Comparison of Infinium MethylationEPIC v2.0 to v1.0 for human population epigenetics: considerations for addressing EPIC version differences in DNA methylation-based tools

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

† These authors contributed equally and should be considered as co-first authors
* Please address all correspondence concerning this article to Dr. Michael S. Kobor

1. BC Children's Hospital Research Institute, 950 West 28th Avenue, Vancouver, BC, V5Z 4H4, Canada
2. Department of Medical Genetics, Faculty of Medicine, University of British Columbia, Vancouver, BC, V6T 1Z3, Canada
3. Robert N. Butler Columbia Aging Center, Mailman School of Public Health, Columbia University, New York, NY 10032, USA
4. Family Medicine Department, Hanoi Medical University, Hanoi, Vietnam
5. Department of Sociology, University of Utah, Salt Lake City, USA
6. Department of Family Studies and Gerontology, Mount Saint Vincent University, Halifax, Canada
7. Canada Research Chair, Global Aging and Community Initiative
8. Duke University School of Medicine, Durham, NC, 27701, USA
9. USC-Office of Population Studies Foundation, Inc., University of San Carlos, Cebu City, Philippines
10. Department of Anthropology, Northwestern University, Evanston, Illinois, USA
11. Program in Child and Brain Development, CIFAR, Toronto, Ontario, Canada
12. Department of Anthropology and Institute for Policy Research, Northwestern University, Evanston, IL 60208, USA
13. Butler Columbia Aging Center, Columbia University Mailman School of Public Health, New York, NY, USA
14. Department of Epidemiology, Columbia University Mailman School of Public Health, New York, NY, USA
15. The Edwin S.H. Leong UBC Healthy Aging Chair-A UBC President's Excellence Chair, University of British Columbia, Canada

Keywords (3-10)

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

Abstract

Background:
The recently launched DNA methylation profiling platform, Illumina MethylationEPIC BeadChip Infinium microarray v2.0 (EPICv2), is highly correlated with measurements obtained from its predecessor MethylationEPIC BeadChip Infinium microarray v1.0 (EPICv1). However, the concordance between the two versions in the context of DNA methylation-based tools, including cell type deconvolution algorithms, epigenetic clocks, and inflammation and lifestyle biomarkers has not yet been investigated.

Findings:
We profiled DNA methylation on both EPIC versions using matched venous blood samples from individuals spanning early to late adulthood across three cohorts. On combining the DNA methylomes of the cohorts, we observed that samples primarily clustered by the EPIC version they were measured on. Within each cohort, when we calculated cell type proportions, epigenetic age acceleration (EAA), rate of aging estimates, and biomarker scores for the matched samples on each version, we noted significant differences between EPICv1 and EPICv2 in the majority of these estimates. These differences were not significant, however, when estimates were adjusted for EPIC version or when EAAs were calculated separately for each EPIC version.

Conclusions:
Our findings indicate that EPIC version differences predominantly explain DNA methylation variation and influence estimates of DNA methylation-based tools, and therefore we recommend caution when combining cohorts run on different versions. We demonstrate the importance of calculating DNA methylation-based estimates separately for each EPIC version or accounting for EPIC version either as a covariate in statistical models or by using version correction algorithms.

Background

Infinium Methylation BeadChip microarrays have been widely used to cost-effectively measure the human DNA methylome in large scale and population-wide studies[1–3]. 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[4]. 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, incorporation of new "nv" probes targeting recurrent somatic cancer mutations, and 65 single nucleotide polymorphism (SNP) sites, compared to the 59 SNPs on EPICv1, for confirmation of sample identity. 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[5]. 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[6,7]. 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[6,7].

Previous iterations of Illumina microarrays (27K, 450K and EPICv1) have been widely used to develop DNA methylation-based tools including cell type deconvolution algorithms[8–10], epigenetic clocks[11–18], interleukin 6 (IL-6) and C-reactive protein (CRP) inflammation markers[19,20], and lifestyle biomarker predictors such as smoking and alcohol use [21]. These tools have been trained on one or more of the previous generation of microarrays, and are based on the property that DNA methylation levels at specific cytosine-guanine dinucleotides (CpGs) highly correlate with cell types, chronological age, and biomarker measures, respectively[11–14,16,17], and a majority of predictive CpGs employed by these tools are retained on EPICv2[6]. It has been previously shown that estimates from some of these DNA methylation-based tools, while highly correlated, are significantly different between 450K and EPICv1[22–24]. With the launch of the recent version of the EPICv2 array, we thus sought to examine whether there are any such differences between EPICv1 and EPICv2, we systematically assessed the concordance of inferred estimates between the two EPIC versions. We also tested whether there were any differences in estimates between EPICv1 and EPICv2 when datasets profiled on the two versions are combined, given that this is an important consideration in meta-analyses which harness statistical power that comes from combining multiple cohorts, and longitudinal studies which include samples profiled on different arrays/versions.

To assess whether EPIC version differences have any implications on DNA methylation measures in the context of DNA methylation-based tools, we examined matched venous blood samples from individuals across three diverse populations spanning early to late adulthood, and profiled DNA methylome using the EPICv1 and EPICv2 arrays. To explore if the high agreement between EPICv1 and EPICv2 reported in cell lines is consistent in primary samples from human populations, we tested the concordance of the two EPIC versions at both array and probe levels and evaluated the potential contribution of EPIC version to overall DNA methylation variation. Next, to investigate whether there are differences in estimates between EPICv1 and EPICv2 from

DNA methylation-based tools built on previous arrays/versions, including immune cell type deconvolution algorithms, epigenetic clocks, inflammation and lifestyle biomarkers, we compared their estimates between the two EPIC versions. We also examined whether EPIC version influenced these estimates by combining EPICv1 and EPICv2 samples, and demonstrated the significance of accounting for EPIC version in these DNA methylation-based estimates.

Results

Unsupervised clustering of samples was primarily determined by EPIC version

To compare the two most recent generations of Illumina MethylationEPIC BeadChip Infinium microarrays, we measured the DNA methylomes of a subset of venous whole blood samples on both EPICv1 and EPICv2 across three cohorts: (i) Vietnam Health and Aging Study (VHAS)[25], (ii) Cebu Longitudinal Health and Nutrition Survey (CLHNS)[26], and (iii) Comprehensive Assessment of Long-term Effects of Reducing Intake of Energy (CALERIE)[27] (Table 1, Figure 1). We first explored whether samples clustered depending on which EPIC version they were measured on or the cohort they belong to, given that the three cohorts were unique in their demographic and biological characteristics such as sex and age range. To do this, we carried out unsupervised hierarchical clustering on all samples from the three cohorts using the 721,378 probes shared between EPICv1 and EPICv2. We observed a clear demarcation of samples first by EPIC version despite cohort-specific characteristics (Figure 2A). When clustering was carried out on subsets of predictive CpGs employed by DNA methylation-based tools, we still noted some separation of samples by EPIC version, although not to the same extent as the clustering on shared probes (Supplementary Figure 1).

Spearman correlation between shared EPICv1 and EPICv2 probes was high at array level and low at the individual probe level

To compare the technical performance of EPICv1 and EPICv2, we first estimated Spearman correlation of probes between matched samples assessed on the two EPIC versions at the array level and at the individual probe level for each cohort independently. We identified overall high array level Spearman correlations between EPICv1 and EPICv2 in the three cohorts (VHAS: Spearman *rho* = 0.969-0.981, CLHNS: 0.973-0.979, and CALERIE: 0.968-0.981). As expected, technical replicate samples measured on the same EPIC version showed marginally higher array level Spearman correlation compared to those measured between versions (within EPICv1: 0.991, within EPICv2: 0.986, and between versions: 0.976). Similarly, samples from the same individual measured on EPICv1 and EPICv2 showed higher absolute β value differences for each probe when compared to differences calculated between technical replicates within each EPIC version. Specifically, we noted that ~75% of probes have absolute β value differences between EPIC versions higher than technical replicates within each EPIC version (VHAS: 74.258% (535,682 probes); CLHNS: 76.365% (550,880 probes); and CALERIE: 76.733% (553,821 probes)). On evaluating probe level Spearman correlations between the two versions, we identified that only a fourth of the investigated probes exhibited Spearman correlation greater than 0.70, specifically, VHAS: 29.220% (210,845 probes), CLHNS: 24.965% (180,093 probes), and CALERIE: 19.287% (139,134 probes)). The remaining ~75% of probes had a low median correlation between EPICv1 and EPICv2, specifically, VHAS: 0.390 (range: -0.648 to 0.991), CLHNS: 0.368 (range: -0.907 to 1.000), and CALERIE: 0.389 (range: -0.744-0.998). One cause for low probe level correlations

may be low inter-sample variability[22]. To assess if observed low individual probe level correlations in the majority of investigated probes can indeed be attributed to low DNA methylation variability across samples, we used a RMSE threshold of < 5% to denote low variability (Figure 2B-D). We observed that of the probes with Spearman correlation < 0.70 between EPICv1 and EPICv2, more than 85% can be attributed to respective RMSE values estimated at less than 5% (VHAS: 89.752% (485,211 probes), CLHNS: 87.994% (476,299 probes), and CALERIE: 91.805% (534,530 probes)). In summary, ~96% of the shared probes between EPICv1 and EPICv2 had either modest-to-high correlation or low correlation explained by low DNA methylation variance, with 87.065% (628,069 probes) commonly identified across three cohorts. When we compared these ~96% probes to platform-bias free and high-confidence mapping probes (n = 542,491) recently identified in cell lines [28], we found that over 95% overlap across our investigated cohorts (VHAS: 96.642% (521,395 probes), CLHNS: 95.011% (512,598 probes), and CALERIE: 97.017% (523,418 probes)).

While the low correlation in the majority of shared EPICv1 and EPICv2 probes has been explained, ~4% probes still remain unexplained by DNA methylation variance (VHAS: 52,322 probes, CLHNS: 64,986 probes; and CALERIE: 47,640 probes; 25,058 common across all cohorts) (Supplementary Table S1). We noted that the vast majority of these probes which have low correlation and high RMSE had % absolute β value differences higher than technical noise measured within each EPIC version (VHAS: 70.217% (36,739 probes), CLHNS: 93.112% (60,510 probes), and CALERIE: 74.733% (35,658 probes)). On comparing these probes to our list of flagged probes with detection *p*-value > 0.01 or beadcount < 3 in greater than 1% of the samples, we identified an overlap of 4,977, 2,141, and 1,292 probes in the three cohorts, respectively. We next sought to determine whether design type switch or cross hybridization of probes to the genome can explain the observed low correlation and high RMSE. We noted that 14, 17, and 15 probes in VHAS, CLHNS, and CALERIE, respectively, overlapped with the 82 probes that have undergone design type switches in EPICv2[6], and a small fraction of these probes overlapped with previously identified as cross hybridizing to multiple genomic locations or mapped to genetic variant sites (742, 1132, and 633 probes in the Peters reference[5], 624, 980, and 537 probes in the Pidsley reference[29], 683, 1058, and 585 probes in the McCartney reference[30], and 1829, 2370, and 1659 probes in the Price reference[31], in the VHAS, CLHNS, and CALERIE cohorts, respectively).

High coverage of epigenetic clock, biomarker predictor, and cell type deconvolution CpGs was noted on EPICv2

To investigate the applicability of widely used DNA methylation-based tools, which were built on previous arrays/versions, on EPICv2, we noted the percentage of predictive CpGs which are absent in either of the EPIC versions and which are replicated on EPICv2. Across the seven epigenetic clocks investigated, we observed that ~77-96% predictive CpGs were retained on EPICv2, with a reintroduction of several predictive CpGs which were absent in EPICv1 on EPICv2 (Horvath pan-tissue: 14 CpGs; Hannum: 4 CpGs; and epigenetic timer of cancer (epiTOC): 25 CpGs). We noted that in all investigated clocks, except epiTOC, ~1-7% of the predictive clock CpGs have replicate probes on EPICv2. Similarly, across the four investigated biomarker predictors, we noted ~89-94% coverage of predictive CpGs on EPICv2, with 0.86-2.86% also including replicate probes. In the CRP predictor, we additionally observed that 5.84% predictive CpGs were absent in EPICv1,

221 with 83 of these CpGs reintroduced in EPICv2. Next, on examining the 1200 pre-selected extended
222 Identifying Optimal DNA methylation Libraries reference (IDOL) probes used in cell type
223 deconvolution[10], we noted that 99.33% were represented on EPICv2, with 4% also having
224 replicate probes on EPICv2 (Table 2). On estimating the correlation of β values of predictive CpGs
225 employed by these DNA methylation-based tools between EPICv1 and EPICv2, weighted by
226 corresponding coefficients, we identified high Spearman correlation ranging from 0.9615-0.9952
227 (Supplementary Figure 2). We also noted that average absolute differences in β values of these
228 predictive CpGs between the EPIC versions range from 0.0157-0.0342, with the Horvath pan-
229 tissue clock having the lowest average absolute difference (MAD) between the two versions
230 (Supplementary Table S2-S3).
231

232 **Immune cell type proportions inferred by IDOL and auto probe selection methods were**
233 **significantly different between EPICv1 and EPICv2**

234 Cellular composition is one of the key contributors to whole blood DNA methylation
235 variation[32,33], and is often included as a covariate in statistical models to account for cell type
236 heterogeneity. Besides that, cellular composition changes have been associated with phenotypes
237 of interest and may in fact be integral in the understanding of underlying biological processes[34–
238 36]. Most studies do not have actual cell counts and rely on predicted values from DNA
239 methylation-based algorithms. Given that these algorithms use references profiled on previous
240 arrays/versions, we tested whether there were differences in cell type proportion estimates obtained
241 from the two EPIC versions using one of the most commonly used cell type deconvolution methods
242 for whole blood[10]. We compared the proportions of twelve immune cell types, namely, basophils
243 (Bas), naïve and memory B cells (Bnv, Bmem), naïve and memory CD4+ T cells (CD4nv,
244 CD4mem), naïve and memory CD8+ T cells (CD8nv, CD8mem), eosinophils (Eos), monocytes
245 (Mono), neutrophils (Neu), natural killer (NK), and T regulatory cells (Treg), inferred from DNA
246 methylation measured on EPICv1 and EPICv2 using the estimatedCellCounts2 functions from
247 *FlowSorted.Blood.EPIC* R package with two commonly used probe selection methods: IDOL and
248 "auto". The IDOL method includes 1200 pre-selected probes, eight of which are absent in EPICv2,
249 and the "auto" method includes 1200 probes independently selected for each EPIC version as
250 described in Methods. On comparing the probes selected for each EPIC version using the "auto"
251 method, we identified a large overlap of 1090 probes (90.83%) between the two EPIC versions,
252 with only 121 EPICv1 (10.983%) and 134 EPICv2 (11.167%) auto-selected probes overlapping
253 with the 1200 pre-selected IDOL probes (Supplementary Figure 3).
254
255 Using the IDOL method, we found Pearson correlations between EPICv1 and EPICv2 to be greater
256 than 0.80 for all predicted cell types, except for Tregs which showed slightly lower correlation
257 (Pearson r = 0.628-0.715). The proportion of CD8nv was zero for all samples in VHAS and
258 CLHNS, but averaged at 0.012 (EPICv1) and 0.010 (EPICv2) in CALERIE. On comparing
259 inferred cell proportions between EPICv1 and EPICv2, we identified significantly higher
260 estimated proportions of Bas, Mono, and NK cells, and lower proportions of CD8mem and Eos in
261 EPICv2 compared to EPICv1, with Cohen's d effect size ranging from small to large depending
262 on the cell type. For all five cell types, MAD between EPIC versions was higher than technical
263 replicate differences within EPIC versions. CD4nv and CD8nv proportions were not significantly
264 different between EPIC versions in any of the investigated cohorts. As for the remaining five cell
265 types, namely Bmem, Bnv, CD4mem, Neu and Tregs, we observed significant proportion

differences between EPIC versions in at least one cohort, however, these findings were not consistent in all three cohorts (Figure 3, Supplementary Table S4).

Using the "auto" method, we found Pearson correlations between EPICv1 and EPICv2 to be greater than 0.80 for all predicted cell types, with the exception of CD4mem and CD4nv which showed only a modest Pearson *r* of 0.788 and 0.778 respectively in VHAS. For all EPICv1 samples, Treg proportions were estimated to be zero in all three cohorts, while CD8nv proportions were estimated to be zero for all samples in VHAS and CLHNS, and 87.5% in CALERIE. Paired *t*-tests showed significantly higher estimated proportions of Bas, Bmem, CD8mem, CD8nv, Mono, NK and Tregs, and lower proportions of CD4mem and CD4nv in EPICv2 compared to EPICv1. Meanwhile, we noted no significant differences in Eos between EPICv1 and EPICv2 across all three cohorts. For all cell types, the MAD between EPIC versions was higher than technical replicate differences within EPIC versions. Among the remaining cell types, we observed higher proportions higher proportions of Bnv (d = 0.13-0.22) and Neu (d = -0.06-0.08) in EPICv2 in CLHNS and CALERIE (Supplementary Figure 4, Supplementary Table S5).

Clock EAAs were significantly different between matched samples on EPICv1 and EPICv2 when samples were combined and EAAs were subsequently calculated

Epigenetic clocks are based on the property that DNA methylation levels at specific CpGs highly correlate with chronological age or age-related outcomes and are often estimated as a linear combination of weighted sum of the coefficients of predictive CpGs[11–14,16,17]. In order to evaluate the concordance between EPICv1 and EPICv2 in the context of DNA methylation-based tools, we compared estimates from widely used first- and second- generation epigenetic clocks, namely Horvath pan-tissue[11], Hannum[12], Horvath skin and blood clocks[13], PhenoAge[14], and GrimAge[15], and rate-based clocks, namely pace of aging calculated from the epigenome (DunedinPACE)[17], DNA methylation estimator of telomere length (DNAmTL)[16], and epiTOC[18] in the three cohorts VHAS, CLHNS, and CALERIE. We identified high Pearson correlations between chronological age and epigenetic age obtained from both EPICv1 (VHAS: Pearson *r* = 0.701-0.825, CLHNS: 0.819-0.912, and CALERIE: 0.853-0.951) and EPICv2 (VHAS: Pearson *r* = 0.583-0.818, CLHNS: 0.807-0.918, and CALERIE: 0.825-0.957) (Supplementary Table S6). Further, EPICv2 epigenetic ages for the first- and second-generation clocks were highly correlated with EPICv1 epigenetic ages (VHAS: Pearson *r* = 0.911-0.980, CLHNS: 0.879-0.990, and CALERIE: 0.915-0.996), with differences in epigenetic ages between technical replicate samples being 0.523-4.206 years and 0.546-2.994 years within EPICv1 and EPICv2 respectively in VHAS, 0.151-3.408 years within EPICv2 in CLHNS (no technical replicates within EPICv1 were available), and 0.13-2.25 years and 0.245-4.142 years within EPICv1 and EPICv2 respectively in CALERIE (Supplementary Table S2). In these three cohorts, we noted significant differences in epigenetic ages between EPICv2 and EPICv1 in all clocks, except for Horvath pan-tissue. We identified significantly lower estimated epigenetic ages for Hannum and Horvath skin and blood clocks and higher estimated epigenetic ages for PhenoAge in EPICv2 when compared to EPICv1, with mean absolute difference (MAD) greater than technical replicate differences. The Hannum clock showed the largest MAD and Cohen's d effect size between EPICv1 and EPICv2 in all three cohorts (VHAS: MAD=21.87, d = 3.40, CLHNS: MAD = 19.81, d = 4.14, and CALERIE: MAD = 21.71, d = 2.86) (Supplementary Table S2, Supplementary Figure 5A).

311 We also estimated epigenetic age acceleration (EAA), a measure of rate of aging which has been
312 associated with health outcomes, by considering samples profiled on each EPIC version separately,
313 as well as by combining matched samples profiled on both versions. When EAA was calculated
314 separately by versions (EPIC versions separate), we noted modest to high correlation of EAA
315 (VHAS: Pearson r = 0.867-0.947, CLHNS: 0.577-0.976, and CLAERIE: 0.622-0.961) and no
316 significant differences between EPICv1 and EPICv2 (Supplementary Table S7, Figure 4B,
317 Supplementary Figure 6B, Supplementary Figure 7B). When clock estimates from the EPIC
318 versions were combined prior to EAA calculation (EPIC versions combined), we observed similar
319 modest to high correlation (VHAS: Pearson r = 0.862-0.947, CLHNS: 0.574-0.958, and
320 CALERIE: 0.632-0.960), but with significant differences between EPICv1 and EPICv2 for all five
321 clock EAAs, with the exception of Horvath pan-tissue (Supplementary Table S7, Figure 4C,
322 Supplementary Figure 6C, Supplementary Figure 7C). Additionally, we identified significantly
323 lower EAAs for Hannum and Horvath skin and blood, and higher EAAs for PhenoAge in EPICv2
324 when compared to EPICv1 (Supplementary Table S7). Next, we tested whether the different EPIC
325 versions indeed contribute to observed EAA differences by first combining epigenetic ages in both
326 EPIC versions and then calculating EAAs by including version as a covariate in the linear
327 regression (EPIC versions combined and adjusted). After EPIC version adjustment, we noted
328 modest to high correlation (VHAS: Pearson r = 0.862-0.947, CLHNS: 0.574-0.958, and
329 CALERIE: 0.632-0.960) and no significant EAA differences between the two versions in the three
330 cohorts (Supplementary Table S7, Figure 4D, Supplementary Figure 6D, Supplementary Figure
331 7D). We repeated the same analyses using epigenetic clocks estimated based on principal
332 components (PC clocks) rather than individual predictive CpGs, as they have been shown to
333 overcome unreliability in clock estimates due to technical noise[37]. We similarly noted that there
334 were no significant EAA differences when EAA was calculated for each EPIC version separately.
335 When EAAs were calculated on combined sets of matched samples, we noted significant
336 differences in PCHorvath skin and blood and PCPhenoAge clocks in the three adult cohorts, and
337 these differences were no longer significant with EPIC version adjustment (Supplementary Figures
338 8-10).
339
340 Given the discrepancies known to exist in epigenetic clock estimations using different
341 normalization methods[23,38,39], we sought to examine whether our results, observed in data
342 normalized by functional normalization, remained consistent irrespective of the choice of
343 normalization method. To test this, we combined the common probes on the two EPIC versions
344 and applied Beta-MIxture Quantile (BMIQ) normalization, and subsequently calculated epigenetic
345 ages and EAAs. We still noted significant differences in epigenetic age and EAA between EPICv1
346 and EPICv2 in Horvath pan-tissue, Horvath skin and blood, PhenoAge, and GrimAge clocks in at
347 least two out of the three cohorts when the two EPIC versions were combined and analyzed
348 (Supplementary Figure 5B and 11C, 12C, and 13C). No significant differences in EAA between
349 the versions were observed when EAAs were calculated separately for EPICv1 and EPICv2 or
350 when the versions were combined and EAAs were adjusted for EPIC version (Supplementary
351 Figure 11-13). Given the contribution of EPIC version in explaining DNA methylation variation,
352 we also examined if applying version correction using batch correction algorithms can circumvent
353 version-specific effects and eliminate the observed differences in DNA methylation-based
354 estimates between EPICv1 and EPICv2. We tested this by combining common probes on EPICv1
355 and EPICv2 after functional normalization, and sequentially correcting for EPIC version, chip and
356 sample position on chip (row) effects as applicable using the ComBat function implemented in the

357 *sva* R package[40]. As expected, when epigenetic ages were calculated using version-corrected
358 input β values, there were no significant differences between the EPIC versions (Supplementary
359 Figure 5C).
360

361 On evaluating the DunedinPACE, DNAmTL, and epiTOC rate-based clocks, we observed a high
362 correlation between EPICv1 and EPICv2 when the two were considered separately, as well as
363 when they were combined and adjusted for EPIC versions (VHAS: Pearson r = 0.936-0.983,
364 CLHNS: 0.838-0.967, and CALERIE: 0.808-0.941) (Supplementary Table S8). When we
365 examined the EPIC versions separately, we noted that DNAmTL estimates were significantly
366 higher in EPICv2 when compared to EPICv1 in all three cohorts and the differences were higher
367 than technical replicate differences within EPIC versions (VHAS: d = 0.96, CLHNS: d = 1.50, and
368 CALERIE: d = 0.65). DunedinPACE estimates were significantly higher in EPICv2 in VHAS (d
369 = 0.47) and CALERIE (d = 0.65), and epiTOC estimates were significantly lower in EPICv2 in
370 CLHNS (d = 0.43) but higher in CALERIE (d = 0.60) (Supplementary Table S8, Figure 5A,
371 Supplementary Figure 6E, and Supplementary Figure 7E). As expected, we observed no significant
372 differences between EPICv1 and EPICv2 when the rate-based clock estimates were adjusted for
373 EPIC version (Supplementary Table S8, Figure 5B, Supplementary Figure 6F, and Supplementary
374 Figure 7F). We observed similar differences when BMIQ normalized data was used as input for
375 rate-based clocks. When EPICv1 and EPICv2 were considered separately, there were significant
376 differences in DunedinPACE and DNAmTL estimates between versions in all three cohorts, and
377 significant differences in epiTOC estimates in CLHNS and CALERIE, and these differences were
378 no longer significant when EPIC version was accounted for as a covariate (Supplementary Figure
379 14) or when input β values were adjusted using version correction algorithms upstream
380 (Supplementary Figure 15A). We also repeated the same analyses using matched capillary blood
381 samples available in the VHAS cohort, as both venous and capillary blood sampling are common
382 in epigenetic research[41]. Consistent with our findings in venous blood samples, we observed
383 significant differences between EPIC versions in epigenetic ages as well as EAAs when EPIC
384 versions were combined and EAAs were calculated in four of the investigated first- and second-
385 generation clocks, two PC clocks, and all three rate-based clocks. When EAAs were calculated
386 separately for each version and when EAAs and rate-based clock estimates were adjusted for EPIC
387 versions, there were no significant differences between the EPIC versions (Supplementary Figures
388 16-19).

389 **Estimated CRP and smoking scores were significantly different between EPICv1 and**
390 **EPICv2**

391 DNA methylation-based estimates of IL-6, CRP, smoking, and alcohol use are widely used as
392 reliable proxies of actual measurements in epigenetic studies[39], and we tested if there were any
393 significant differences in these estimates between the two EPIC versions. We observed high
394 correlation of IL-6 (VHAS: Pearson r = 0.902, CLHNS: 0.912, and CALERIE: 0.85), CRP
395 (VHAS: Pearson r = 0.993, CLHNS: 0.992, and CALERIE: 0.944), and smoking scores (VHAS:
396 Pearson r = 0.981, CLHNS: 0.99, and CALERIE: 0.72), and modest correlation of alcohol scores
397 (VHAS: Pearson r = 0.789, CLHNS: 0.76, and CALERIE: 0.75) between EPICv1 and EPICv2.
398 When estimates were calculated separately by versions (EPIC versions separate), we identified
399 significantly higher CRP scores in EPICv2 compared to EPICv1 in VHAS (d = 0.26) and CLHNS
400 (d = 0.27), with differences higher than technical replicate differences, while there were no

401 significant differences in IL-6 and alcohol scores in either cohort. We noted significantly higher
402 IL-6 score and alcohol score in EPICv2 compared to EPICv1 in CALERIE (d = 0.49 and 1.28,
403 respectively), and significantly lower smoking scores in CLHNS (d = 0.30). When estimates from
404 the EPIC versions were first combined and then adjusted for version, as expected, we noted no
405 significant differences between EPICv1 and EPICv2 in any of the DNA methylation-based
406 predictors tested in the three cohorts (Supplementary Table S3, Supplementary Figure 20). Our
407 observation of significant EPIC version differences in estimates remained consistent when BMIQ
408 normalized β was used as input (Supplementary Figure 21), and these differences were no longer
409 significant when EPIC version was adjusted for as a covariate during estimate calculation or when
410 version-correction was applied upstream (Supplementary Figure 15B).

411 **Discussion**

412 The newly developed Illumina EPICv2 microarray provides wider coverage of the DNA
413 methylome compared to its predecessor EPICv1, with some reported differences in overall probe
414 content. While good correlations between the two EPIC versions have been previously reported
415 [5–7], a recent study demonstrated that differing probe content between versions influences
416 epigenetic age estimates, although this was not evaluated in matched samples profiled on EPICv1
417 and EPICv2[42]. To that end, we measured DNA methylation on EPICv1 and EPICv2 in matched
418 venous blood samples from adults across three geographically diverse populations, and
419 comprehensively examined the concordance between versions in a wide range of DNA
420 methylation-based tools.
421
422 We performed unsupervised hierarchical clustering on all samples across the three cohorts using
423 the shared probes between EPICv1 and EPICv2 and observed that samples clustered based on the
424 EPIC version they were measured on, despite cohort-specific characteristics. While examining
425 subsets of probes employed by the different DNA methylation-based tools, we noted a similar
426 clustering pattern indicating EPIC version as the predominant driver of DNA methylation
427 variation, followed by biological characteristics such as sex and age, which are known to influence
428 DNA methylation. Further, when we combined matched EPICv1 and EPICv2 samples in each
429 cohort and calculated cell type proportions, epigenetic ages, and biomarker scores, we observed
430 significant systematic differences in the majority of these estimated measures between the EPIC
431 versions, although some of these differences and their corresponding effect sizes were cohort-
432 specific. These differences were no longer significant when EPIC version was adjusted for during
433 the calculation of estimates, and in the case of epigenetic clocks specifically, when EAAs were
434 calculated in an EPIC version-specific manner.
435
436 On comparing the two EPIC versions, we observed overall high correlations at the array level and
437 low correlation at the individual probe level, in agreement with previous findings[5–7]. Only ~25%
438 of the shared probes between EPICv1 and EPICv2 showed modest-to-high probe level correlation,
439 with an additional ~67% showing low correlation explained by low DNA methylation variance.
440 The majority of these probes have also been reported previously as platform-bias free and high-
441 confidence mapping in cell lines[28], indicating their reliability between the two EPIC versions
442 and potentially across different sample types. In the remaining ~4% of probes, the vast majority
443 had notable differences in β values between EPICv2 and EPICv1, suggesting that these observed
444 differences may stem from EPIC version beyond technical noise, consistent with previous findings
445 comparing 450K and EPICv1[22,43,44]. While non-specific probe hybridization to the genome,

446 underlying genetic variation, or technical differences such as altered EPICv2 probe design may
447 explain the low concordance between the two EPIC versions[5,6,29–31], such probes accounted
448 for only a small proportion. We therefore recommend cautious interpretation of DNA methylation
449 readouts from these probes owing to their differences between the versions, specifically when
450 integrating EPICv1 and EPICv2 data for analyses.
451
452 We observed generally high correlation of cell type proportion estimates between EPICv1 and
453 EPICv2, however, significant proportion differences in specific immune cell types were observed.
454 Using pre-selected IDOL and auto methods of probe selection, we found that the proportions of
455 five to nine cell types were significantly different between EPICv1 and EPICv2, with monocytes
456 and NK cells consistently identified in both methods. Within each probe selection method, our
457 observations were generally consistent across all three investigated cohorts indicating systematic
458 differences between the two EPIC versions that are not driven by cohort-specific characteristics.
459 Given that EPICv2 cell type proportions were predicted using the reference library created by
460 profiling purified immune cell types from peripheral blood on EPICv1[10], and similar differences
461 in predicted cell type proportions have been observed previously between 450K and EPICv1[22],
462 the observed differences are likely due to version incompatibility between the reference library
463 used and input samples for EPICv2. These effects may be compounded by the proportional nature
464 of cell type estimates, implying that any changes in the proportion of even one cell type will
465 influence the proportions of other cell types. Together, our findings indicate inherent EPIC version
466 differences influencing cell type proportion estimations, and the need to investigate whether EPIC
467 version-specific cell type reference libraries would enhance the accuracy of cell type
468 deconvolution in EPICv2, and whether the detected EPIC version differences in cell type
469 proportions would affect downstream analyses such as epigenome wide association studies.
470
471 Our epigenetic clock analyses showed significant differences in estimates of epigenetic age
472 between EPICv1 and EPICv2 in five investigated clocks despite the high correlation of shared
473 clock CpGs, suggesting systematic differences in epigenetic ages depending on which EPIC
474 version the samples were profiled on. Given that there are varying clock CpGs absent on either
475 EPIC version, it is possible that the subsequent imputation performed by the clock algorithms
476 could influence obtained epigenetic ages and explain some of the observed variation. While using
477 common probes between the two versions as input for clock calculation, however, the variation in
478 epigenetic age estimates was still observed, suggesting that the proportion of missing clock CpGs
479 does not account for the EPIC version differences. This is in line with a previous study that missing
480 values do not compromise clock estimates in first-generation clocks[23]. Although in our analyses
481 the differences in Horvath pan-tissue epigenetic ages between EPICv1 and EPICv2 did not reach
482 statistical significance with functional normalization, these differences were significant when we
483 applied BMIQ normalization, and in both normalization methods, epigenetic ages were higher in
484 EPICv2 compared to EPICv1. Our Horvath pan-tissue clock findings therefore affirm previous
485 reports that choice of normalization method influences clock estimates[23,38,39]. Nevertheless,
486 we still observe significant epigenetic age differences in all other investigated first- and second-
487 generation clocks in both functional and BMIQ normalization, highlighting the contribution of
488 systematic EPIC version differences to epigenetic ages irrespective of the normalization method
489 used. It has been noted that the stochastic changes to single-CpG DNA methylation level as a
490 function of aging explain a non-negligible amount of variation in established epigenetic
491 clocks[45]. However, the individual-level stochastic variation in epigenetic ages is not the focus

of our analyses and cannot be measured using the methods employed in this study, and is likely to be minimal relative to cohort and other batch effects.

We also tested whether there were EPIC version differences in EAAs, which are more robust to data preprocessing methods compared to epigenetic ages[23,38,39]. We noted no EAA differences in the investigated first- and second-generation clocks when EAAs were calculated separately for the two EPIC versions, regardless of observed systematic shifts in epigenetic ages, lending additional support to the robustness of EAA as an epigenetic age measure when calculating in a version-specific manner. On the other hand, with EAA calculation being relative to input samples and observed epigenetic age differences between EPICv1 and EPICv2, we expectedly observed EAA differences when samples from both EPIC versions were combined and then EAA was calculated. These differences were no longer significant when we calculated EAA by regressing epigenetic age on chronological age and additionally adjusting for EPIC version. In contrast, calculating epigenetic age differences, by subtracting chronological age from epigenetic age, would yield significant differences between EPICv1 and EPICv2 due to the observed systematic EPIC version differences in epigenetic ages, once again highlighting the appropriateness of EAA in epigenetic age analyses. Our findings of significant differences between EPIC versions hold true in rate-based clock estimates and biomarker scores as well, and these differences were no longer significant with EPIC version adjustment, thus emphasizing correcting for EPIC version in these tools as an effective strategy to obtain comparable estimates.

Overall, our results indicate that epigenetic age differences can be ascribed to EPIC version, which will influence interpretation of findings if not accounted for. While adding EPIC version as a covariate in statistical models may be an option to handle version-specific discrepancies, this approach is not as straightforward in cases when EPIC version is confounded with the variable of interest. Alternatively, as noted in our investigations, version correction by ComBat may also circumvent observed version differences. However, employing such tools when study designs are unbalanced can introduce false biological signals, and is not appropriate when biological variables are confounded with EPIC versions[46–49]. Although the impact of applying batch correction on DNA methylation-based estimates is not well studied, it is known that data normalization methods can introduce variation in epigenetic clock estimates[23,39]. Therefore, it is reasonable to speculate that implementing additional data preprocessing steps such as version correction will result in biased estimates, specifically when biological variables are not evenly distributed across the two versions, highlighting an area that warrants further investigation.

Our study improves the current understanding of the applicability of DNA methylation-based tools for EPICv1 and EPICv2, however, there are several limitations. While our analyses may be limited by the modest sample sizes, our study includes matched samples collected from adults with wide age range and across diverse geographical regions. We also recognize that the CLHNS cohort contains female only, however, the consistency of our findings across these demographically heterogeneous cohorts provides support that the identified differences were not driven by any cohort-specific characteristics. We also acknowledge that the time between EPICv1 and EPICv2 measurement varied by approximately one month for VHAS and two years in CLHNS. While the time between array quantification was not consistent, it offers a more realistic representation of research settings where samples are collected and quantified across multiple batches on different EPIC versions. Next, given that the training datasets of the investigated epigenetic clocks were

538 predominantly composed of European populations, the suitability of investigated clocks may not
539 be well established in our investigated Asian and African cohorts[50–52]. However, the
540 differences in epigenetic clock estimates we observed in the VHAS and CLHNS cohorts was
541 similar to our observations in the CALERIE cohort comprising individuals of European and
542 African descent, indicating that our findings were not entirely driven by population-specific
543 effects. Additionally, any population-specific effects which may influence obtained estimates are
544 likely to influence both EPIC versions in a comparable manner. In our analyses, we observed high
545 correlations between clock estimates and chronological age and performed all comparisons
546 between EPICv1 and EPICv2 in a cohort-specific manner, thereby circumventing any potential
547 population-specific effects on calculated clock estimates. Further, while we compared EPICv2 cell
548 type proportions and biomarker scores to EPICv1, we recognize that actual cell counts and
549 biomarker measurements would be more appropriate as the ground truth and would be useful in
550 validating our DNA methylation-based estimates. However, the primary aim of the current work
551 is to investigate the concordance in estimates obtained from EPICv2 and EPICv1. Furthermore,
552 high concordance between actual cell counts or biomarker scores and bioinformatically predicted
553 estimates for EPICv1 has been previously reported[10,19–21], therefore it may be reasonable to
554 rely on these DNA methylation-based estimates. We are also cognizant of the influence of
555 preprocessing methods including normalization and batch correction options on DNA
556 methylation-based estimates[23,38,39]. We have addressed this to a certain degree by employing
557 two different normalization methods, namely functional normalization and BMIQ, as input in our
558 analyses. Our findings in the majority of the tested epigenetic clocks and biomarker predictors
559 remained significant with BMIQ normalization, implying that EPIC version effects still persisted
560 across the two investigated normalization methods. While applying version correction resolved
561 some of these EPIC version differences, it is plausible that currently available batch correction
562 algorithms can result in biased estimates as previously mentioned. Future work to investigate or
563 develop other preprocessing methods could be helpful in harmonizing EPICv1 and EPCIv2 and
564 potentially resolving observed EPIC version differences.

565 **Conclusions**

566 With the rapid generation of DNA methylation data profiled on the newer iterations of the Illumina
567 microarrays, such as EPICv2 and Infinium Methylation Screening Array (MSA), integrating
568 samples across these platforms poses a challenge, owing to discrepancies in probe content among
569 arrays/versions. Our findings highlight significant differences in the new EPICv2 compared to
570 EPICv1, demonstrate the influence of EPIC version on DNA methylation-based tools, and provide
571 recommendations to minimize technical variation which may arise from inconsistencies between
572 versions. We therefore encourage caution when harmonizing and interpreting DNA methylation
573 data when combining multiple arrays/versions to ensure reliability and reproducibility in
574 epigenetics analyses.

575 **Methods**

576 **Description of cohorts**

577 To compare the performance of the two most recent generations of MethylationEPIC BeadChip
578 Infinium microarrays in the context of DNA methylation-based clocks, biomarkers and cell type
579 proportion estimates, we measured the DNA methylomes of a subset of venous whole blood

samples on both EPICv1 and EPICv2 selected from three cohorts: (i) Vietnam Health and Aging Study (VHAS)[25], (ii) Cebu Longitudinal Health and Nutrition Survey (CLHNS)[26], and (iii) Comprehensive Assessment of Long-term Effects of Reducing Intake of Energy (CALERIE)[27]. The VHAS cohort additionally includes matched capillary blood samples, randomized using the same array design on both EPIC versions, such that the concordance between EPICv1 and EPICv2 can be assessed in capillary blood as well (n = 24 × 2 blood collection methods × 2 versions). The time between EPICv1 and EPICv2 array quantification was approximately one month and two years for VHAS and CLHNS respectively. EPICv1 and EPICv2 array quantification was carried out at the same time for CALERIE. Given that the demographic characteristics across the cohorts are different (Table 1), we performed all analyses independently on the cohorts and reported the findings in a cohort-specific manner, excluding the unsupervised hierarchical clustering analyses as described below.

DNA methylation profiling, sample and probe quality control

Using similar protocols for all three cohorts, DNA was extracted from samples, bisulfite converted using EZ-96 DNA Methylation kits (Zymo Research, Irvine, CA), hybridized to the EPICv1 and EPICv2 arrays, and scanned with the Illumina iScan 2000 to obtain IDAT files that capture raw DNA methylation intensities. IDATs were read using *minfi* R package to obtain β values that represent DNA methylation intensities for each CpG site ranging from 0 (fully unmethylated) to 1 (fully methylated). Technical replicates derived from the same sample were quantified to monitor technical variation within each EPIC version, independently for each cohort. Sample quality control checks were performed as described in previous publications[34,53]. Blood samples collected in the three cohorts passed all 17 Illumina quality control metrics in the *ewastools* R package[54,55], and detection *p*-value, beadcount, average methylated and unmethylated intensity metrics in the *minfi* R package (Supplementary Table S9). We also performed sample identity checks using unsupervised hierarchical clustering analysis on the 57 SNP probes that are common to both EPIC versions, and noted that matched samples from the same individual assessed on EPICv1 and EPICv2 grouped together.

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

DNA methylation data preprocessing and replicate probe analyses

To account for color and probe-type bias, we performed functional normalization (funnorm) with background correction and dye-bias normalization (noob) in the *minfi* R package[56,57] independently on EPICv1 and EPICv2 samples for each cohort. Cohort-specific technical replicate sample correlations were used to monitor preprocessing. Between technical replicate samples on the same EPIC version, improved Spearman correlations of whole array β values and reduced root mean square error (RMSE) were observed as processing progressed from raw (VHAS: Spearman *rho* = 0.9832, RMSE = 0.0306; CLHNS: 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).

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, and CALERIE, and subsequently applying Beta-MIxture Quantile (BMIQ) normalization implemented in the *wateRmelon* R package[58]. Due to their type I and type II design switch between EPICv1 and EPICv2[6], 82 probes were removed prior to BMIQ normalization. To account for any systematic bias in DNA methylation measurements[40] 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[40]. We applied Spearman correlation to examine concordance of probes between the two EPIC versions, while in all other analyses we applied Pearson correlation to evaluate linear relationships between variables.

Given that there are certain probes on EPICv2 having two or more replicates (replicate probes), we characterized their distribution across the genome, and compared three strategies to collapse them into a single β value (based on detection *p*-value, mean and median). Our analyses identified that collapsed β values of EPICv2 replicate probes obtained using all three methods were highly correlated to corresponding EPICv1 probes, therefore EPICv2 replicate probes with lowest detection *p*-value were chosen as the representative probe based on previous recommendation[6] (Supplementary Material), and this approach was used for all the reported subsequent analyses.

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

Cellular composition in heterogeneous tissue such as whole blood is one of the key contributors to the variation in DNA methylation profiles of bulk tissue[32,33]. In the absence of complete cell count data for the study samples, we estimated proportions of twelve immune cell types, basophils (Bas), naïve and memory B cells (Bnv, Bmem), naïve and memory CD4+ T cells (CD4nv, CD4mem), naïve and memory CD8+ T cells (CD8nv, CD8mem), eosinophils (Eos), monocytes (Mono), neutrophils (Neu), natural killer (NK), and T regulatory cells (Treg) from matched venous blood samples measured on the two EPIC versions. We used two methods of probe selection to estimate these cell type proportions: (i) the extended Identifying Optimal DNA methylation Libraries reference (IDOL), with probes not represented on EPICv2 removed from the reference before cell type proportions estimation in EPICv2 samples, and (ii) the "auto" method which selects the top 100 probes with F-stat *p*-value < 1E-8 for each cell type with the greatest magnitude of methylation difference, both implemented in the *FlowSorted.Blood.EPIC* R package with noob corrected values as recommended[10].

Estimation of epigenetic age and epigenetic age acceleration

We compared the performance of eight commonly used epigenetic clocks between EPICv1 and EPICv2 in VHAS, CLHNS, and CALERIE. Epigenetic age of first generation clocks including Horvath pan-tissue[11], Hannum[12], and Horvath skin and blood clocks[13], and second generation clocks including PhenoAge[14], and GrimAge[15] were obtained from the online DNA Methylation Age Calculator (https://dnamage.genetics.ucla.edu/new). For these clocks, we calculated epigenetic age acceleration (EAA), a measure of rate of aging commonly used in epigenetic clock investigations, by employing three approaches independently for each cohort.

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

To evaluate if there are significant differences in principal components (PC) clock[37] estimates, which are more robust to technical noise, by EPIC version, we first estimated epigenetic ages by the R script provided (https://github.com/MorganLevineLab/PC-Clocks) and then calculated EAAs using the approaches mentioned above.

As for the rate-based clocks, DunedinPACE was calculated using the DunedinPACE R package[17], epiTOC was obtained using the *getEpiTOC* function in the *cgageR* R package[18], and DNA methylation-based estimator of telomere length (DNAmTL)[16] were obtained from the online DNA Methylation Age Calculator (https://dnamage.genetics.ucla.edu/new). Unlike EAA calculation, there were no secondary measures calculated from these rate-based estimates in our analyses, estimates remained the same when calculated by the EPIC versions separate or combined approach, therefore, we employed only two approaches for each cohort:

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

Using paired *t*-tests and applying a Bonferroni multiple test correction, we evaluated differences in epigenetic clock estimations and EAAs (as well as biomarker predictor scores and cell type proportions described below) between matched samples assessed on EPICv1 and EPICv2 in a cohort-specific manner. Statistical significance was defined as Bonferroni adjusted p-value < 0.05. Effect sizes were measured by 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 recommended benchmarks [59].

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

707 Among other DNA methylation-based tools are inflammation, smoking and alcohol score
708 predictors, which provide biomarker measures that correlate with levels of inflammatory markers
709 (IL-6, CRP), smoking and alcohol use, respectively. DNA methylation-derived scores of IL-6,
710 CRP, smoking and alcohol use were calculated as a weighted sum of coefficients derived from
711 published lists of predictive CpGs[19–21]. Being that these biomarkers were trained on previous
712 Illumina arrays, we sought out to determine the correlation of derived DNA methylation-based
713 scores between EPICv1 and EPICv2, and compare these scores without version adjustment (EPIC
714 version separate) and with version adjustment (EPIC versions combined and adjusted) using
715 approaches similar to the rate-based epigenetic clocks.

716 **References**

717

718 1. Battram T, Yousefi P, Crawford G, Prince C, Sheikhali Babaei M, Sharp G, et al. The EWAS
719 Catalog: a database of epigenome-wide association studies. Wellcome Open Res. 2022;7:41.

720 2. Kuzawa CW, Ryan CP, Adair LS, Lee NR, Carba DB, MacIsaac JL, et al. Birth weight and
721 maternal energy status during pregnancy as predictors of epigenetic age acceleration in young
722 adults from metropolitan Cebu, Philippines. Epigenetics. 2022;17:1535–45.

723 3. Li M, Zou D, Li Z, Gao R, Sang J, Zhang Y, et al. EWAS Atlas: a curated knowledgebase of
724 epigenome-wide association studies. Nucleic Acids Res. 2019;47:D983–8.

725 4. Infinium MethylationEPIC v2.0 BeadChip. Infinium MethylationEPIC v2.0 BeadChip. 2023;

726 5. Peters TJ, Meyer B, Ryan L, Achinger-Kawecka J, Song J, Campbell EM, et al. Characterisation
727 and reproducibility of the HumanMethylationEPIC v2.0 BeadChip for DNA methylation profiling.
728 BMC Genomics. 2024;25:251.

729 6. Kaur D, Lee SM, Goldberg D, Spix NJ, Hinoue T, Li H-T, et al. Comprehensive evaluation of
730 the Infinium human MethylationEPIC v2 BeadChip. Epigenetics Commun. 2023;3:6.

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

734 8. Houseman EA, Molitor J, Marsit CJ. Reference-free cell mixture adjustments in analysis of
735 DNA methylation data. Bioinforma Oxf Engl. 2014;

736 9. Salas LA, Koestler DC, Butler RA, Hansen HM, Wiencke JK, Kelsey KT, et al. An optimized
737 library for reference-based deconvolution of whole-blood biospecimens assayed using the Illumina
738 HumanMethylationEPIC BeadArray. Genome Biol. 2018;19:64.

739 10. Salas LA, Zhang Z, Koestler DC, Butler RA, Hansen HM, Molinaro AM, et al. Enhanced cell
740 deconvolution of peripheral blood using DNA methylation for high-resolution immune profiling.
741 Nat Commun. 2022;13:761.

742 11. Horvath S. DNA methylation age of human tissues and cell types. Genome Biol.
743 2013;14:R115.

744 12. Hannum G, Guinney J, Zhao L, Zhang L, Hughes G, Sadda S, et al. Genome-wide Methylation
745 Profiles Reveal Quantitative Views of Human Aging Rates. Mol Cell. 2013;49:359–67.

746 13. Horvath S, Oshima J, Martin GM, Lu AT, Quach A, Cohen H, et al. Epigenetic clock for skin
747 and blood cells applied to Hutchinson Gilford Progeria Syndrome and *ex vivo* studies. Aging.
748 2018;10:1758–75.

749 14. Levine ME, Lu AT, Quach A, Chen BH, Assimes TL, Bandinelli S, et al. An epigenetic
750 biomarker of aging for lifespan and healthspan. Aging. 2018;10:573–91.

751 15. Lu AT, Quach A, Wilson JG, Reiner AP, Aviv A, Raj K, et al. DNA methylation GrimAge
752 strongly predicts lifespan and healthspan. Aging. 2019;11:303–27.

753 16. Lu AT, Seeboth A, Tsai P-C, Sun D, Quach A, Reiner AP, et al. DNA methylation-based
754 estimator of telomere length. Aging. 2019;11:5895–923.

755 17. Belsky DW, Caspi A, Corcoran DL, Sugden K, Poulton R, Arseneault L, et al. DunedinPACE,
756 a DNA methylation biomarker of the pace of aging. Deelen J, Tyler JK, Suderman M, Deelen J,
757 editors. eLife. 2022;11:e73420.

758 18. Yang Z, Wong A, Kuh D, Paul DS, Rakyan VK, Leslie RD, et al. Correlation of an epigenetic
759 mitotic clock with cancer risk. Genome Biol. 2016;17:205.

760 19. Stevenson AJ, Gadd DA, Hillary RF, McCartney DL, Campbell A, Walker RM, et al. Creating
761 and Validating a DNA Methylation-Based Proxy for Interleukin-6. J Gerontol A Biol Sci Med Sci.
762 2021;76:2284–92.

763 20. Wielscher M, Mandaviya PR, Kuehnel B, Joehanes R, Mustafa R, Robinson O, et al. DNA
764 methylation signature of chronic low-grade inflammation and its role in cardio-respiratory
765 diseases. Nat Commun. 2022;13:2408.

766 21. McCartney DL, Hillary RF, Stevenson AJ, Ritchie SJ, Walker RM, Zhang Q, et al. Epigenetic
767 prediction of complex traits and death. Genome Biol. 2018;19:136.

768 22. Solomon O, MacIsaac J, Quach H, Tindula G, Kobor MS, Huen K, et al. Comparison of DNA
769 methylation measured by Illumina 450K and EPIC BeadChips in blood of newborns and 14-year-
770 old children. Epigenetics. 2018;13:655–64.

771 23. McEwen LM, Jones MJ, Lin DTS, Edgar RD, Husquin LT, MacIsaac JL, et al. Systematic
772 evaluation of DNA methylation age estimation with common preprocessing methods and the
773 Infinium MethylationEPIC BeadChip array. Clin Epigenetics. 2018;10:123.

774 24. Dhingra R, Kwee LC, Diaz-Sanchez D, Devlin RB, Cascio W, Hauser ER, et al. Evaluating
775 DNA methylation age on the Illumina MethylationEPIC Bead Chip. PLOS ONE.
776 2019;14:e0207834.

25. Korinek K, Teerawichitchainan B, Zimmer Z, Brindle E, Nguyen TKC, Nguyen HM, et al. Design and measurement in a study of war exposure, health, and aging: protocol for the Vietnam health and aging study. BMC Public Health. 2019;19:1351.

26. Adair LS, Popkin BM, Akin JS, Guilkey DK, Gultiano S, Borja J, et al. Cohort Profile: The Cebu Longitudinal Health and Nutrition Survey. Int J Epidemiol. 2011;40:619–25.

27. Rochon J, Bales CW, Ravussin E, Redman LM, Holloszy JO, Racette SB, et al. Design and conduct of the CALERIE study: comprehensive assessment of the long-term effects of reducing intake of energy. J Gerontol A Biol Sci Med Sci. 2011;66:97–108.

28. Chen BH, Zhou W. mLiftOver: Harmonizing Data Across Infinium DNA Methylation Platforms [Internet]. bioRxiv; 2024 [cited 2024 Apr 8]. p. 2024.03.18.585415. Available from: https://www.biorxiv.org/content/10.1101/2024.03.18.585415v1

29. Pidsley R, Zotenko E, Peters TJ, Lawrence MG, Risbridger GP, Molloy P, et al. Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. Genome Biol. 2016;17:208.

30. McCartney DL, Walker RM, Morris SW, McIntosh AM, Porteous DJ, Evans KL. Identification of polymorphic and off-target probe binding sites on the Illumina Infinium MethylationEPIC BeadChip. Genomics Data. 2016;9:22–4.

31. Price EM, Cotton AM, Lam LL, Farré P, Emberly E, Brown CJ, et al. Additional annotation enhances potential for biologically-relevant analysis of the Illumina Infinium HumanMethylation450 BeadChip array. Epigenetics Chromatin. 2013;6:4.

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

33. Zheng SC, Beck S, Jaffe AE, Koestler DC, Hansen KD, Houseman AE, et al. Correcting for cell-type heterogeneity in epigenome-wide association studies: revisiting previous analyses. Nat Methods. 2017;14:216–7.

34. Merrill SM, Gladish N, Fu MP, Moore SR, Konwar C, Giesbrecht GF, et al. Associations of peripheral blood DNA methylation and estimated monocyte proportion differences during infancy with toddler attachment style. Attach Hum Dev. 2023;25:132–61.

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

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

811 37. Higgins-Chen AT, Thrush KL, Wang Y, Minteer CJ, Kuo P-L, Wang M, et al. A computational
812 solution for bolstering reliability of epigenetic clocks: Implications for clinical trials and
813 longitudinal tracking. Nat Aging. 2022;2:644–61.

814 38. Ori APS, Lu AT, Horvath S, Ophoff RA. Significant variation in the performance of DNA
815 methylation predictors across data preprocessing and normalization strategies. Genome Biol.
816 2022;23:225.

817 39. Engelbrecht H-R, Merrill SM, Gladish N, MacIsaac JL, Lin DTS, Ecker S, et al. Sex
818 differences in epigenetic age in Mediterranean high longevity regions. Front Aging.
819 2022;3:1007098.

820 40. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch
821 effects and other unwanted variation in high-throughput experiments. Bioinforma Oxf Engl.
822 2012;28:882–3.

823 41. Hubens WHG, Maié T, Schnitker M, Bocova L, Puri D, Wessiepe M, et al. Targeted DNA
824 Methylation Analysis Facilitates Leukocyte Counts in Dried Blood Samples. Clin Chem.
825 2023;69:1283–94.

826 42. Garma LD, Quintela-Fandino M. Epigenetic age prediction drifts resulting from next-
827 generation methylation arrays [Internet]. 2024 [cited 2024 Jul 1]. Available from:
828 http://biorxiv.org/lookup/doi/10.1101/2024.06.07.597709

829 43. Sugden K, Hannon EJ, Arseneault L, Belsky DW, Corcoran DL, Fisher HL, et al. Patterns of
830 Reliability: Assessing the Reproducibility and Integrity of DNA Methylation Measurement.
831 Patterns. 2020;1:100014.

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

835 45. Meyer DH, Schumacher B. Aging clocks based on accumulating stochastic variation. Nat
836 Aging. 2024;1–15.

837 46. Price EM, Robinson WP. Adjusting for Batch Effects in DNA Methylation Microarray Data,
838 a Lesson Learned. Front Genet [Internet]. 2018 [cited 2021 Jun 22];9. Available from:
839 https://www.frontiersin.org/articles/10.3389/fgene.2018.00083/full

840 47. Buhule OD, Minster RL, Hawley NL, Medvedovic M, Sun G, Viali S, et al. Stratified
841 randomization controls better for batch effects in 450K methylation analysis: a cautionary tale.
842 Front Genet. 2014;5:354.

843 48. Nygaard V, Rødland EA, Hovig E. Methods that remove batch effects while retaining group
844 differences may lead to exaggerated confidence in downstream analyses. Biostat Oxf Engl.
845 2016;17:29–39.

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

849 50. Lin W-Y. Epigenetic clocks derived from western samples differentially reflect Taiwanese
850 health outcomes. Front Genet. 2023;14:1089819.

851 51. Horvath S, Gurven M, Levine ME, Trumble BC, Kaplan H, Allayee H, et al. An epigenetic
852 clock analysis of race/ethnicity, sex, and coronary heart disease. Genome Biol. 2016;17:171.

853 52. Cronjé HT, Nienaber-Rousseau C, Min JL, Green FR, Elliott HR, Pieters M. Comparison of
854 DNA methylation clocks in Black South African men. Epigenomics. 2021;13:437–49.

855 53. Konwar C, Asiimwe R, Inkster AM, Merrill SM, Negri GL, Aristizabal MJ, et al. Risk-focused
856 differences in molecular processes implicated in SARS-CoV-2 infection: corollaries in DNA
857 methylation and gene expression. Epigenetics Chromatin. 2021;14:54.

858 54. Murat K, Grüning B, Poterlowicz PW, Westgate G, Tobin DJ, Poterlowicz K. Ewastools:
859 Infinium Human Methylation BeadChip pipeline for population epigenetics integrated into
860 Galaxy. GigaScience. 2020;9:giaa049.

861 55. Illumina Inc. BeadArray Controls Reporter Software Guide. 2015;

862 56. Fortin J-P, Labbe A, Lemire M, Zanke BW, Hudson TJ, Fertig EJ, et al. Functional
863 normalization of 450k methylation array data improves replication in large cancer studies. Genome
864 Biol. 2014;15:503.

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

867 58. Teschendorff AE, Marabita F, Lechner M, Bartlett T, Tegner J, Gomez-Cabrero D, et al. A
868 beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium
869 450 k DNA methylation data. Bioinformatics. 2013;29:189–96.

870 59. Cohen J. Statistical Power Analysis for the Behavioral Sciences. 2nd ed. New York: Routledge;
871 1988.

872

873

874 **Acknowledgements**

875 We thank Dr. Sarah Merrill, Maggie Fu, and Dr. Kimberly Schmidt for their helpful feedback
876 during manuscript preparation. We would also like to thank all cohort participants for their
877 involvement and providing biological samples for research. This research utilized the
878 *FlowSorted.BloodExtended.EPIC* software packages developed at Dartmouth College, which are

879 governed by the licensing terms provided by Dartmouth Technology Transfer. D.W.B. and M.S.K.
880 are fellows of the CIFAR CBD Network.

881 **Funding**

882 This work received supported from the Canadian Institutes of Health Research (CIHR) / Project
883 Grant (Sponsor Reference Number: PJT-175309), National Institute on Aging of the National
884 Institutes of Health under Award Number R01AG052537, National Institutes of Health / Research
885 Project Grant (R01) (Sponsor Reference Number: 5R01AG061006-05), and US National Institute
886 on Aging Grants R01AG061378.
887

888 **Ethics approval and consent to participate**

889 This study was approved by the University of British Columbia Research Ethics Boards (H18-
890 03136) and the University of Utah Institutional Review Board (IRB 00098861).

891 **Competing interests**

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

895 **Authors' contributions**

896 Conceptualization: B.C.Z, M.S.J., C.K.; biological data collection/generation: T.K.T., K.K., Z.Z.,
897 K.M.H., D.W.B., T.W.M., C.W.K., M.S.K.; DNA methylation data generation: J.L.M., K.D.,
898 M.S.K.; DNA methylation data curation and processing: B.C.Z., M.S.J., H-R.E.; Phenotypic data
899 curation: T.K.T., K.K., Z.Z., K.M.H., D.W.B., T.W.M, C.W.K.; Data analysis: B.C.Z., M.S.J.;
900 Interpretation of results and manuscript writing: M.S.J., B.C.Z., C.K.; Figures/visualization:
901 B.C.Z., M.S.J.; Manuscript review: all authors; Cohort Owners/Custodians: T.K.T., K.K., Z.Z.,
902 K.M.H., D.W.B., T.W.M., C.W.K.
903
904
905

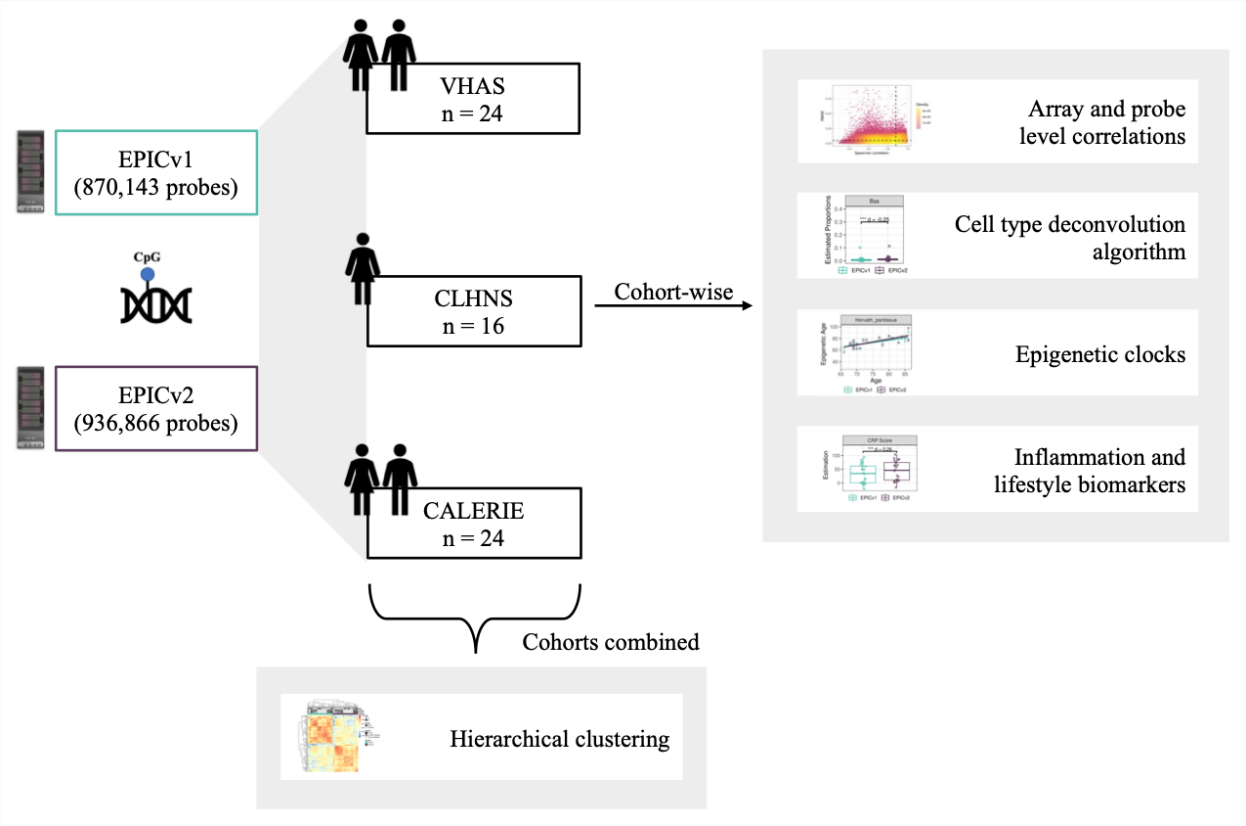

Figure 1. Overview of study design and analyses.

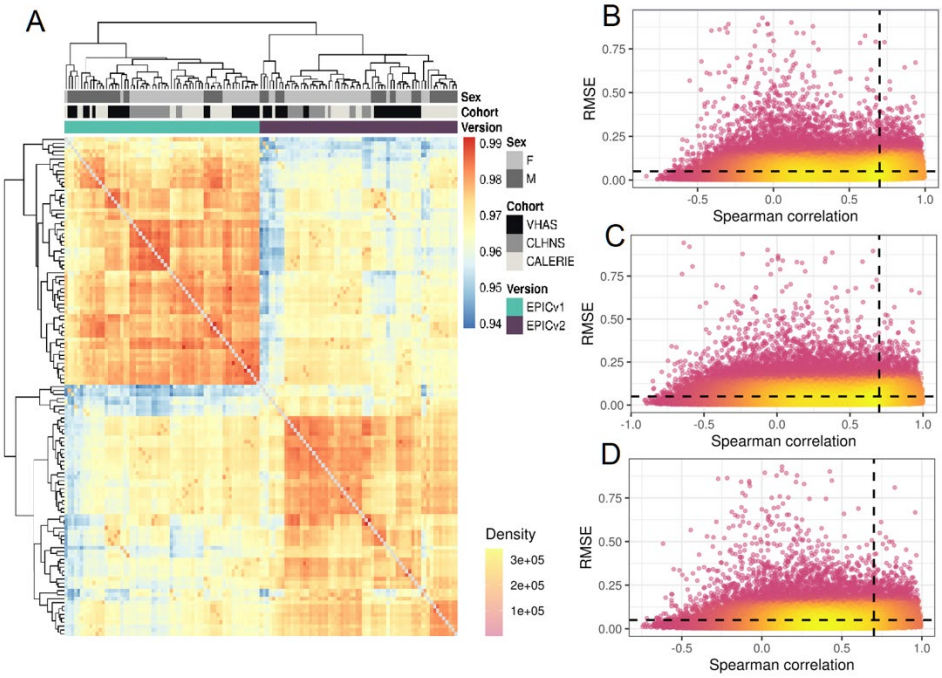

Figure 2. (A) Hierarchical clustering of 721,378 probes common between EPICv1 and EPICv2 of matched VHAS, CLHNS, and CALERIE samples; blue to red color range denotes Spearman correlation from 0.94-1.00. **(B-D)** Spearman correlation and root mean square error (RMSE) of 721,378 probes common between EPICv1 and EPICv2 investigated on **(B)** VHAS, **(C)** CLHNS, and **(D)** CALERIE. The X-axis represents the Spearman correlation and Y-axis represents the RMSE of probes common to both arrays. Dashed horizontal line indicates the RMSE threshold set at 0.05, and the vertical dashed line indicates the correlation threshold set at 0.70. The colors indicate the density of points, such that pink is low density and yellow is high density. Probes unique to either array are not shown.

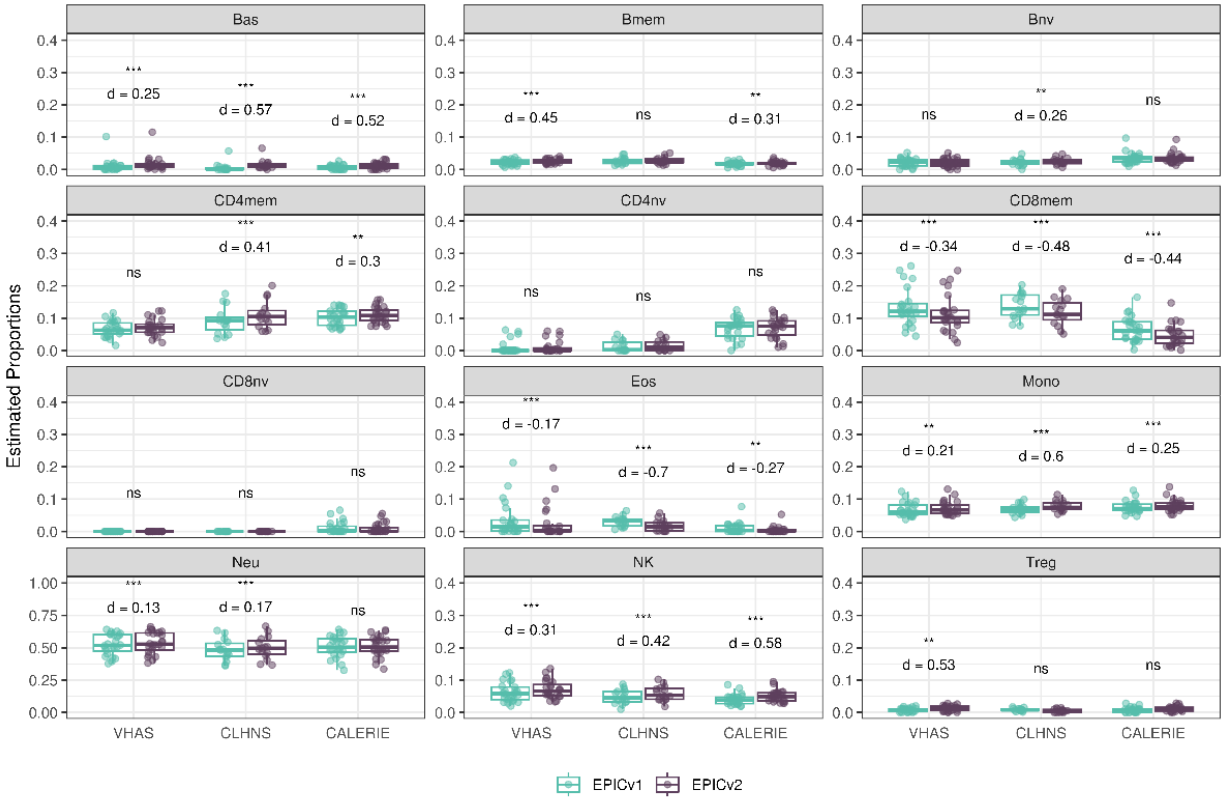

Figure 3. Differences in DNA methylation-based immune cell type proportions estimated using the IDOL reference on matched samples assessed on EPICv1 and EPICv2 in VHAS, CLHNS, and CALERIE. Statistical significance was defined as Bonferroni adjusted *p*-value <0.05. ** denotes Bonferroni *p* <0.05, *** 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.

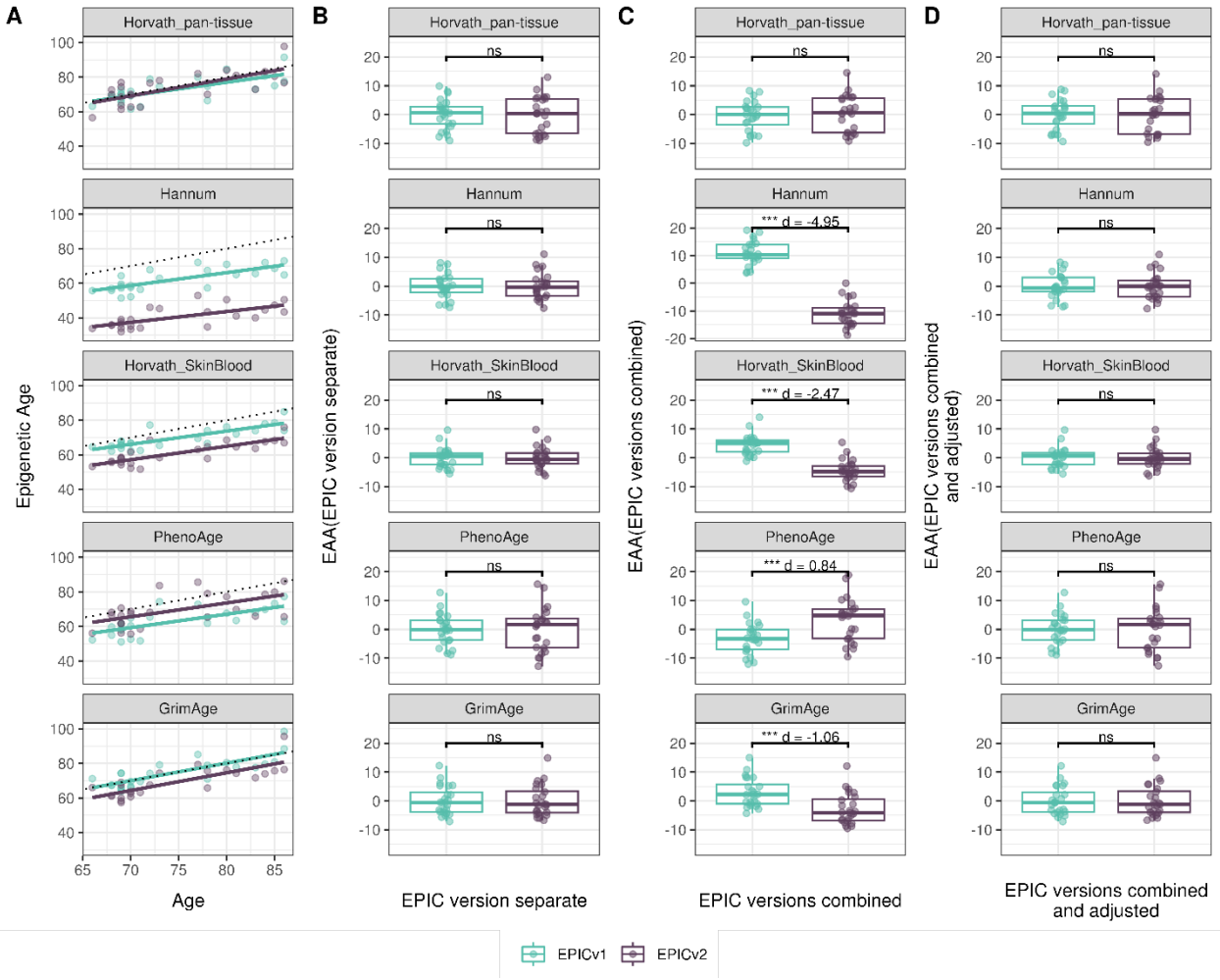

Figure 4. Epigenetic ages on matched samples assessed on EPICv1 and EPICv2 in VHAS. **(A)** Scatter plot of Horvath pan-tissue, Hannum, Horvath skin and blood, PhenoAge, and GrimAge clock ages (Y axis) and chronological age (X axis) with dotted line indicating x=y, coloured by EPIC version. **(B-D)** Boxplots comparing EPICv1 and EPICv2 EAAs calculated by considering (B) EPIC versions separately, (C) combined, and (D) combined and EPIC version adjusted. 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.

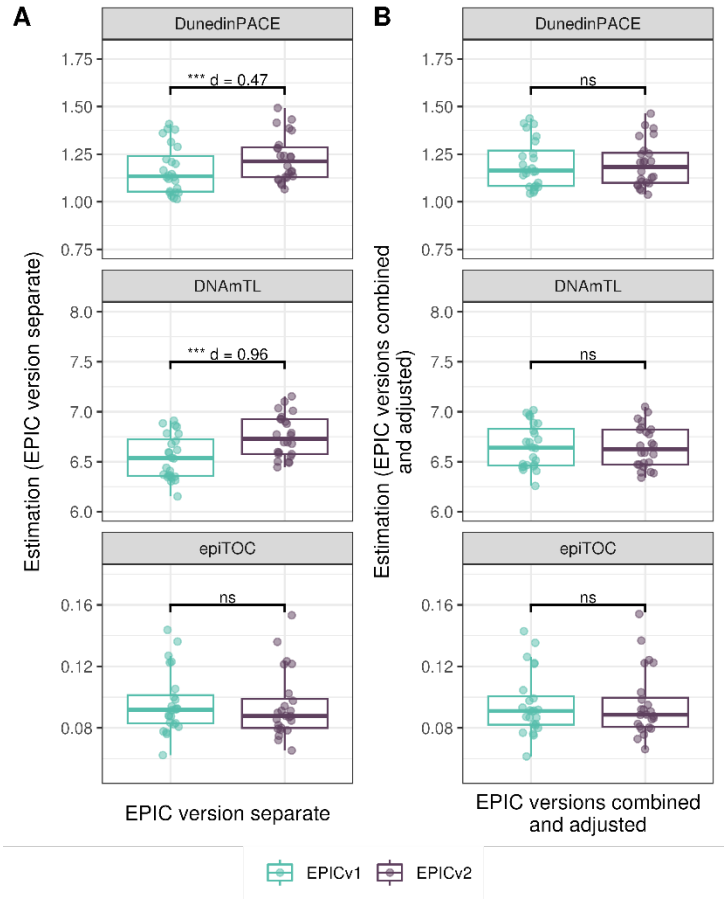

Figure 5. Rate-based clock estimates on matched samples assessed on EPICv1 and EPICv2 in VHAS. Boxplots comparing DunedinPACE, DNAmTL and epiTOC estimates calculated by considering EPIC versions separately **(A)** and combined and EPIC version adjusted **(B)**, between EPICv1 and EPICv2. Statistical 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.

Table 1. Demographics of the three investigated cohorts.

	VHAS	CLHNS	CALERIE
Samples (n)	48 (24 matched venous and capillary samples)	16	24
Country of origin	Vietnam	Philippines	United States of America
Biological sex (% Female)	36.4%	100%	54.2%
Age in years (mean (SD, range))	74.6 (6.7, 66-86)	48.4 (6.1, 40-63)	36.7(8.1, 21.3-51)

954 **Table 2.** Summary of predictive CpGs of DNA methylation-based clocks, biomarker predictors,
955 and cell type deconvolution in EPICv1 and EPICv2.

Predictors	**CpGs**	**Probes absent in EPICv1 (%)**	**Probes absent in EPICv2 (%)**	**Probes absent in EPICv1 and reintroduced in EPICv2**	**EPICv2 replicate probes**	**Training arrays; training age range (year)**	**Training sample size**	**Probes absent in Infinium Methylation Screening Array (MSA)**	**Probes absent in 450K**
Horvath pan-tissue*	353	17 (4.82%)	13 (3.68%)	14	4(1.13%)	27K and 450K Age: 0-100	7844	6(1.7%)	0
Hannum*	71	6 (8.45%)	7 (9.86%)	4	5(7.04%)	450K Age: 19-101	656	3(4.23%)	0
Horvath SkinBlood	391	0	17 (4.35%)	0	9(2.3%)	450K and EPICv1 Age: 0-94	896	3(0.77%)	0
PhenoAge	513	0	18 (3.51%)	0	7(1.36%)	27K, 450K, EPICv1 Age: 18-100	9926	3(0.58%)	0
GrimAge**	1030	NA	NA	0	NA	450K and EPICv1 Age: NA (mean 66)	1731	NA	NA
PC Clocks***	78464	0	5801 (7.39%)	0	650 (0.83%)	450K and EPICv1 Age: 0-101	NA	51108 (65.14%)	0

DunedinPACE	173	0	29 (16.76%)	0	2(1.16%)	EPICv1 Age 38 and 45	1037	2(1.16%)	0
DNAmTL	140	0	31 (22.14%)	0	3(2.14%)	450K and EPICv1 Age: 22-93	2256	2(1.43%)	0
epiTOC*	385	31 (8.05%)	26 (6.75%)	25	0	450K Age: 19-101	656	13(3.38%)	0
IL-6 score	35	0	3 (8.57%)	0	1(2.86%)	450K and EPICv1 Age: 67-78	875	12(34.29%)	0
CRP score*	1765	103 (5.84%)	96 (5.44%)	83	41 (2.32%)	27K, 450K and EPICv1 Age: 16-75	22774	357 (20.23%)	0
Smoking score	233	0	23 (9.87%)	0	2(0.86%)	EPICv1 Age: 18-99	5087	27(11.59%)	0
Alcohol score	450	0	49 (10.89%)	0	6(1.33%)	EPICv1 Age: 18-99	5087	47(10.44%)	0
IDOL	1200	0	8 (0.67%)	0	48(4%)	EPICv1 Age: 19-58	NA	20(1.67%)	741 (61.75%)

* CpGs commonly absent in both EPICv1 and EPICv2: Horvath pan-tissue: 3 CpGs; Hannum: 2 CpGs; EpiTOC: 6 CpGs; CRP: 20 CpGs. ** Clock CpGs for GrimAge are not publicly available. *** PC Clocks CpGs include the total CpGs required to calculate PC versions of Horvath pan-tissue, Hannum, Horvath SkinBlood, PhenoAge, and GrimAge.