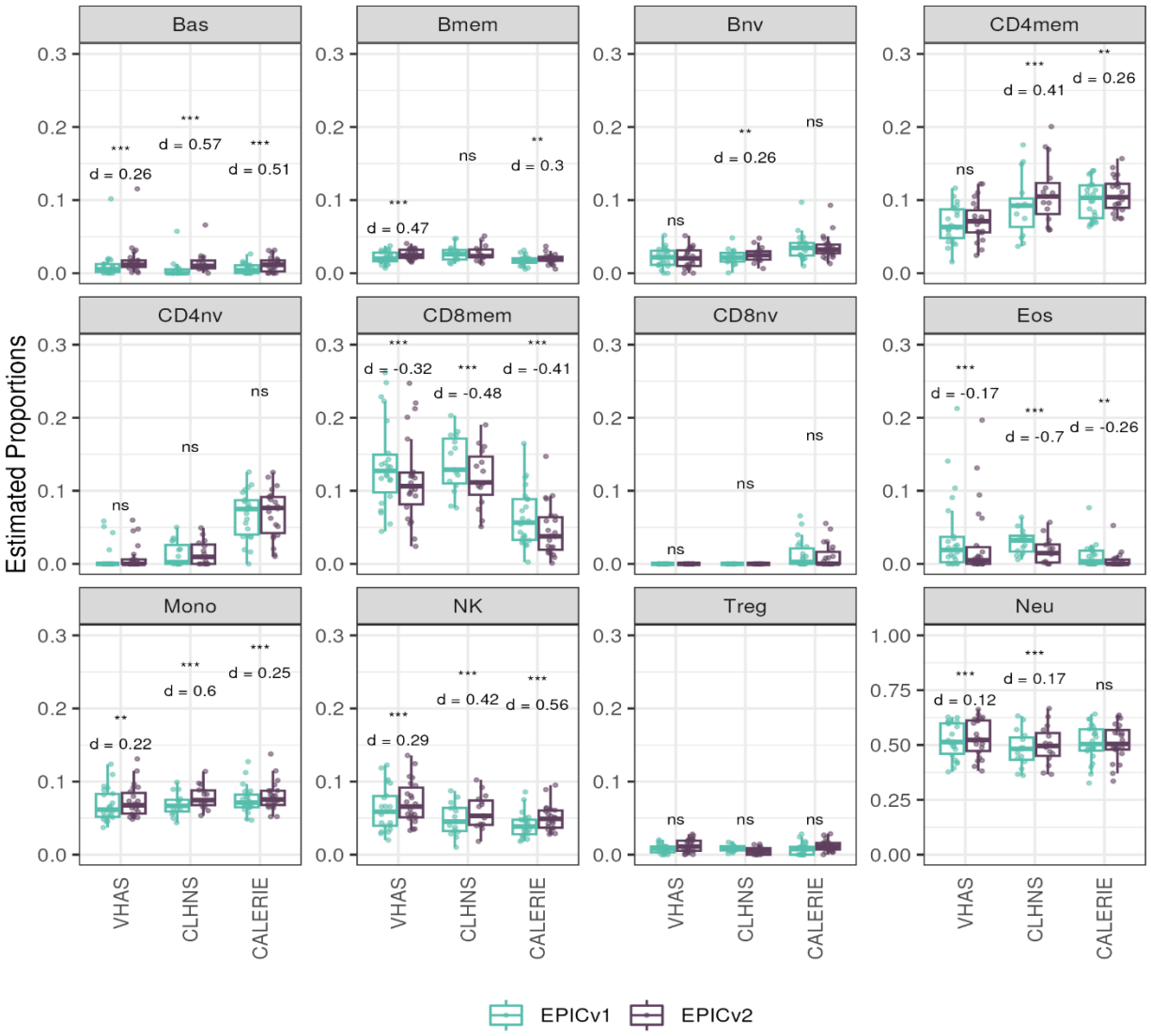


Figure 3. Differences in DNA methylation-based immune cell type proportions estimated using the IDOL reference on matched samples assessed on EPICv1 and EPICv2 in VHAS, CLHNS, and CALERIE. Paired t-tests were performed to compare cell type proportions between EPICv1 and EPICv2, and p-values were derived. Statistical significance was defined as Bonferroni adjusted *p*-value <0.05. ** denotes Bonferroni *p* <0.05, *** denotes Bonferroni *p* <0.001, "ns" denotes "not significant", and "d" denotes effect size measured using Cohen's d. A positive Cohen's d indicates higher estimates in EPICv2 compared to EPICv1.

294 **Table 2.** Summary of predictive CpGs of DNA methylation-based clocks, biomarker predictors, and cell
295 type deconvolution in EPICv1, EPICv2, Infinium Methylation Screening Array (MSA), and 450K.

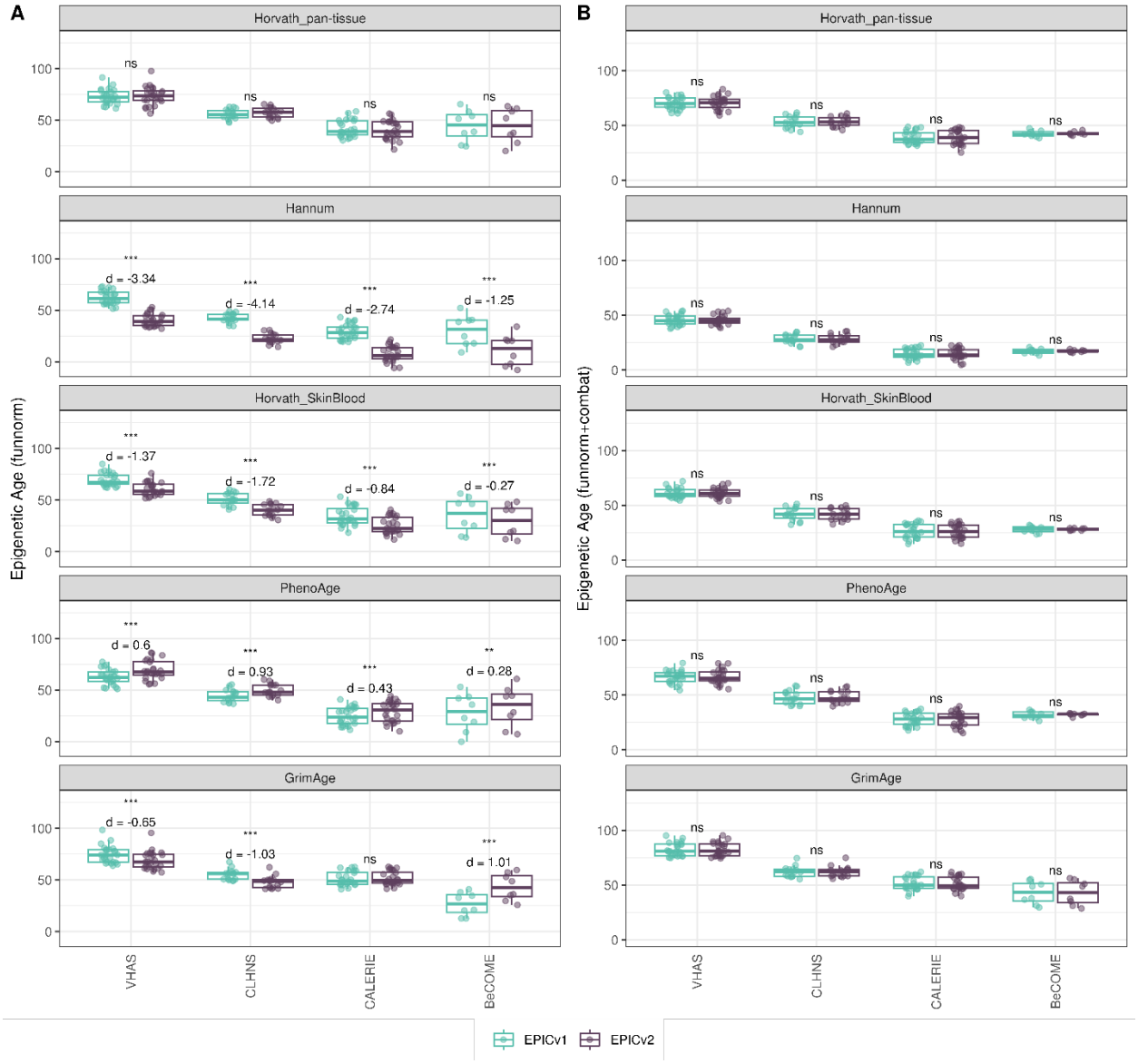
Tools	CpGs	EPICv1 absent CpGs	EPICv2 absent CpGs	EPICv2 replicate probes	MSA absent CpGs	450K absent CpGs	Training arrays (sample size, age in years)
Horvath pan-tissue*	353	19 (5.38%)	13 (3.68%)	4 (1.13%)	6 (1.7%)	0	27K, 450K (n=7844, 0-100)
Hannum*	71	6 (8.45%)	7 (9.86%)	5 (7.04%)	3 (4.23%)	0	450K (n=656, 9-101)
Horvath Skin and Blood	391	0	17 (4.35%)	9 (2.3%)	3 (0.77%)	0	450K, EPICv1 (n=896, 0-94)
PhenoAge	513	0	18 (3.51%)	7 (1.36%)	3 (0.58%)	0	27K, 450K, EPICv1 (n=9926, 18-100)
GrimAge**	1030	NA	NA	NA	NA	NA	450K, EPICv1 (n=1731, mean 66)
PC Clocks***	78464	0	5801 (7.39%)	650 (0.83%)	51108 (65.14%)	0	450K, EPICv1 (n=NA, 0-101)
DunedinPACE	173	0	29 (16.76%)	2 (1.16%)	2 (1.16%)	0	EPICv1 (n=1037, 38 and 45)
DNAmTL	140	0	31 (22.14%)	3 (2.14%)	2 (1.43%)	0	450K, EPICv1 (n=2256, 22-93)
epiTOC*	385	31 (8.05%)	26 (6.75%)	0	13 (3.38%)	0	450K (n=656, 19-101)
IL-6 score	35	0	3 (8.57%)	1 (2.86%)	12 (34.29%)	0	450K, EPICv1 (n=875, 67-78)
CRP score*	1765	104 (5.89%)	96 (5.44%)	41 (2.32%)	357 (20.23%)	0	27K, 450K, EPICv1 (n=22774, 16-75)
Smoking score	233	0	23 (9.87%)	2 (0.86%)	27 (11.59%)	0	EPICv1 (n=5087, 18-99)
Alcohol score	450	1 (0.22%)	49 (10.89%)	6 (1.33%)	47 (10.44%)	0	EPICv1 (n=5087, 18-99)
IDOL	1200	0	8 (0.67%)	48 (4%)	20 (1.67%)	741 (61.75%)	EPICv1 (n=56, 19-58)

296 * There are CpGs commonly absent in both EPICv1 and EPICv2 in these clocks: Horvath pan-tissue: 3
297 CpGs; Hannum: 2 CpGs; epiTOC: 6 CpGs; CRP score: 20 CpGs. There are probes absent in EPICv1 and

298 are reintroduced in EPICv2 in these clocks: Horvath pan-tissue: 14 CpGs; Hannum: 4 CpGs; epiTOC: 25
299 CpGs; CRP score: 83 CpGs.
300 ** Clock CpGs for GrimAge are not publicly available.
301 *** PC Clocks CpGs include the total CpGs required to calculate PC versions of Horvath pan-tissue,
302 Hannum, Horvath Skin and Blood, PhenoAge, and GrimAge.

303 **Epigenetic ages and EAAs were significantly different between EPIC versions depending on analysis**
304 **method**

305 Epigenetic clocks are based on the property that DNA methylation levels at specific CpGs highly
306 correlate with chronological age or age-related outcomes(Horvath 2013; Hannum et al. 2013; Horvath et
307 al. 2018; Levine et al. 2018; Lu et al. 2019b; Belsky et al. 2022). In order to evaluate the concordance
308 between EPICv1 and EPICv2 in the context of these DNA methylation-based tools, we compared estimates
309 from seven widely used first- and second- generation epigenetic clocks in VHAS, CLHNS, and CALERIE
310 (see Methods). Across these seven epigenetic clocks, ~77-96% of predictive CpGs were retained on EPICv2
311 (Table 2 and Supplementary material). The first and second generation clocks have comparable probe
312 coverage on both EPICv2 and the MSA, while the PC clocks have a much reduced probe coverage (35%)
313 on the MSA. We identified high Pearson correlations of 0.807-0.996 between chronological age and
314 epigenetic age obtained from both EPICv1 and EPICv2 (Supplementary Table S12). EPICv2 epigenetic
315 ages for all clocks were moderately to highly correlated with EPICv1 epigenetic ages (0.583-0.996), with
316 differences in epigenetic ages between technical replicates being 0.151-4.206 (Supplementary Table S13),
317 although we noted significant differences in epigenetic ages between EPICv2 and EPICv1 in Hannum,
318 Horvath skin and blood, and PhenoAge, with these differences ranging in effect size from 0.267-4.137
319 (Figure 4A, Supplementary Table S13). Consistent with these results, in the external validation cohort
320 BeCOME, EPICv1 and EPICv2 epigenetic ages were highly correlated with chronological age and yet there
321 were significant epigenetic age differences between versions in all epigenetic clocks except again Horvath
322 pan-tissue (Supplementary Tables S12-13, Supplementary Fig. S9A).


A
B
Horvath_pan-tissue
Hannum
Horvath_SkinBlood
PhenoAge
GrimAge
Epigenetic Age (funnorm)
Epigenetic Age (funnorm+combat)
ns
d = -3.34
d = -4.14
d = -2.74
d = -1.25
d = -1.37
d = -1.72
d = -0.84
d = -0.27
d = 0.6
d = 0.93
d = 0.43
d = 0.28
d = -0.65
d = -1.03
d = 1.01
VHAS
CLHNS
CALERIE
BeCOME
EPICv1
EPICv2

323

324 **Figure 4.** Differences in epigenetic ages between EPICv1 and EPICv2 using the Horvath pan-
325 tissue, Hannum, Horvath skin and blood, PhenoAge, and GrimAge clocks in VHAS, CLHNS, and
326 CALERIE when using **(A)** functional normalization and **(B)** functional normalization with batch-
327 correction for EPIC version, chip and row. Paired t-tests were performed to compare estimates
328 between EPICv1 and EPICv2, and *p*-values were derived. Statistical significance was defined as
329 Bonferroni adjusted p-value <0.05. ** denotes Bonferroni p <0.05, *** denotes Bonferroni
330 p<0.001, "ns" denotes "not significant", and "d" denotes effect size measured using Cohen's d. A
331 positive Cohen's d indicates estimates in EPICv2 compared to EPICv1.

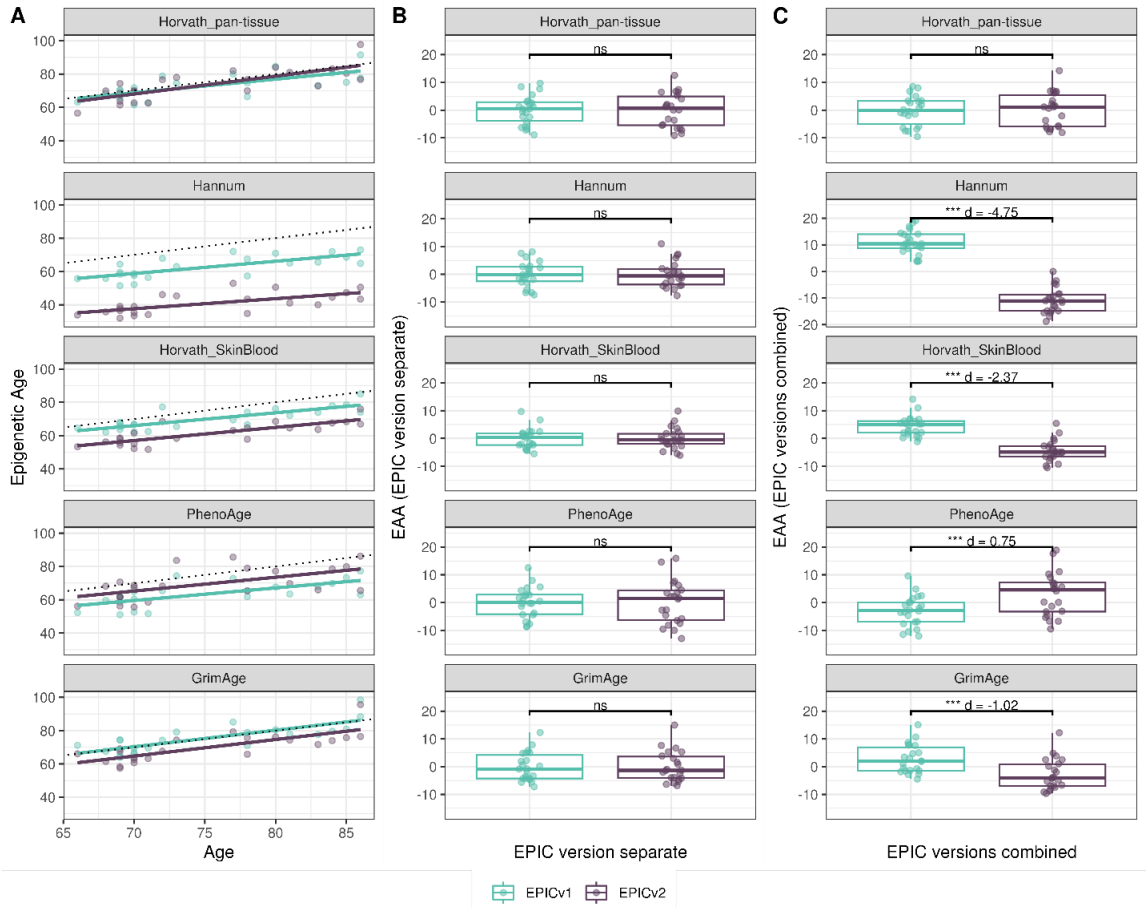
332

333

334 We also estimated epigenetic age acceleration (EAA), a measure of rate of aging which has been
335 associated with health outcomes, by considering samples profiled on each EPIC version separately and in
336 a combined manner (see Methods). Irrespective of how matched samples on the two EPIC versions were
337 considered for EAA calculation, we noted modest to high correlation (0.574-0.960) of EAA between the
338 EPIC versions. When EAA was calculated separately by versions, we noted no significant differences
339 between EPICv1 and EPICv2 (Supplementary Table S12, Figure 5B, Supplementary Fig. S10B and S11B).
340 When clock estimates from the EPIC versions were combined prior to EAA calculation, there were
341 significant differences between EPICv1 and EPICv2 for all clock EAAs except Horvath pan-tissue
342 (Supplementary Table S14, Figure 5C, Supplementary Fig. S10C and S11C). To next test whether the
343 different EPIC versions contribute to observed EAA differences, we combined epigenetic ages in both EPIC
344 versions and then calculated EAAs by including version as a covariate in the linear regression. In doing so,
345 we noted that there were no significant EAA differences between the two EPIC versions, akin to when EAA
346 was calculated separately by version (Supplementary Table S14). We repeated the same analyses using
347 epigenetic clocks estimated based on principal components (PC clocks) rather than individual predictive
348 CpGs, as they have been shown to overcome unreliability in clock estimates due to technical noise(Higgins-

349 Chen et al. 2022). Once again, we noted that there were no significant EAA differences when EAA was
350 calculated for each EPIC version separately. When EAAs were calculated on combined sets of matched
351 samples, we noted significant differences in PCHorvath skin and blood and PCPhenoAge, and these
352 differences were corrected with EPIC version adjustment (Supplementary Fig. S12-14). Consistent with
353 other cohorts, in BeCOME, we noted similar findings in epigenetic ages, EAAs, and PC clocks estimates
354 (Supplementary Table S14, Supplementary Fig. S9 and S14).

355 On next evaluating the DunedinPACE, DNAmTL, and epiTOC clocks, we observed a high
356 correlation (0.699-0.983) of clock estimates between EPICv1 and EPICv2 (Supplementary Table S15).
357 Unlike the epigenetic clocks above, when we examined the EPIC versions separately, we noted that
358 DNAmTL estimates were significantly different between versions in all four cohorts, while DunedinPACE
359 and epiTOC estimates were significantly different in three out of the four cohorts (Supplementary Table
360 S15, Figure 6, and Supplementary Fig. S9-11).
361



362

363 **Figure 5**. Epigenetic ages on matched samples assessed on EPICv1 and EPICv2 in VHAS. **(A)** Scatter plot
364 of Horvath pan-tissue, Hannum, Horvath skin and blood, PhenoAge, and GrimAge clock ages (Y axis) and
365 chronological age (X axis) with dotted line indicating x=y, coloured by EPIC version. **(B-C)** Boxplots
366 comparing EPICv1 and EPICv2 EAAs calculated by considering (B) EPIC versions separately and (C)
367 combined. Paired t-tests were performed to compare estimates between EPICv1 and EPICv2, and p-values
368 were derived. Statistical significance was defined as Bonferroni adjusted *p*-value <0.05. ** denotes
369 Bonferroni *p* <0.05, *** denotes Bonferroni *p* <0.001, "ns" denotes "not significant", and "d" denotes effect
370 size measured using Cohen's d. A positive Cohen's d indicates higher estimates in EPICv2 compared to
371 EPICv1.

372

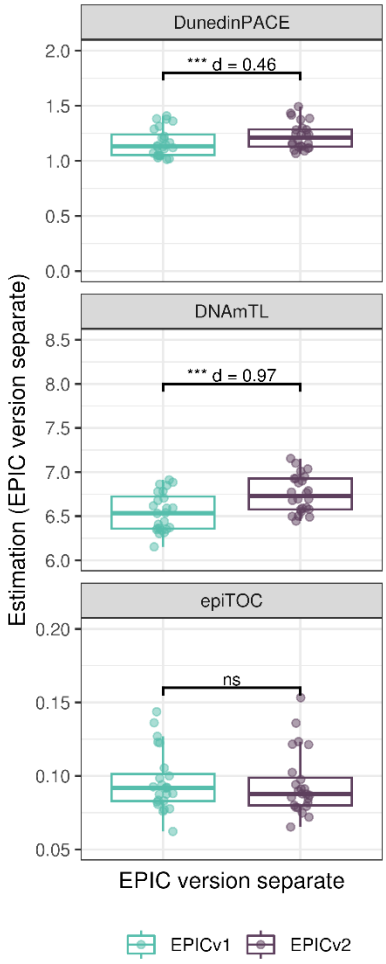


Figure 6. Rate-based and other clock estimates on matched samples assessed on EPICv1 and EPICv2 in VHAS. Boxplots comparing DunedinPACE, DNAmTL and epiTOC estimates calculated by considering EPIC versions separately between EPICv1 and EPICv2. Paired t-tests were performed to compare estimates between EPICv1 and EPICv2, and p-values were derived. Statistical significance was defined as Bonferroni adjusted *p*-value <0.05. ** denotes Bonferroni *p* <0.05, *** denotes Bonferroni *p* <0.001, "ns" denotes "not significant", and "d" denotes effect size measured using Cohen's d. A positive Cohen's d indicates estimates in EPICv2 compared to EPICv1.

Estimated CRP and smoking scores were significantly different between EPIC versions

DNA methylation-based estimates of IL-6, CRP, smoking, and alcohol use are widely used as reliable proxies of actual measurements in epigenetic studies(Engelbrecht et al. 2022). Approximately 90%

384 predictive CpGs were retained on EPICv2, while only 66-90% of these probes were retained on the MSA
385 (Table 2 and Supplementary material). Estimates of these inflammation and lifestyle biomarkers were
386 modestly to highly correlated (0.788-0.993) between EPICv1 and EPICv2 (Supplementary Table S16).
387 When estimates were calculated separately by versions (EPIC versions separate), we identified significantly
388 different IL-6, CRP, alcohol, and smoking scores between EPICv1 and EPICv2 in at least one of the four
389 cohorts (Supplementary Fig. S16). When estimates from the EPIC versions were first combined and then
390 adjusted for version, as expected, we noted no significant differences between EPICv1 and EPICv2 in any
391 of these four predictors (Supplementary Table S16).

392 **EPIC version differences remained significant irrespective of the choice of normalization method,**
393 **while upstream batch correction eliminated such differences**

394 To investigate the influence of data preprocessing methods, we repeated the clustering and PCA
395 analyses using raw and functional normalized data with batch correction by ComBat. When using raw data,
396 we noted a pattern similar as previously observed using functional normalized data (Fig. 2B and
397 Supplementary Fig. S17A). Expectedly, when functional normalized data was adjusted for version, chip,
398 and position on chip (row), we observed that matched samples clustered by individual and not by EPIC
399 version (Supplementary Fig. S17B). When examining the contribution of EPIC version to PC1 (~98% of
400 the total DNA methylation variation) across different stages of data processing—from raw to functional
401 normalized, and then version-corrected—we observed a reduction in the relative contribution of EPIC
402 version while other biological variables had increased contribution. When batch correction was applied on
403 functional normalized data, EPIC version contribution was reduced to zero with no significant association
404 to PC1 (Supplementary Fig. S5).

405 Given the effect of data normalization on epigenetic clock estimations(McEwen et al. 2018; Ori et
406 al. 2022), we examined whether normalization methods affect the observed differences between EPIC
407 versions, obtained after applying functional normalization (funnorm). To test this, we combined the shared
408 probes on the two EPIC versions and applied Beta-MIxture Quantile (BMIQ) normalization, and

409 subsequently calculated epigenetic ages, EAAs, and inflammation and lifestyle biomarker estimates.
410 Overall, the differences in estimates of these DNA methylation-based tools between EPIC versions when
411 BMIQ normalization was applied were consistent with those obtained in functional normalized data,
412 specifically in Horvath skin and blood and PhenoAge which were consistent across the three in-house
413 cohorts (Supplementary Fig. S18-23).

414 Next, we also examined if applying EPIC version correction using batch correction algorithms can
415 circumvent version-specific effects and eliminate the observed differences in DNA methylation-based
416 estimates between EPICv1 and EPICv2, given the observed contribution of EPIC version in explaining
417 DNA methylation variation. We tested this by combining common probes on EPICv1 and EPICv2 after
418 functional normalization or BMIQ normalization, and sequentially adjusted for EPIC version, chip, and
419 row effects as applicable using the ComBat function implemented in the *sva* R package(Leek et al. 2012).
420 As expected, when epigenetic clocks, inflammation and lifestyle biomarkers were calculated using version-
421 corrected input β values, there were no significant differences in estimates between the EPIC versions
422 (Figure 4, Supplementary Fig. S18, S24-25).

423
424

425 **Discussion**

426 Various generations of DNA methylation arrays, largely manufactured by Illumina, have had
427 tremendous impact on the field of epigenetics, enabling large population studies of human health and
428 disease. The most recently released Illumina EPICv2 microarray provides wider coverage of the DNA
429 methylome compared to its predecessor EPICv1, yet at the same time eliminates approximately one-fifth
430 of the previous probes. This then might create challenges for the utility and transferability of the myriad
431 bioinformatic tools developed for the EPICv1, especially if some of the missing probes were included in a
432 particular tool. Here, using matched venous blood samples from 67 adults across four geographically
433 diverse populations, we comprehensively examined the concordance between EPIC versions in a wide

434 range of commonly used DNA methylation-based tools. Overall, our study identified small but significant
435 differences between the two EPIC versions, and provided some insights into possible remediation
436 approaches. First, we observed overall high array level concordance but variable probe level concordance,
437 with a subset of probes displaying poor agreement between versions despite high inter-sample variability.
438 Second, EPIC version differences were identified across all cohorts in DNA methylation-based tools,
439 including epigenetic clocks, inflammation and lifestyle biomarkers, and cell type predictors. Third,
440 discordance between EPIC versions was dampened by either accounting for EPIC version in statistical
441 models as a covariate or using batch correction tools such as ComBat. Fourth, irrespective of the choice of
442 normalization method EPIC version differences still persisted. Overall, our work emphasizes careful
443 consideration of research settings when i) samples across both EPIC versions are harmonized, and ii)
444 findings derived from one EPIC version are compared to those derived from the other version.

445 Initial comparisons of EPIC versions that were confined to immortalized cell lines and a limited
446 number of human primary samples reported generally good array level correlations between the two EPIC
447 versions(Kaur et al. 2023; Noguera-Castells et al. 2023; Peters et al. 2024). However, these studies were
448 not designed to investigate concordance at the probe level, and did not evaluate the performance of DNA
449 methylation-based tools such as epigenetic clocks, cell types, and biomarker predictors, particularly across
450 geographically diverse settings, which are highly relevant for human population epigenetic studies. Using
451 a population-based framework, we performed unsupervised clustering on sample-to-sample Pearson
452 correlations across all cohorts, focusing on the shared probes between EPICv1 and EPICv2 to capture linear
453 associations. Our analyses revealed both technical and biological influences on DNA methylation profiles
454 across both individual and combined cohort analyses. As expected, sex emerged as a primary biological
455 driver of sample clustering, followed by a significant technical effect of EPIC version, with samples
456 clustering according to the version on which they were profiled. As anticipated, samples from the external
457 validation cohort BeCOME clustered closely together and showed relative separation from the other
458 cohorts, reflecting differences in sample handling and processing inherent to different facilities. To assess
459 systematic, technical and potentially nonlinear effects (Bishara and Hittner 2015), we also performed

460 clustering on sample-to-sample Spearman rank correlations. In contrast to the Pearson correlation-based
461 findings, Spearman correlation-based clustering revealed clear sample separation first by EPIC version
462 across both individual and combined cohorts, regardless of cohort-specific characteristics.

463 Biological variables, such as sample relatedness and cell type composition, are well-established
464 key drivers of DNA methylation variation (Gaunt et al. 2016; Hawe et al. 2022; Bergstedt et al. 2022),
465 typically exhibiting a larger effect than batch variables. Our PCA, which combined all four cohorts to assess
466 the linear relationship between DNA methylation and each variable, expectedly showed a significant
467 contribution of sample relatedness and estimated cell type proportions on PC1, with a smaller yet significant
468 contribution of EPIC version. Given our study design, which involved matched samples obtained from the
469 same DNA aliquot, it was expected that sample relatedness and cell type proportions would have a larger
470 contribution than EPIC version. However, even in this controlled setting, EPIC version was still a notable
471 contributor to PC1, though to a smaller percentage than both investigated biological variables. When the
472 same analysis was performed on an individual cohort-basis, we noted that both EPIC version and donor
473 were associated with PC1, albeit with varying percentage contributions depending on the cohort, suggesting
474 that sample sizes are likely a determinant of the relative contribution of each variable.

475 On examining whether the observed unsupervised clustering by EPIC version was reflective of
476 variable probe-level concordance between the two versions, we noted that a substantial proportion of shared
477 probes (over 50%) were poorly-concordant across all four investigated cohorts. While a small proportion
478 of these poorly-concordant probes, skewed toward lower inter-sample variability, could be attributed to
479 their non-specific probe hybridization to the genome, underlying genetic variation, or technical differences
480 such as altered EPICv2 probe design(Kaur et al. 2023; McCartney et al. 2016; Peters et al. 2024; Pidsley et
481 al. 2016; Price et al. 2013), the majority remain unexplained by these factors. Affirming our observation
482 that a large proportion of shared probes exhibit low-concordance, previous studies in 450K and EPICv1
483 have reported similar low correlations and intraclass correlation/reliability for the majority of the shared
484 probes(Logue et al. 2017; Sugden et al. 2020; Olstad et al. 2022). Similarly, studies evaluating probe
485 reliability within the same platform also reported a modest mean correlation of 0.381 and 0.361 across all

486 probes in EPICv1-EPICv1 and 450K-450K respectively(Zhang et al. 2024; Bose et al. 2014), suggesting
487 that a substantial proportion of probes even within the same platform/version exhibit low correlations. More
488 recent studies comparing EPICv1 and EPICv2 have only estimated array-level concordance using a single
489 sample from cell lines(Kaur et al. 2023; Noguera-Castells et al. 2023; Chen and Zhou 2024) or probe-level
490 concordance estimated on sample size comparable to our study(Lussier et al. 2024). While these studies
491 were either cell line-based or limited to a single pediatric cohort, our analysis includes a wide range of
492 sample sizes, age groups, and diverse geographical regions, providing a more comprehensive and
493 representative evaluation of EPIC version differences. This approach better reflects the scale and
494 complexity of human studies and offers valuable insights into the impact of EPIC version variation in
495 human epigenetics studies.

496 We observed significant systematic differences in measures derived from DNA methylation-based
497 tools. In the context of cell type prediction algorithms, there were significant differences in proportions of
498 various cell types, when using either the IDOL or auto prediction methods. These differences may be
499 reflective of version incompatibility between the reference library used, built on previous Illumina
500 microarrays, and EPICv2(Salas et al. 2022). While proportional differences between EPIC versions were
501 minimal, within the RMSE between observed and predicted proportions of the cell-type deconvolution
502 algorithm(Salas et al. 2022), these differences were systematic as the change in direction was consistent
503 across the investigated cohorts. Further, DNA methylation studies often report molecular effects as small
504 as 2%(Breton et al. 2017), making these minimal differences pertinent to the interpretation of findings.
505 Specifically, in studies where cell types are the primary variable of interest, even such small differences in
506 predicted cell type proportions - confounded by the EPIC versions - can make it difficult to separate true
507 biological signals from version differences, potentially leading to spurious findings. Similarly, in the
508 epigenetic clock and biomarker predictors, we observed discordance in estimates obtained between EPICv1
509 and EPICv2 as previously noted(Lussier et al. 2024; Tay et al. 2025), though these observations were
510 generally consistent across cohorts and tools with some nuanced differences. It has been shown that this
511 discordance between versions may be due to the varying number of predictive CpGs absent on EPICv2

512 (Garma and Quintela-Fandino 2024; Pourcq et al. 2024). Additionally, this discordance between versions
513 may be due to the weights of these missing predictive CpGs and the influence of subsequent imputation
514 performed by the clock algorithms to account for this missingness. However, our analyses using BMIQ
515 normalization which only used shared probes between the versions and thus had the same absent predictive
516 CpGs in both versions, still showed differences in estimates and EAAs. Similarly, when we applied a new
517 clock trained on the common probeset of 450K, EPICv1 and EPICv2 (Garma and Quintela-Fandino 2024),
518 we still noted significant differences in clock estimates between versions (Supplementary Material,
519 Supplementary Table S17, Supplementary Fig. S26), supporting the notion that absent CpGs are not the
520 only contributor to the observed version differences.

521 Although epigenetic clocks are often analyzed as direct comparisons of epigenetic ages as described
522 above, it has been noted that a regression-based measure of epigenetic age, EAA, is more appropriate in
523 most cases as it is robust to data preprocessing(Horvath 2013; McEwen et al. 2018; Engelbrecht et al. 2022;
524 Ori et al. 2022). On calculating EAAs by taking the common approach of including all samples in a cohort
525 irrespective of version, we noted that EAA estimates were different between versions, suggesting that this
526 method of calculating EAA is still sensitive to inherent variation between EPICv1 and EPICv2. Using an
527 alternate approach, when we calculated EAA estimates separately for samples profiled on each EPIC
528 version, we found that these differences were no longer significant. This indicates that EAAs may be a
529 suitable measure of epigenetic aging when version is taken into consideration while calculating EAA.
530 Similarly, we noted consistent differences between versions in inflammation and lifestyle biomarker
531 predictor scores, emphasizing again that it is important to account for systematic version differences when
532 calculating DNA methylation-based measures.

533 Including EPIC version as a covariate in statistical models or employing version correction using
534 ComBat may be options to handle version-specific discrepancies, however, these approaches are not
535 appropriate in cases when EPIC version is fully or partially confounded with biological variables. Our
536 matched and balanced study design allowed us to correct for version using ComBat, which expectedly
537 dampened version differences. In cases of unbalanced study designs, any adjustment for EPIC version can

538 incorrectly remove important biological variation as well as introduce false biological signals(Price and
539 Robinson 2018; Buhule et al. 2014; Nygaard et al. 2016; Zindler et al. 2020). This ideal - and perhaps even
540 unrealistic - scenario starkly contrasts with a typical meta-analysis, where a researcher might wish to
541 combine results from multiple distinct sample groups measured on different EPIC versions. Furthermore,
542 another research setting that presents a similar challenge is a longitudinal study design, where the aim is to
543 explore DNA methylation changes over time. In such cases, when samples from each timepoint are
544 measured on different EPIC versions, it inevitably results in a confound between the variable of interest,
545 i.e., timepoint and EPIC version, thereby not enabling any version correction. If samples from the previous
546 timepoint are still available, one feasible approach is to include a small number of these samples profiled
547 on the previous EPIC version alongside samples from the new timepoint to be profiled on EPICv2 such that
548 any between version differences can be monitored. It is also known that DNA methylation-based tools are
549 sensitive to normalization methods (McEwen et al. 2018; Ori et al. 2022; Engelbrecht et al. 2022), and it is
550 important to explore the contribution of such approaches to observed EPIC version differences noted in the
551 present study. By employing two distinct normalization methods, namely funnorm and BMIQ, we found
552 similar differences in DNA methylation-based estimates based on version, irrespective of normalization
553 method used.

554 On the strength of our matched sample design utilizing four different human cohorts, our study
555 improves the current understanding of the applicability of DNA methylation-based tools for EPICv1 and
556 EPICv2, however, there are several limitations. Our analyses may be limited by relatively modest sample
557 sizes within each cohort, despite this, our study includes matched samples collected from adults with a wide
558 age range and across diverse geographical regions. While the time between EPICv1 and EPICv2
559 measurement varied across the cohorts, it offers a more realistic representation of research settings where
560 samples are collected and quantified across multiple batches on different EPIC versions. Next, given that
561 the training datasets of the investigated epigenetic clocks were predominantly composed of European
562 populations, the suitability of investigated clocks may not be well established in our investigated Asian
563 cohorts, VHAS and CLHNS respectively (Lin 2023; Horvath et al. 2016; Cronjé et al. 2021). However,

564 differences in these cohorts were similar to CALERIE and BeCOME, which are comprised primarily of
565 individuals of European descent, indicating that our findings were not limited to specific populations.
566 Further, we compared EPICv2 cell type proportions and biomarker scores to EPICv1, although we
567 recognize that actual cell counts and biomarker measurements would be more appropriate as the ground
568 truth and would be useful in validating our DNA methylation-based estimates. In spite of this, the primary
569 aim of the current work is to investigate the concordance in estimates between versions, and not to evaluate
570 the accuracy of biomarker prediction. Finally, although our study focused on venous blood, a commonly
571 used tissue in epigenetics research, we speculate that these EPIC version differences will hold true in other
572 tissue types as well. While multiple tissue types were not collected in all of these cohorts, we had access to
573 matched capillary and venous blood samples in VHAS, which allowed us to at least test and confirm the
574 consistency of our findings in another blood preparation.

575 With the rapid generation of DNA methylation data profiled on the newer iterations of the Illumina
576 microarrays, integrating samples across these platforms poses a challenge, owing to discrepancies in probe
577 content among arrays/versions. Our findings highlight differences in the new EPICv2 compared to EPICv1,
578 demonstrate the influence of EPIC version on the majority of commonly used DNA methylation-based
579 tools, and provide possible remediation approaches to account for these EPIC version discrepancies. We
580 therefore encourage careful consideration when harmonizing and interpreting DNA methylation data across
581 multiple arrays/versions to ensure reliability and reproducibility in epigenetics analyses.

582 **Methods**

583 **Description of cohorts**

584 To compare the performance of the two most recent generations of MethylationEPIC BeadChip Infinium
585 microarrays in the context of DNA methylation-based clocks, biomarkers and cell type proportion
586 estimates, we measured the DNA methylomes of a subset of venous whole blood samples on both EPICv1
587 and EPICv2 selected from the (i) Vietnam Health and Aging Study (VHAS)(Korinek et al. 2019), (ii) Cebu

588 Longitudinal Health and Nutrition Survey (CLHNS)(Adair et al. 2011), and (iii) Comprehensive
589 Assessment of Long-term Effects of Reducing Intake of Energy (CALERIE)(Rochon et al. 2011) cohorts.
590 The VHAS cohort additionally includes matched capillary blood samples, randomized using the same array
591 design on both EPIC versions, such that the concordance between EPICv1 and EPICv2 can be assessed in
592 capillary blood as well (n = 24 × 2 blood collection methods × 2 versions). Samples from VHAS, CLHNS
593 and CALERIE were processed in the same facility (Kobor Lab, University of British Columbia (UBC) and
594 BC Children's Hospital Research Institute, Vancouver, Canada). This study was approved by UBC
595 Research Ethics Boards (H18-03136) and the University of Utah Institutional Review Board (IRB
596 00098861). To compare our findings from the three cohorts processed in-house to those from an external
597 facility, we used an independent validation cohort, BeCOME. In this dataset, DNA methylation was
598 measured on EPICv1 and EPICv2 using matched samples at the Max Planck Institute of Psychiatry in
599 Munich, Germany (Table 1). Matched blood samples on EPICv1 and EPICv2 were derived from the DNA
600 extracted from the same aliquot collected at a single timepoint, reducing the likelihood that immune cell
601 fractions differ significantly between the matched samples, or that they contribute to the observed EPIC
602 version differences. The time between EPICv1 and EPICv2 array quantification was approximately one
603 month and two years for VHAS and CLHNS respectively. EPICv1 and EPICv2 array quantification was
604 carried out at the same time for CALERIE and BeCOME. Given that the demographic characteristics
605 across the cohorts are different (Table 1), we performed all analyses independently on the cohorts and
606 reported the findings in a cohort-specific manner, excluding the unsupervised clustering analyses as
607 described below.

608 **DNA methylation profiling, sample and probe quality control**

609 Using similar protocols for all three cohorts, DNA was extracted from samples, bisulfite converted
610 using EZ-96 DNA Methylation kits (Zymo Research, Irvine, CA), hybridized to the MethylationEPIC
611 BeadChip Infinium microarray v1.0 B5 (EPICv1) and Infinium MethylationEPIC v2.0 (EPICv2) arrays,
612 and scanned with the Illumina iScan 2000 to obtain IDAT files that capture raw DNA methylation

613 intensities. IDATs were read using *minfi* R package to obtain β values that represent DNA methylation
614 intensities for each CpG site ranging from 0 (fully unmethylated) to 1 (fully methylated). Technical
615 replicates derived from the same sample after bisulfite conversation were quantified to monitor technical
616 variation within each EPIC version, independently for each cohort. Both VHAS and CALERIE included
617 two technical replicates each on EPICv1 and EPICv2, and CLHNS included one technical replicate on
618 EPICv2 and no technical replicate on EPICv1. The external validation cohort BeCOME did not include any
619 technical replicates on either EPIC version. Sample quality control checks were performed as described in
620 previous publications(Merrill et al. 2023; Konwar et al. 2021). Blood samples collected in the three cohorts
621 and the external validation cohort passed all 17 Illumina quality control metrics in the *ewastools* R
622 package(Murat et al. 2020; Illumina Inc. 2015), and detection *p*-value, beadcount, average methylated and
623 unmethylated intensity metrics in the *minfi* R package (Supplementary Table S18). We also performed
624 sample identity checks with unsupervised hierarchical clustering using complete linkage with Euclidean
625 distance on sample-to-sample Spearman correlations, calculated using the 57 single nucleotide
626 polymorphism (SNP) probes that are common to both EPIC versions to confirm whether matched samples
627 cluster together.

628 In order to identify EPICv1 and EPICv2 probes of poor quality, we performed quality control
629 checks using the detectionP and beadcount functions in the *minfi* and *wateRmelon* R packages, respectively.
630 Probes with detection p-value > 0.01 or beadcount < 3 in greater than 1% of the samples were flagged
631 (VHAS- EPICv1: 59,233, EPICv2: 46,735, common to both versions: 4,826; CLHNS-EPICv1: 23,793,
632 EPICv2: 17,587, common to both versions: 991; CALERIE-EPICv1: 29,608, EPICv2: 24,872, common to
633 both versions: 1,292; BeCOME-EPICv1: 12,511, EPICv2: 20411, common to both versions: 2,714), but all
634 probes in EPICv1 and EPICv2 were retained for subsequent analyses.

635 **Unsupervised clustering analyses on DNA methylation data using a two-step approach**

636 To perform unsupervised hierarchical clustering on samples, we employed a two-step approach, as
637 applied in previous studies comparing array platforms and other -omic studies (Cheung et al. 2020;

638 Levenstien et al. 2003; Koch et al. 2018). Specifically, we first calculated sample-to-sample (array level)
639 Pearson or Spearman correlations using the 721,378 probes shared between EPICv1 and EPICv2 in pair-
640 wise manner for all the samples in the four cohorts, then we clustered the samples by unsupervised
641 hierarchical clustering using complete linkage with Euclidean distance on sample-to-sample correlations,
642 using the hclust function implemented in the *stats* R package (R Core Team 2022) and visualized by
643 pheatmap in the *pheatmap* R package(Kolde 2019). We first performed clustering on each of the four
644 cohorts separately, then by combining the three cohorts processed in the in-house facility (VHAS, CLHNS
645 and CALERIE), and finally by combining all four cohorts regardless of processing facility. Grouping
646 cohorts once again in a similar manner, we performed principal component analyses (PCA) using the shared
647 probes between the EPIC versions. We tested the association between each of the top five PC loadings and
648 EPIC version, cohort, and sex by employing one-way analysis of variance (ANOVA) or t-tests and applying
649 a Bonferroni multiple test correction. The percentage of DNA methylation variation explained by each
650 variable (EPIC version, cohort, and sex) was calculated as the ANOVA or t-test R-squared value. It should
651 be noted that while performing the PCA on each cohort independently, we did not test the association
652 between PC loadings and sex in the CLHNS cohort since this cohort comprises of only females; for the
653 other three cohorts, associations with EPIC version, cohort, and sex were tested.

654 **DNA methylation data preprocessing and replicate probe analyses**

655 To account for color and probe-type bias, we performed functional normalization (funnorm) with
656 background correction and dye-bias normalization (noob) in the *minfi* R package(Fortin et al. 2014; Triche
657 et al. 2013) independently on EPICv1 and EPICv2 samples for each cohort. Cohort-specific technical
658 replicate sample correlations were used to monitor preprocessing (technical replicates were not available in
659 the BeCOME external validation cohort). Between technical replicate samples on the same EPIC version,
660 improved Spearman correlations of whole array β values and reduced root mean square error (RMSE) were
661 observed as processing progressed from raw (VHAS: Spearman *rho* = 0.9832, RMSE = 0.0306; CLHNS:
662 Spearman *rho* = 0.9890, RMSE = 0.0235; CALERIE: Spearman *rho* = 0.9859, RMSE = 0.0349) to funnorm

normalized data (VHAS: Spearman *rho* = 0.9851, RMSE=0.0288; CLHNS: Spearman *rho* = 0.9917, RMSE = 0.0231, CALERIE: Spearman *rho* = 0.9887, RMSE = 0.0223). Array level reliability of within- and between-EPIC version technical replicates were assessed by intra-class correlation coefficients (ICC) with the two-way random-effects model as previously described(Sugden et al. 2020; Koo and Li 2016). Specifically, array level ICC was calculated using the technical replicates as the repeated measures ("raters") and each probe represented a "target" or "subject". Within-EPIC version ICCs were calculated based on the technical replicates within the same EPIC version, while between-EPIC version ICC were based on the matched technical replicates on EPICv1 and EPICv2.

To calculate estimates of DNA methylation-based tools, noob corrected data was used as input for cell type deconvolution, and funnorm normalized data was used as input for epigenetic clocks and biomarkers. In addition, to test whether normalization methods influence DNA methylation-based tools, we compared estimates calculated by combining the common probes on EPICv1 and EPICv2 after noob correction in a cohort-wise manner for VHAS, CLHNS, CALERIE, and BeCOME, and subsequently applying Beta-MIxture Quantile (BMIQ) normalization implemented in the *wateRmelon* R package(Teschendorff et al. 2013). Due to their type I and type II design switch between EPICv1 and EPICv2(Kaur et al. 2023), 82 probes were removed prior to BMIQ normalization. To account for any systematic bias in DNA methylation measurements(Leek et al. 2012) and subsequently test whether there are differences in DNA methylation-based estimates between EPIC versions, we applied batch correction for EPIC version, chip and row on funnorm normalized data using the ComBat function implemented in the *sva* R package(Leek et al. 2012). We applied Pearson correlation to evaluate linear relationships to age in the epigenetic clock analyses, while in all analyses we applied Spearman correlation. Absolute β value differences of within EPIC version technical replicates on each of the common probes were used to determine technical noise of β value per probe. Absolute differences of within EPIC version technical replicates in DNA methylation-based tool estimates (cell type deconvolution algorithms, epigenetic clocks, and inflammation and lifestyle biomarkers) were used to indicate within EPIC version technical error.

688 Technical replicates were removed prior to calculating correlations and preforming statistical tests
689 comparing EPICv1 and EPICv2.

690 Given that there are certain probes on EPICv2 having two or more replicates (replicate probes), we
691 characterized their distribution across the genome, and compared three strategies to collapse them into a
692 single β value (based on detection *p*-value, mean and median). Our analyses identified that collapsed β
693 values of EPICv2 replicate probes obtained using all three methods were highly correlated to corresponding
694 EPICv1 probes, therefore EPICv2 replicate probes with lowest detection *p*-value were chosen as the
695 representative probe based on previous recommendation(Kaur et al. 2023) (Supplementary Material), and
696 this approach was used for all the reported subsequent analyses.

697 **Estimation of immune cell type proportions using DNA methylation-based cell type deconvolution**

698 Cellular composition in heterogeneous tissue such as whole blood is one of the key contributors to
699 the variation in DNA methylation profiles of bulk tissue(Jones et al. 2017; Zheng et al. 2017). In the absence
700 of complete cell count data for the study samples, we estimated proportions of twelve immune cell types,
701 basophils (Bas), naïve and memory B cells (Bnv, Bmem), naïve and memory CD4+ T cells (CD4nv,
702 CD4mem), naïve and memory CD8+ T cells (CD8nv, CD8mem), eosinophils (Eos), monocytes (Mono),
703 neutrophils (Neu), natural killer (NK), and T regulatory cells (Treg) from matched venous blood samples
704 measured on the two EPIC versions. We used two methods of probe selection to estimate these cell type
705 proportions: (i) the extended Identifying Optimal DNA methylation Libraries reference (IDOL), with
706 probes not represented on EPICv2 removed from the reference before cell type proportions estimation in
707 EPICv2 samples, and (ii) the auto method which selects the top 100 probes with F-stat *p*-value < 1E-8 for
708 each cell type with the greatest magnitude of methylation difference, both implemented in the
709 *FlowSorted.Blood.EPIC* R package with noob corrected values as recommended(Salas et al. 2022).

710 **Estimation of epigenetic age and epigenetic age acceleration**

711 We compared the performance of eight commonly used epigenetic clocks between EPICv1 and EPICv2
712 in VHAS, CLHNS, CALERIE, and the external validation cohort BeCOME. Epigenetic clocks analyses on
713 capillary blood samples in the VHAS cohort were also performed (Supplementary material, Supplementary
714 Fig. S27-29). Epigenetic age of first generation clocks including Horvath pan-tissue(Horvath 2013),
715 Hannum(Hannum et al. 2013), and Horvath skin and blood clocks(Horvath et al. 2018), and second
716 generation clocks including PhenoAge(Levine et al. 2018), and GrimAge(Lu et al. 2019a) were obtained
717 from the online DNA Methylation Age Calculator (https://dnamage.genetics.ucla.edu/new). Missing clock
718 CpG β value imputation was performed by the clock algorithms. For these clocks, we calculated epigenetic
719 age acceleration (EAA), a measure of rate of aging commonly used in epigenetic clock investigations, by
720 employing three approaches independently for each cohort.

721 1. **EPIC version separate:** we separated epigenetic ages of EPICv1 and EPICv2 samples and then
722 calculated EAA independently on EPICv1 and EPICv2 samples by extracting residuals from the
723 linear regression model: Epigenetic age ~ chronological age.
724 2. **EPIC versions combined:** we first combined epigenetic ages of EPICv1 and EPICv2 samples and
725 then calculated EAA by extracting residuals from the linear regression model: Epigenetic age ~
726 chronological age.
727 3. **EPIC versions combined and version adjusted:** we first combined epigenetic ages of EPICv1
728 and EPICv2 samples and calculated EAA by extracting residuals from the linear regression model:
729 Epigenetic age ~ chronological age + EPIC version.

730 To evaluate if there are significant differences in principal components (PC) clock(Higgins-Chen et al.
731 2022) estimates, which are more robust to technical noise, by EPIC version, we first estimated epigenetic
732 ages by the R script provided (https://github.com/MorganLevineLab/PC-Clocks) and then calculated EAAs
733 using the approaches mentioned above.
734

735 DunedinPACE, a rate-based clock, was calculated using the DunedinPACE R package(Belsky et al.
736 2022), and the two other clocks, epiTOC and DNA methylation-based estimator of telomere length
737 (DNAmTL)(Lu et al. 2019b) were obtained using the *getEpiTOC* function in the *cgageR* R package(Yang
738 et al. 2016) and the online DNA Methylation Age Calculator (https://dnamage.genetics.ucla.edu/new),
739 respectively. Missing clock CpG β value imputation was performed by the clock algorithms. Unlike EAA
740 calculation, there were no secondary measures calculated from these three clock estimates in our analyses.
741 Estimates remained the same when calculated by the EPIC versions separate or combined approach,
742 therefore, we employed only two approaches for each cohort:

743 1. **EPIC version separate:** rate-based and other clock estimates were calculated for EPICv1 and
744 EPICv2 samples without EPIC version adjustment.
745 2. **EPIC versions combined and version adjusted:** rate-based and other clock estimates were first
746 calculated for EPICv1 and EPICv2 samples, and were subsequently adjusted for EPIC versions by
747 regressing out EPIC versions using linear regression model: Rate estimate ~ EPIC version.

748 Using paired *t*-tests and applying a Bonferroni multiple test correction, we evaluated differences in
749 epigenetic clock estimations and EAAs (as well as biomarker predictor scores and cell type proportions
750 described below) between matched samples assessed on EPICv1 and EPICv2 in a cohort-specific manner.
751 Statistical significance was defined as Bonferroni adjusted p-value < 0.05. Effect sizes were measured by
752 Cohen's d, and classified as "small" (d = 0.2-0.49), "medium" (d = 0.5-0.79), and "large" ($d \geq 0.8$) based on
753 recommended benchmarks(Cohen 1988).

754 **Estimation of DNA methylation-based inflammation, smoking, and alcohol scores**

755 Among other DNA methylation-based tools are inflammation, smoking and alcohol score predictors, which
756 provide biomarker measures that correlate with levels of inflammatory markers (IL-6, CRP), smoking and
757 alcohol use, respectively. DNA methylation-derived scores of IL-6, CRP, smoking and alcohol use were

758 calculated as a weighted sum of coefficients derived from published lists of predictive CpGs(McCartney et
759 al. 2018; Stevenson et al. 2021; Wielscher et al. 2022). Being that these biomarkers were trained on previous
760 Illumina arrays, we sought out to determine the correlation of derived DNA methylation-based scores
761 between EPICv1 and EPICv2, and compare these scores without version adjustment (EPIC version
762 separate) and with version adjustment (EPIC versions combined and adjusted) using approaches similar to
763 the rate-based and other epigenetic clocks.

764 **Data Availability**

765 Raw data generated in this study are deposited in the Gene Expression Omnibus database (GEO) with
766 accession GSE286313. The EPICv2 preprocessing pipeline and analysis scripts in R are available on
767 GitHub https://github.com/kobor-lab/EPICv2_QC_preprocessing, and https://github.com/kobor-
768 lab/EPICv1v2_comparison_manuscript.

769 **Conflict of interest**

770 D.W.B. is listed as an inventor of the Duke University and University of Otago invention DunedinPACE,
771 which is licensed to TruDiagnostic. The other authors declare that they have no competing interests.

772 **Acknowledgements**

773 We thank Dr. Sarah Merrill, Maggie Fu, and Dr. Kimberly Schmidt for their helpful feedback during
774 manuscript preparation. We would also like to thank all cohort participants for their involvement and
775 providing biological samples for research, and cohort study teams for their efforts. We wish to thank
776 Monika Rex-Haffner and the whole BeCOME study team for their technical assistance. This research
777 utilized the *FlowSorted.BloodExtended.EPIC* software packages developed at Dartmouth College, which
778 are governed by the licensing terms provided by Dartmouth Technology Transfer. D.W.B. and M.S.K. are
779 fellows of the CIFAR CBD Network. This work received supported from the Canadian Institutes of Health
780 Research (CIHR) / Project Grant (Sponsor Reference Number: PJT-175309), National Institute on Aging

of the National Institutes of Health under Award Number R01AG052537, National Institutes of Health / Research Project Grant (R01) (Sponsor Reference Number: 5R01AG061006-05), US National Institute on Aging Grants R01AG061378, Alexander von Humboldt Foundation for the German Chancellor Fellowship, the National Institute of General Medical Sciences (Award Number T32GM144273), BC Children's Hospital Research Investigator Grant Award Program, BC Children's Hospital Research Institute Establishment Award, and Edwin S.H. Leong Centre for Healthy Aging.

Author contributions

Conceptualization: B.C.Z, M.S.J., C.K., K.Korthauer, M.S.K.; Data curation: B.C.Z, M.S.J., N.Y., C.P.R., H-R.E., J.W., A.A.H., K.Korinek, K.M.H., N.R.L., D.C., E.B.B., T.K.T., T.W.M., C.W.K., Z.Z., D.W.B., M.S.K.; DNA methylation data generation: J.L.M., K.D., M.S.K. (Kobor Lab facility), E.B.B. (MPIP); Formal analysis: B.C.Z., M.S.J., C.K.; Methodology: B.C.Z, M.S.J., C.K., K.Korthauer, M.S.K.; Resources: T.K.T., K. Korinek, K.M.H., N.R.L., D.C., E.B.B., T.W.M., C.W.K., Z.Z., D.W.B., M.S.K.; Visualization: B.C.Z., M.S.J.; Writing – original draft: M.S.J., C.K., B.C.Z., M.S.K.; Writing – review & editing: all authors; Cohort Owners/Custodians: T.K.T., K. Korinek, K.M.H., N.R.L., E.B.B., T.W.M., C.W.K., Z.Z., D.W.B.

References

Adair LS, Popkin BM, Akin JS, Guilkey DK, Gultiano S, Borja J, Perez L, Kuzawa CW, McDade T, Hindin MJ. 2011. Cohort Profile: The Cebu Longitudinal Health and Nutrition Survey. *Int J Epidemiol* **40**: 619–625.

Battram T, Yousefi P, Crawford G, Prince C, Sheikhali Babaei M, Sharp G, Hatcher C, Vega-Salas MJ, Khodabakhsh S, Whitehurst O, et al. 2022. The EWAS Catalog: a database of epigenome-wide association studies. *Wellcome Open Res* **7**: 41.

Belsky DW, Caspi A, Corcoran DL, Sugden K, Poulton R, Arseneault L, Baccarelli A, Chamarti K, Gao X, Hannon E, et al. 2022. DunedinPACE, a DNA methylation biomarker of the pace of aging eds. J. Deelen, J.K. Tyler, M. Suderman, and J. Deelen. *eLife* **11**: e73420.

806 Bergstedt J, Azzou SAK, Tsuo K, Jaquaniello A, Urrutia A, Rotival M, Lin DTS, MacIsaac JL,
807 Kobor MS, Albert ML, et al. 2022. The immune factors driving DNA methylation variation
808 in human blood. *Nat Commun* **13**: 5895.

809 Bishara AJ, Hittner JB. 2015. Reducing Bias and Error in the Correlation Coefficient Due to
810 Nonnormality. *Educ Psychol Meas* **75**: 785–804.

811 Bose M, Wu C, Pankow JS, Demerath EW, Bressler J, Fornage M, Grove ML, Mosley TH, Hicks
812 C, North K, et al. 2014. Evaluation of microarray-based DNA methylation measurement
813 using technical replicates: the Atherosclerosis Risk In Communities (ARIC) Study. *BMC*
814 *Bioinformatics* **15**: 312.

815 Breton CV, Marsit CJ, Faustman E, Nadeau K, Goodrich JM, Dolinoy DC, Herbstman J, Holland
816 N, LaSalle JM, Schmidt R, et al. 2017. Small-Magnitude Effect Sizes in Epigenetic End
817 Points are Important in Children's Environmental Health Studies: The Children's
818 Environmental Health and Disease Prevention Research Center's Epigenetics Working
819 Group. *Environ Health Perspect* **125**: 511–526.

820 Brückl TM, Spoormaker VI, Sämann PG, Brem A-K, Henco L, Czamara D, Elbau I, Grandi NC,
821 Jollans L, Kühnel A, et al. 2020. The biological classification of mental disorders
822 (BeCOME) study: a protocol for an observational deep-phenotyping study for the
823 identification of biological subtypes. *BMC Psychiatry* **20**: 213.

824 Buhule OD, Minster RL, Hawley NL, Medvedovic M, Sun G, Viali S, Deka R, McGarvey ST,
825 Weeks DE. 2014. Stratified randomization controls better for batch effects in 450K
826 methylation analysis: a cautionary tale. *Front Genet* **5**: 354.

827 Chen BH, Zhou W. 2024. mLiftOver: harmonizing data across Infinium DNA methylation
828 platforms. *Bioinformatics* **40**: btae423.

829 Cheung K, Burgers MJ, Young DA, Cockell S, Reynard LN. 2020. Correlation of Infinium
830 HumanMethylation450K and MethylationEPIC BeadChip arrays in cartilage. *Epigenetics*
831 **15**: 594–603.

832 Cohen J. 1988. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Routledge, New
833 York.

834 Cronjé HT, Nienaber-Rousseau C, Min JL, Green FR, Elliott HR, Pieters M. 2021. Comparison of
835 DNA methylation clocks in Black South African men. *Epigenomics* **13**: 437–449.

836 Dhingra R, Kwee LC, Diaz-Sanchez D, Devlin RB, Cascio W, Hauser ER, Gregory S, Shah S,
837 Kraus WE, Olden K, et al. 2019. Evaluating DNA methylation age on the Illumina
838 MethylationEPIC Bead Chip. *PLOS ONE* **14**: e0207834.

839 Engelbrecht H-R, Merrill SM, Gladish N, MacIsaac JL, Lin DTS, Ecker S, Chrysohoou CA, Pes
840 GM, Kobor MS, Rehkopf DH. 2022. Sex differences in epigenetic age in Mediterranean
841 high longevity regions. *Front Aging* **3**: 1007098.

842 Fortin J-P, Labbe A, Lemire M, Zanke BW, Hudson TJ, Fertig EJ, Greenwood CM, Hansen KD.
843 2014. Functional normalization of 450k methylation array data improves replication in
844 large cancer studies. *Genome Biol* **15**: 503.

845 Garma LD, Quintela-Fandino M. 2024. Applicability of epigenetic age models to next-generation
846 methylation arrays. *Genome Med* **16**: 116.

847 Gaunt TR, Shihab HA, Hemani G, Min JL, Woodward G, Lyttleton O, Zheng J, Duggirala A,
848 McArdle WL, Ho K, et al. 2016. Systematic identification of genetic influences on
849 methylation across the human life course. *Genome Biol* **17**: 61.

850 Goldberg DC, Cloud C, Lee SM, Barnes B, Gruber S, Kim E, Pottekat A, Westphal M, McAuliffe
851 L, Majournie E, et al. 2024. MSA: scalable DNA methylation screening BeadChip for high-
852 throughput trait association studies. *bioRxiv* 2024.05.17.594606.

853 Hannum G, Guinney J, Zhao L, Zhang L, Hughes G, Sadda S, Klotzle B, Bibikova M, Fan J-B,
854 Gao Y, et al. 2013. Genome-wide Methylation Profiles Reveal Quantitative Views of
855 Human Aging Rates. *Mol Cell* **49**: 359–367.

856 Hawe JS, Wilson R, Schmid KT, Zhou L, Lakshmanan LN, Lehne BC, Kühnel B, Scott WR,
857 Wielscher M, Yew YW, et al. 2022. Genetic variation influencing DNA methylation
858 provides insights into molecular mechanisms regulating genomic function. *Nat Genet* **54**:
859 18–29.

860 Higgins-Chen AT, Thrush KL, Wang Y, Minteer CJ, Kuo P-L, Wang M, Niimi P, Sturm G, Lin J,
861 Moore AZ, et al. 2022. A computational solution for bolstering reliability of epigenetic
862 clocks: Implications for clinical trials and longitudinal tracking. *Nat Aging* **2**: 644–661.

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

864 Horvath S, Gurven M, Levine ME, Trumble BC, Kaplan H, Allayee H, Ritz BR, Chen B, Lu AT,
865 Rickabaugh TM, et al. 2016. An epigenetic clock analysis of race/ethnicity, sex, and
866 coronary heart disease. *Genome Biol* **17**: 171.

867 Horvath S, Oshima J, Martin GM, Lu AT, Quach A, Cohen H, Felton S, Matsuyama M, Lowe D,
868 Kabacik S, et al. 2018. Epigenetic clock for skin and blood cells applied to Hutchinson
869 Gilford Progeria Syndrome and *ex vivo* studies. *Aging* **10**: 1758–1775.

870 Houseman EA, Molitor J, Marsit CJ. 2014. Reference-free cell mixture adjustments in analysis of
871 DNA methylation data. *Bioinforma Oxf Engl*.

872 Illumina Inc. 2015. BeadArray Controls Reporter Software Guide.

873 Infinium MethylationEPIC v2.0 BeadChip. 2023. Infinium MethylationEPIC v2.0 BeadChip.

874 Jones MJ, Moore SR, Kobor MS. 2017. Principles and Challenges of Applying Epigenetic
875 Epidemiology to Psychology. 30.

876 Kaur D, Lee SM, Goldberg D, Spix NJ, Hinoue T, Li H-T, Dwaraka VB, Smith R, Shen H, Liang
877 G, et al. 2023. Comprehensive evaluation of the Infinium human MethylationEPIC v2
878 BeadChip. *Epigenetics Commun* **3**: 6.

879 Koch CM, Chiu SF, Akbarpour M, Bharat A, Ridge KM, Bartom ET, Winter DR. 2018. A
880 Beginner's Guide to Analysis of RNA Sequencing Data. *Am J Respir Cell Mol Biol* **59**:
881 145–157.

882 Kolde R. 2019. pheatmap: Pretty Heatmaps.

883 Kong Y, Rastogi D, Seoighe C, Greally JM, Suzuki M. 2019. Insights from deconvolution of cell
884 subtype proportions enhance the interpretation of functional genomic data. *PLOS ONE* **14**:
885 e0215987.

886 Konwar C, Asiimwe R, Inkster AM, Merrill SM, Negri GL, Aristizabal MJ, Rider CF, MacIsaac
887 JL, Carlsten C, Kobor MS. 2021. Risk-focused differences in molecular processes
888 implicated in SARS-CoV-2 infection: corollaries in DNA methylation and gene
889 expression. *Epigenetics Chromatin* **14**: 54.

890 Koo TK, Li MY. 2016. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients
891 for Reliability Research. *J Chiropr Med* **15**: 155–163.

892 Korinek K, Teerawichitchainan B, Zimmer Z, Brindle E, Nguyen TKC, Nguyen HM, Tran KT.
893 2019. Design and measurement in a study of war exposure, health, and aging: protocol for
894 the Vietnam health and aging study. *BMC Public Health* **19**: 1351.

895 Kuzawa CW, Ryan CP, Adair LS, Lee NR, Carba DB, MacIsaac JL, Dever K, Atashzay P, Kobor
896 MS, McDade TW. 2022. Birth weight and maternal energy status during pregnancy as
897 predictors of epigenetic age acceleration in young adults from metropolitan Cebu,
898 Philippines. *Epigenetics* **17**: 1535–1545.

899 Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. 2012. The sva package for removing batch
900 effects and other unwanted variation in high-throughput experiments. *Bioinforma Oxf Engl*
901 **28**: 882–883.

902 Levenstien MA, Yang Y, Ott J. 2003. Statistical significance for hierarchical clustering in genetic
903 association and microarray expression studies. *BMC Bioinformatics* **4**: 62.

904 Levine ME, Lu AT, Quach A, Chen BH, Assimes TL, Bandinelli S, Hou L, Baccarelli AA, Stewart
905 JD, Li Y, et al. 2018. An epigenetic biomarker of aging for lifespan and healthspan. *Aging*
906 **10**: 573–591.

907 Li M, Zou D, Li Z, Gao R, Sang J, Zhang Y, Li R, Xia L, Zhang T, Niu G, et al. 2019. EWAS
908 Atlas: a curated knowledgebase of epigenome-wide association studies. *Nucleic Acids Res*
909 **47**: D983–D988.

910 Lin W-Y. 2023. Epigenetic clocks derived from western samples differentially reflect Taiwanese
911 health outcomes. *Front Genet* **14**: 1089819.

912 Logue MW, Smith AK, Wolf EJ, Maniates H, Stone A, Schichman SA, McGlinchey RE, Milberg
913 W, Miller MW. 2017. The correlation of methylation levels measured using Illumina 450K
914 and EPIC BeadChips in blood samples. *Epigenomics* **9**: 1363–1371.

915 Lu AT, Quach A, Wilson JG, Reiner AP, Aviv A, Raj K, Hou L, Baccarelli AA, Li Y, Stewart JD,
916 et al. 2019a. DNA methylation GrimAge strongly predicts lifespan and healthspan. *Aging*
917 **11**: 303–327.

918 Lu AT, Seeboth A, Tsai P-C, Sun D, Quach A, Reiner AP, Kooperberg C, Ferrucci L, Hou L,
919 Baccarelli AA, et al. 2019b. DNA methylation-based estimator of telomere length. *Aging*
920 **11**: 5895–5923.

921 Lussier AA, Schuurmans IK, Großbach A, MacIsaac J, Dever K, Koen N, Zar HJ, Stein DJ, Kobor
922 MS, Dunn EC. 2024. Technical variability across the 450K, EPICv1, and EPICv2 DNA
923 methylation arrays: lessons learned for clinical and longitudinal studies. *Clin Epigenetics*
924 **16**: 166.

925 McCartney DL, Hillary RF, Stevenson AJ, Ritchie SJ, Walker RM, Zhang Q, Morris SW,
926 Bermingham ML, Campbell A, Murray AD, et al. 2018. Epigenetic prediction of complex
927 traits and death. *Genome Biol* **19**: 136.

928 McCartney DL, Walker RM, Morris SW, McIntosh AM, Porteous DJ, Evans KL. 2016.
929 Identification of polymorphic and off-target probe binding sites on the Illumina Infinium
930 MethylationEPIC BeadChip. *Genomics Data* **9**: 22–24.

931 McEwen LM, Jones MJ, Lin DTS, Edgar RD, Husquin LT, MacIsaac JL, Ramadori KE, Morin
932 AM, Rider CF, Carlsten C, et al. 2018. Systematic evaluation of DNA methylation age
933 estimation with common preprocessing methods and the Infinium MethylationEPIC
934 BeadChip array. *Clin Epigenetics* **10**: 123.

935 Merrill SM, Gladish N, Fu MP, Moore SR, Konwar C, Giesbrecht GF, MacIssac JL, Kobor MS,
936 Letourneau NL. 2023. Associations of peripheral blood DNA methylation and estimated
937 monocyte proportion differences during infancy with toddler attachment style. *Attach Hum*
938 *Dev* **25**: 132–161.

939 Murat K, Grüning B, Poterlowicz PW, Westgate G, Tobin DJ, Poterlowicz K. 2020. Ewastools:
940 Infinium Human Methylation BeadChip pipeline for population epigenetics integrated into
941 Galaxy. *GigaScience* **9**: giaa049.

942 Noguera-Castells A, García-Prieto CA, Álvarez-Errico D, Esteller M. 2023. Validation of the new
943 EPIC DNA methylation microarray (900K EPIC v2) for high-throughput profiling of the
944 human DNA methylome. *Epigenetics* **18**: 2185742.

945 Nygaard V, Rødland EA, Hovig E. 2016. Methods that remove batch effects while retaining group
946 differences may lead to exaggerated confidence in downstream analyses. *Biostat Oxf Engl*
947 **17**: 29–39.

948 Olstad EW, Nordeng HME, Sandve GK, Lyle R, Gervin K. 2022. Low reliability of DNA
949 methylation across Illumina Infinium platforms in cord blood: implications for replication
950 studies and meta-analyses of prenatal exposures. *Clin Epigenetics* **14**: 80.

951 Ori APS, Lu AT, Horvath S, Ophoff RA. 2022. Significant variation in the performance of DNA
952 methylation predictors across data preprocessing and normalization strategies. *Genome*
953 *Biol* **23**: 225.

954 Peters TJ, Meyer B, Ryan L, Achinger-Kawecka J, Song J, Campbell EM, Qu W, Nair S, Loi-Luu
955 P, Stricker P, et al. 2024. Characterisation and reproducibility of the
956 HumanMethylationEPIC v2.0 BeadChip for DNA methylation profiling. *BMC Genomics*
957 **25**: 251.

958 Pidsley R, Zotenko E, Peters TJ, Lawrence MG, Risbridger GP, Molloy P, Van Djik S,
959 Muhlhausler B, Stirzaker C, Clark SJ. 2016. Critical evaluation of the Illumina
960 MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling.
961 *Genome Biol* **17**: 208.

962 Pourcq SD, Kuo P-L, Moore AZ, Bandinelli S, Horvath S, Ferrucci L, Santini V. 2024.
963 Compatibility and comparative analysis of chronological and biological aging between the
964 legacy 450K and the EPIC v2.0 arrays. 2024.10.17.618672.
965 https://www.biorxiv.org/content/10.1101/2024.10.17.618672v1 (Accessed November 6,
966 2024).

967 Price EM, Cotton AM, Lam LL, Farré P, Emberly E, Brown CJ, Robinson WP, Kobor MS. 2013.
968 Additional annotation enhances potential for biologically-relevant analysis of the Illumina
969 Infinium HumanMethylation450 BeadChip array. *Epigenetics Chromatin* **6**: 4.

970 Price EM, Robinson WP. 2018. Adjusting for Batch Effects in DNA Methylation Microarray Data,
971 a Lesson Learned. *Front Genet* **9**: 83.

972 R Core Team. 2022. R: A language and environment for statistical computing. R Foundation for
973 Statistical Computing. https://www.R-project.org/.

974 Rochon J, Bales CW, Ravussin E, Redman LM, Holloszy JO, Racette SB, Roberts SB, Das SK,
975 Romashkan S, Galan KM, et al. 2011. Design and conduct of the CALERIE study:
976 comprehensive assessment of the long-term effects of reducing intake of energy. *J*
977 *Gerontol A Biol Sci Med Sci* **66**: 97–108.

978 Ryan CP, Jones MJ, Edgar RD, Lee NR, Kobor MS, McDade TW, Kuzawa CW. 2022. Immune
979 cell type and DNA methylation vary with reproductive status in women: possible pathways
980 for costs of reproduction. *Evol Med Public Health* **10**: 47–58.

981 Salas LA, Koestler DC, Butler RA, Hansen HM, Wiencke JK, Kelsey KT, Christensen BC. 2018.
982 An optimized library for reference-based deconvolution of whole-blood biospecimens
983 assayed using the Illumina HumanMethylationEPIC BeadArray. *Genome Biol* **19**: 64.

984 Salas LA, Zhang Z, Koestler DC, Butler RA, Hansen HM, Molinaro AM, Wiencke JK, Kelsey
985 KT, Christensen BC. 2022. Enhanced cell deconvolution of peripheral blood using DNA
986 methylation for high-resolution immune profiling. *Nat Commun* **13**: 761.

987 Solomon O, MacIsaac J, Quach H, Tindula G, Kobor MS, Huen K, Meaney MJ, Eskenazi B,
988 Barcellos LF, Holland N. 2018. Comparison of DNA methylation measured by Illumina
989 450K and EPIC BeadChips in blood of newborns and 14-year-old children. *Epigenetics*
990 **13**: 655–664.

991 Stevenson AJ, Gadd DA, Hillary RF, McCartney DL, Campbell A, Walker RM, Evans KL, Harris
992 SE, Spires-Jones TL, McRae AF, et al. 2021. Creating and Validating a DNA Methylation-
993 Based Proxy for Interleukin-6. *J Gerontol A Biol Sci Med Sci* **76**: 2284–2292.

994 Sugden K, Hannon EJ, Arseneault L, Belsky DW, Corcoran DL, Fisher HL, Houts RM,
995 Kandaswamy R, Moffitt TE, Poulton R, et al. 2020. Patterns of Reliability: Assessing the
996 Reproducibility and Integrity of DNA Methylation Measurement. *Patterns* **1**: 100014.

997 Tay JH, Chew YE, Wang W, Lim ZM, Guan L, Dorajoo R, Kennedy BK, Brooke R, Gordevicius
998 J, Horvath S, et al. 2025. DNAm age differences between infinium methylationEPICv1 vs
999 EPICv2 in buffy coat, PBMC, and saliva samples. *Commun Biol* **8**: 1–8.

1000 Teschendorff AE, Marabita F, Lechner M, Bartlett T, Tegner J, Gomez-Cabrero D, Beck S. 2013.
1001 A beta-mixture quantile normalization method for correcting probe design bias in Illumina
1002 Infinium 450 k DNA methylation data. *Bioinformatics* **29**: 189–196.

1003 Triche TJ, Weisenberger DJ, Van Den Berg D, Laird PW, Siegmund KD. 2013. Low-level
1004 processing of Illumina Infinium DNA Methylation BeadArrays. *Nucleic Acids Res* **41**: e90.

1005 Wielscher M, Mandaviya PR, Kuehnel B, Joehanes R, Mustafa R, Robinson O, Zhang Y, Bodinier
1006 B, Walton E, Mishra PP, et al. 2022. DNA methylation signature of chronic low-grade
1007 inflammation and its role in cardio-respiratory diseases. *Nat Commun* **13**: 2408.

1008 Yang Z, Wong A, Kuh D, Paul DS, Rakyan VK, Leslie RD, Zheng SC, Widschwendter M, Beck
1009 S, Teschendorff AE. 2016. Correlation of an epigenetic mitotic clock with cancer risk.
1010 *Genome Biol* **17**: 205.

1011 Zhang W, Young JI, Gomez L, Schmidt MA, Lukacsovich D, Varma A, Chen XS, Kunkle B,
1012 Martin ER, Wang L. 2024. Critical evaluation of the reliability of DNA methylation probes
1013 on the Illumina MethylationEPIC v1.0 BeadChip microarrays. *Epigenetics* **19**: 2333660.

1014 Zheng SC, Beck S, Jaffe AE, Koestler DC, Hansen KD, Houseman AE, Irizarry RA, Teschendorff
1015 AE. 2017. Correcting for cell-type heterogeneity in epigenome-wide association studies:
1016 revisiting previous analyses. *Nat Methods* **14**: 216–217.

1017 Zindler T, Frieling H, Neyazi A, Bleich S, Friedel E. 2020. Simulating ComBat: how batch
1018 correction can lead to the systematic introduction of false positive results in DNA
1019 methylation microarray studies. *BMC Bioinformatics* **21**: 271.

1020