Discrepancies in readouts between Infinium MethylationEPIC v2.0 and v1.0 reflected in DNA methylation-based tools: implications and considerations for human population epigenetic studies

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

† 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. 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. Department Genes and Environment, Max Planck Institute of Psychiatry, Munich, 80804, Germany
4. International Max Planck Research School for Translational Psychiatry (IMPRS-TP), Munich, 80804, Germany

5. Robert N. Butler Columbia Aging Center, Mailman School of Public Health, Columbia University, New York, NY 10032, USA
6. Harvard Medical School/ MIT Institute of Technology MD-PhD program, Boston, Massachusetts, MA 02115, USA
7. Family Medicine Department, Hanoi Medical University, Hanoi, Vietnam
8. Department of Sociology, University of Utah, Salt Lake City, Utah, UT 84112, USA
9. Department of Family Studies and Gerontology, Mount Saint Vincent University, Halifax, NS, B3M 2J6, Canada
10. Canada Research Chair, Global Aging and Community Initiative, Canada
11. USC-Office of Population Studies Foundation, Inc., University of San Carlos, Cebu City, Philippines
12. Department of Anthropology, Northwestern University, Evanston, Illinois, IL 60208 USA
13. Program in Child and Brain Development, CIFAR, Toronto, Ontario, Canada
14. Department of Anthropology and Institute for Policy Research, Northwestern University, Evanston, Illinois, IL 60208, USA
15. Duke University School of Medicine, Durham, NC, 27701, USA
16. Department of Epidemiology, Columbia University Mailman School of Public Health, New York, NY 10032, USA
17. Department of Statistics, Faculty of Science, University of British Columbia, Vancouver, BC, V6T 1Z4, Canada
18. Edwin S.H. Leong Centre for Healthy Aging and Department of Medical Genetics, University of British Columbia, Vancouver, BC, V6T 1Z4, Canada

Keywords

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. To address this, we profiled DNA methylation on both EPIC versions using matched venous blood samples from individuals spanning early to late adulthood across four cohorts.

Findings:

Within each cohort, samples primarily clustered by the EPIC version they were measured on. High concordance between EPIC versions at the array level, but variable concordance at the individual probe level was noted. Significant differences between versions in estimates from DNA methylation-based tools were observed, irrespective of the normalization method, with some nuanced differences across cohorts and tools. Adjusting for EPIC version or calculating estimates separately for each version largely mitigated these version-specific discordances.

Conclusions:

Our work illustrates the importance of accounting for EPIC version differences in research scenarios, especially in meta-analyses and longitudinal studies, when samples profiled across

different versions are harmonized. Alongside DNA methylation-based tools, our observations also have implications in interpretation of epigenome-wide association studies (EWAS) findings, when results obtained from one version are compared to another, particularly for probes that are poorly concordant between versions.

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, 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[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 Infinium microarrays (27K, 450K and EPICv1) have been widely used to develop DNA methylation-based bioinformatic tools including cell type deconvolution algorithms[8–10], a rapidly increasing and diverse set of 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]. 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[11–14,16,17], and many, but not all, of predictive CpGs employed by these tools are retained on EPICv2[6]. 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[22–24]. 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.

Here, we used an unprecedented set of 67 primary human population samples and five technical replicates to systematically assess the concordance of inferred estimates derived from a broad set of DNA methylation-based tools between EPICv1 and EPICv2. We also tested whether EPIC version differences might affect meta-analyses, which harness statistical power that comes from combining multiple cohorts, and longitudinal studies, which often include samples profiled on different arrays/versions. Specifically, we created sets of matched venous blood samples from 67 individuals across a diverse collection of four populations spanning early to late adulthood, and profiled the DNA methylome using the EPICv1 and EPICv2 arrays. Using this unique dataset, we tested the concordance of the two EPIC versions at both array and probe levels and illustrated the potential contribution of EPIC version to overall DNA methylation variation. To explore if these version differences are reflected in DNA methylation-based tools, including immune cell

type deconvolution algorithms, epigenetic clocks, inflammation and lifestyle biomarkers, we compared their estimates between the two EPIC versions, and confirmed that these differences persisted irrespective of data preprocessing methods. We also demonstrated different remediation methods to account for these EPIC version discrepancies in epigenetic investigations. Collectively, this work encourages careful consideration while harmonizing data profiled on the two EPIC versions and comparing epigenome-wide association studies' (EWAS) findings from one version to another.

Results

Unsupervised clustering analyses 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 in the Kobor Lab, Vancouver, Canada (in-house facility): (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], and an external cohort processed in a different facility (Max Planck Institute of Psychiatry in Munich, Germany): Biological Classification of Mental Disorders study (BeCOME)[28] (Table 1, Figure 1). These four cohorts represented distinct demographic and biological characteristics such as sex and age range. Using unsupervised hierarchical clustering in a cohort-wise manner across the 721,378 probes shared between EPICv1 and EPICv2, we noted that samples exclusively clustered by EPIC version (Figure 2B and 2C). Consistently, when smaller subsets of predictive CpGs employed by DNA methylation-based tools were used for clustering, 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). When clustering was performed on all the three cohorts processed by the in-house facility, we similarly observed a clear demarcation of samples first by EPIC version, followed by cohort despite their distinct characteristics (Figure 2A). Perhaps not surprisingly, when samples from BeCOME were combined with the other three cohorts, we observed samples clustering by the facility they were processed in, while still separating by EPIC version within the two facilities (Supplementary Figure 2). Our findings remained consistent when we performed unsupervised clustering using principal component analysis (PCA). The top three PC loadings, corresponding to >98% of DNA methylation variation, were significantly associated with EPIC version (Supplementary Figure 3). Overall, EPIC version was the primary contributor to clustering of samples and explained a large proportion of DNA methylation variation.

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

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

between- and within-version using intra-class correlation coefficients (ICC), and found slightly lower agreement in between-versions technical replicates (0.9947-0.9958) than within-version (0.9972-0.9983) (Supplementary Table 1). Finally, we assessed the error in technical replicates between and within EPIC versions using root mean square root error (RMSE). As expected, the mean RMSE between-versions (0.0367-0.0416) was slightly higher than the within-version (0.0214-0.0264) (Supplementary Table 1).

Extending beyond the technical replicates to all the matched samples profiled on each version, we noted high array level Spearman correlation ranging from 0.968 to 0.981 between EPICv1 and EPICv2, though this was at the array level (Supplementary Table 2). In contrast to the strong array level concordance, Spearman correlations at the probe level, calculated as the concordance of β values of individual probes across samples between versions, were not as congruent. Specifically, only ~25% the probes per cohort had Spearman correlations greater than 0.70 while the remaining ~75% probes showed a low mean correlation of <0.30 (Supplementary Table 3). One plausible explanation for low probe level correlations of the majority of probes may be low inter-sample variability. It has been previously reported that probes with a narrow range of DNA methylation β values across samples tend to have poor correlations[22]. To test this, we calculated pooled standard deviation (SD) across samples on a cohort basis, and defined the first lower quartile (≈ 0.01 for each cohort) as the pooled SD threshold to denote low variability (Supplementary Table 2). Supporting our hypothesis, we found that probes with Spearman correlations ≤ 0.70 had significantly lower pooled SD or low variability compared to probes with higher correlations (Supplementary Table 3). Specifically, of the probes with Spearman correlation ≤ 0.70, ~30% exhibited low inter-sample variability; in contrast, of probes with

Spearman correlation > 0.70, only 0.069-5.577% exhibited low variability (Supplementary Table 6). We also cross-referenced these poorly-concordant probes, which we defined as probes with low correlation despite their high variability, with low quality and unreliable probes identified in previous annotations of the Illumina arrays[5,29–31]. Specifically, 0.11-2.39% of these poorly-concordant probes overlapped with previously identified as cross hybridizing to multiple genomic locations or mapped to genetic variant sites across all cohorts [5,29–31] (Supplementary Table 4). Of the 82 probes that have undergone design type switches in EPICv2[6], 27-40 probes also overlapped with our poorly-concordant probes in the four cohorts. Overall, only a small fraction (~10%) of these poorly-concordant probes overlapped with probes of bad quality, design type switch[6], and cross hybridization to the genome[5,29–31] (Supplementary Table 5), while the rest still remained unexplained. In contrast, of the highly correlated or low variability probes, ~80% overlapped with platform-bias free and high-confidence mapping probes recently identified in cell lines[32], confirming the agreement of the readouts at these probes (Supplementary Table 5).

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

of shared probes was noted when the Spearman correlation threshold was reduced to 0.50 (Supplementary Table 6). In summary, we observed high concordance at the array level between versions; however, on the probe level, there were a subset of probes with low concordance between versions regardless of high inter-sample variability.

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

Cellular composition is a key contributor to whole blood DNA methylation variation and has been associated with disease phenotypes and often included as a covariate in statistical models to account for cell type heterogeneity[33–35]. Most studies do not measure actual cell counts and rely on predicted values from DNA methylation-based algorithms. Given that these algorithms use references profiled on previous arrays/versions, we tested whether there were differences in cell type proportions estimated from the two EPIC versions using one of the most commonly used cell type deconvolution methods for whole blood[10]. We compared the proportions of twelve immune cell types using the *FlowSorted.Blood.EPIC* R package with two commonly used probe selection methods: IDOL and auto. The IDOL method includes 1200 pre-selected probes, which has over 99% coverage in EPICv2 (Table 2 and Supplementary material), and the auto method includes 1200 probes independently selected for each EPIC version as described in Methods. On comparing the probes selected for each EPIC version using the auto method, we identified a large overlap of over 90% probes between the two EPIC versions, with only 10% these auto-selected probes overlapping with the pre-selected IDOL probes (Supplementary Figure 4).

We found high Spearman correlations between cell type proportions inferred on EPICv1 and EPICv2, with an average correlation of 0.883 by the IDOL method and 0.890 by the auto method (Supplementary Tables 7-8). Despite their high correlation, we identified significant differences in estimated proportions of five and nine cell types between EPICv1 and EPICv2 using the IDOL and auto methods, respectively, across the three in-house cohorts, with four cell types (Bas, CD8mem, Mono, NK) commonly identified in both methods. Consistent findings were observed in BeCOME with the auto method; two additional cell types (CD8mem and NK) showed significant proportional differences between versions when the IDOL method was used (Supplementary Figure 5, Supplementary Tables 7-8). Overall, estimated proportions were significantly different between the EPIC versions across multiple cell types, and these findings remained consistent irrespective of the probe selection method, albeit with varying effect sizes depending on cell type and cohort (Figure 3, Supplementary Figure 5-6, and Supplementary Tables 7-8).

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

Epigenetic clocks are based on the property that DNA methylation levels at specific CpGs highly correlate with chronological age or age-related outcomes[11–14,16,17]. In order to evaluate the concordance between EPICv1 and EPICv2 in the context of these DNA methylation-based tools, we compared estimates from seven widely used first- and second- generation epigenetic clocks in VHAS, CLHNS, and CALERIE (see Methods). Across these seven epigenetic clocks, ~77-96% of predictive CpGs were retained on EPICv2 (Table 2 and Supplementary material). We identified high Pearson correlations of 0.807-0.996 between chronological age and epigenetic age obtained from both EPICv1 and EPICv2 (Supplementary Table 9). EPICv2 epigenetic ages

for all clocks were highly correlated with EPICv1 epigenetic ages (0.879-0.996), with differences in epigenetic ages between technical replicates being 0.151-4.206 (Supplementary Table 10), although we noted significant differences in epigenetic ages between EPICv2 and EPICv1 in Hannum, Horvath skin and blood, and PhenoAge, with these differences ranging in effect size from 0.267-4.137 (Supplementary Table 10, Supplementary Figure 7A). Consistent with these results, in the external validation cohort BeCOME, EPICv1 and EPICv2 epigenetic ages were highly correlated with chronological age and yet there were significant epigenetic age differences between versions in all epigenetic clocks except again Horvath pan-tissue (Supplementary Tables 9-10, Supplementary Figure 8A).

We also estimated epigenetic age acceleration (EAA), a measure of rate of aging which has been associated with health outcomes, by considering samples profiled on each EPIC version separately and in a combined manner (see Methods). Irrespective of how matched samples on the two EPIC versions were considered for EAA calculation, we noted modest to high correlation (0.574-0.960) of EAA between the EPIC versions. When EAA was calculated separately by versions, we noted no significant differences between EPICv1 and EPICv2 (Supplementary Table 11, Figure 4B, Supplementary Figures 9B and 10B). When clock estimates from the EPIC versions were combined prior to EAA calculation, there were significant differences between EPICv1 and EPICv2 for all clock EAAs except Horvath pan-tissue (Supplementary Table 11, Figure 4C, Supplementary Figures 9C and 10C). To next test whether the different EPIC versions contribute to observed EAA differences, we combined epigenetic ages in both EPIC versions and then calculated EAAs by including version as a covariate in the linear regression. In doing so, we noted that there were no significant EAA differences between the two EPIC

versions, akin to when EAA was calculated separately by version (Supplementary Table 11, Figure 4D, Supplementary Figures 9D and 10D). We repeated the same analyses using epigenetic clocks estimated based on principal components (PC clocks) rather than individual predictive CpGs, as they have been shown to overcome unreliability in clock estimates due to technical noise[36]. Once again, we noted that there were no significant EAA differences when EAA was calculated for each EPIC version separately. When EAAs were calculated on combined sets of matched samples, we noted significant differences in PCHorvath skin and blood and PCPhenoAge, and these differences were no longer significant with EPIC version adjustment (Supplementary Figures 11-13). Consistent with other cohorts, in BeCOME, we noted similar findings in epigenetic ages, EAAs, and PC clocks estimates (Supplementary Table 11, Supplementary Figure 8 and 14).

On next evaluating the DunedinPACE, DNAmTL, and epiTOC clocks, we observed a high correlation (0.699-0.983) of clock estimates between EPICv1 and EPICv2 (Supplementary Table 12). Unlike the epigenetic clocks above, when we examined the EPIC versions separately, we noted that DNAmTL estimates were significantly different between versions in all four cohorts, while DunedinPACE and epiTOC estimates were significantly different in three out of the four cohorts (Supplementary Tables 11-12, Figure 5, and Supplementary Figures 8-10).

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[37]. Similar to other tools, approximately 90% predictive CpGs were retained on EPICv2 (Table 2 and Supplementary

material) and estimates between EPICv1 and EPICv2 were modestly to highly correlated (0.788-0.993) (Supplementary Table 13). When estimates were calculated separately by versions (EPIC versions separate), we identified significantly different IL-6, CRP, alcohol, and smoking scores between EPICv1 and EPICv2 in at least one of the four cohorts. When estimates from the EPIC versions were first combined and then adjusted for version, as expected, we noted no significant differences between EPICv1 and EPICv2 in any of these four predictors (Supplementary Table 13, Supplementary Figure 15).

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

Given the effect of data normalization on epigenetic clock estimations[23,38], we examined whether normalization methods affect the observed differences between EPIC versions, obtained after applying functional normalization (funnorm). To test this, we combined the shared probes on the two EPIC versions and applied Beta-MIxture Quantile (BMIQ) normalization, and subsequently calculated epigenetic ages, EAAs, and inflammation and lifestyle biomarker estimates. Overall, the differences in estimates of these DNA methylation-based tools between EPIC versions when BMIQ normalization was applied were consistent with those obtained in functional normalized data, specifically in Horvath skin and blood and PhenoAge which were consistent across the three in-house cohorts (Supplementary Figures 7, 16-21).

Next, we also examined if applying EPIC version correction using batch correction algorithms can circumvent version-specific effects and eliminate the observed differences in DNA methylation-based estimates between EPICv1 and EPICv2, given the observed contribution of

EPIC version in explaining DNA methylation variation. We tested this by combining common probes on EPICv1 and EPICv2 after functional normalization or BMIQ normalization, and sequentially adjusted for EPIC version, chip and sample position on chip (row) effects as applicable using the ComBat function implemented in the *sva* R package[39]. As expected, when epigenetic clocks, inflammation and lifestyle biomarkers were calculated using version-corrected input β values, there were no significant differences in estimates between the EPIC versions (Supplementary Figures 7, 22-24).

Discussion

Various generations of DNA methylation arrays, largely manufactured by Illumina, have had tremendous impact on the field of epigenetics, enabling large population studies of human health and disease. The most recently released Illumina EPICv2 microarray provides wider coverage of the DNA methylome compared to its predecessor EPICv1, yet at the same time eliminates approximately one-fifth of the previous probes. This then might create challenges for the utility and transferability of the myriad bioinformatic tools developed for the EPICv1, especially if some of the missing probes were included in a particular tool. Here, using matched venous blood samples from 67 adults across four geographically diverse populations, we comprehensively examined the concordance between EPIC versions in a wide range of commonly used DNA methylation-based tools. Overall, our study identified notable differences between the two EPIC versions at various levels, and provided some insights into possible remediation approaches. First, we observed overall high array level concordance but variable probe level concordance, with a subset of probes displaying poor agreement between versions despite high inter-sample variability. Second, EPIC version differences were identified across all cohorts in DNA

methylation-based tools, including epigenetic clocks, inflammation and lifestyle biomarkers, and cell type predictors. Third, discordance between EPIC versions was dampened by the manner of statistical accounting of EPIC version. Fourth, irrespective of the choice of normalization method EPIC version differences still persisted. Overall, our work emphasizes careful consideration of research settings when i) samples across both EPIC versions are harmonized, and ii) findings derived from one EPIC version are compared to those derived from the other version.

Initial comparisons of EPIC versions that were confined to immortalized cell lines and a limited number of human primary samples reported generally good array level correlations between the two EPIC versions[5–7]. However, these studies were not powered to investigate concordance at the probe level nor did they compare DNA methylation-based tools such as epigenetic clocks, cell types, and biomarker predictors, particularly across geographically diverse settings, which are highly relevant for human population epigenetic studies. Using this population-based framework, unsupervised clustering on samples across the three in-house cohorts (VHAS, CLHNS, and CALERIE) using the shared probes between EPICv1 and EPICv2 showed that samples clustered based on the EPIC version on which they were measured, despite cohort-specific characteristics. Perhaps not surprisingly, given the differences in sample handling and processing inherent to different experimental facilities, inclusion of the external validation cohort BeCOME showed that samples clustered by facility first. However, performing these analyses on each cohort separately, we again found that EPIC version was the principal driver of DNA methylation variation, irrespective of the facility in which these samples were processed in. This separation by EPIC version may be due to the variable concordance of individual probes between the two versions. This was in general agreement with previous work which showed low

reliability across shared probes measured on previous iterations of Illumina arrays, namely 450K and EPICv1[22,40,41]. While a small proportion of these poorly-concordant probes, skewed toward lower inter-sample variability, could be attributed to their non-specific probe hybridization to the genome, underlying genetic variation, or technical differences such as altered EPICv2 probe design[5,6,29–31], the majority remain unexplained by these factors. Further, these poorly-concordant probes will have implications in interpretation of EWAS results, where validation of findings from one EPIC version to another may be challenging.

We observed significant systematic differences in measures derived from DNA methylation-based tools. In the context of cell type prediction algorithms, there were significant differences in proportions of various cell types, when using either the IDOL or auto prediction methods. These differences may be reflective of version incompatibility between the reference library used, built on previous Illumina microarrays, and EPICv2[10]. Further, they may be compounded by the compositional nature of cell type estimates, meaning that any changes in the proportion of even one cell type will influence the proportions of other cell types. Similarly, in the epigenetic clock and biomarker predictors, we noted discordance in estimates obtained between EPICv1 and EPICv2, though these observations were generally consistent across cohorts and tools with some nuanced differences. It is reasonable to speculate that this discordance between versions may be due to the varying number and weights of predictive CpGs absent on either EPIC version, and the influence of subsequent imputation performed by the clock algorithms to account for this missingness. However, our analyses using BMIQ normalization which only used shared probes between the versions and thus had the same absent predictive CpGs in both versions, still showed

differences in estimates and EAAs, suggesting that absent CpGs are not the main contributor to the version differences.

Although epigenetic clocks are often analyzed as direct comparisons of epigenetic ages as described above, it has been noted that a regression-based measure of epigenetic age, EAA, is more appropriate in most cases as it is robust to data preprocessing[11,23,37,38]. On calculating EAAs by taking the common approach of including all samples in a cohort irrespective of version, we noted that EAA estimates were different between versions, suggesting that this method of calculating EAA is still sensitive to inherent variation between EPICv1 and EPICv2. Using an alternate approach, when we calculated EAA estimates separately for samples profiled on each EPIC version, we found that these differences were no longer significant. This indicates that EAAs may be a suitable measure of epigenetic aging when version is taken into consideration while calculating EAA. Similarly, we noted consistent differences between versions in inflammation and lifestyle biomarker predictor scores, emphasizing again that it is important to account for systematic version differences when calculating DNA methylation-based measures.

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

introduce false biological signals[42–45]. This ideal - and perhaps even unrealistic - scenario starkly contrasts with a typical meta-analysis, where a researcher might wish to combine results from multiple distinct sample groups measured on different EPIC versions. Furthermore, another research setting that presents a similar challenge is a longitudinal study design, where the aim is to explore DNA methylation changes over time. In such cases, when samples from each timepoint are measured on different EPIC versions, it inevitably results in a confound between the variable of interest, i.e., timepoint and EPIC version, thereby not enabling any version correction. If samples from the previous timepoint are still available, one feasible approach is to include a small number of these samples profiled on the previous EPIC version alongside samples from the new timepoint to be profiled on EPICv2 such that any between version differences can be monitored. It is also known that DNA methylation-based tools are sensitive to normalization methods [23,37,38], and it is important to explore the contribution of such approaches to observed EPIC version differences noted in the present study. By employing two distinct normalization methods, namely funnorm and BMIQ, we found similar differences in DNA methylation-based estimates based on version, irrespective of normalization method used.

On the strength of our matched sample design utilizing four different human cohorts, our study improves the current understanding of the applicability of DNA methylation-based tools for EPICv1 and EPICv2, however, there are several limitations. Our analyses may be limited by relatively modest sample sizes within each cohort, despite this, our study includes matched samples collected from adults with a wide age range and across diverse geographical regions. While the time between EPICv1 and EPICv2 measurement varied across the cohorts, 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 predominantly composed of European populations, the suitability of investigated clocks may not be well established in our investigated Asian cohorts, VHAS and CLHNS respectively [46–48]. However, differences in these cohorts were similar to CALERIE and BeCOME, which are comprised primarily of individuals of European descent, indicating that our findings were not limited to specific populations. Further, we compared EPICv2 cell type proportions and biomarker scores to EPICv1, although we recognize that actual cell counts and biomarker measurements would be more appropriate as the ground truth and would be useful in validating our DNA methylation-based estimates. In spite of this, the primary aim of the current work is to investigate the concordance in estimates between versions, and not to evaluate the accuracy of biomarker prediction. Finally, although our study focused on venous blood, a commonly used tissue in epigenetics research, we speculate that these EPIC version differences will hold true in other tissue types as well. While multiple tissue types were not collected in all of these cohorts, we had access to matched capillary and venous blood samples in VHAS, which allowed us to at least test and confirm the consistency of our findings in another blood preparation.

Conclusions

With the rapid generation of DNA methylation data profiled on the newer iterations of the Illumina microarrays, integrating samples across these platforms poses a challenge, owing to discrepancies in probe content among arrays/versions. Our findings highlight differences in the new EPICv2 compared to EPICv1, demonstrate the influence of EPIC version on the majority of commonly used DNA methylation-based tools, and provide possible remediation approaches to minimize technical variation which may arise from inconsistencies between versions. We

therefore encourage careful consideration when harmonizing and interpreting DNA methylation data across multiple arrays/versions to ensure reliability and reproducibility in epigenetics analyses.

Methods

Description of cohorts

To compare the performance of the two most recent generations of MethylationEPIC BeadChip Infinium microarrays in the context of DNA methylation-based clocks, biomarkers and cell type proportion estimates, we measured the DNA methylomes of a subset of venous whole blood samples on both EPICv1 and EPICv2 selected from the (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] cohorts. 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). Samples from VHAS, CLHNS and CALERIE were processed in the same facility (Kobor Lab, University of British Columbia and BC Children's Hospital Research Institute, Vancouver, Canada). To compare our findings from the three cohorts processed in-house to those from an external facility, we used an independent validation cohort, BeCOME. In this dataset, DNA methylation was measured on EPICv1 and EPICv2 using matched samples at the Max Planck Institute of Psychiatry in Munich, Germany (Table 1). 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 and BeCOME. 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 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 MethylationEPIC BeadChip Infinium microarray v1.0 B5 (EPICv1) and Infinium MethylationEPIC v2.0 (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 after bisulfite conversation were quantified to monitor technical variation within each EPIC version, independently for each cohort. Both VHAS and CALERIE included two technical replicates each on EPICv1 and EPICv2, and CLHNS included one technical replicate on EPICv2 and no technical replicate on EPICv1. The external validation cohort BeCOME did not include any technical replicates on either EPIC version. Sample quality control checks were performed as described in previous publications[33,49]. Blood samples collected in the three cohorts and the external validation cohort passed all 17 Illumina quality control metrics in the *ewastools* R package[50,51], and detection *p*-value, beadcount, average methylated and unmethylated intensity metrics in the *minfi* R package (Supplementary Table 14). We also performed sample identity checks using unsupervised hierarchical clustering analysis on the 57 single nucleotide

polymorphism (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; BeCOME-EPICv1: 12,511, EPICv2: 20411, common to both versions: 2,714), but all probes in EPICv1 and EPICv2 were retained for subsequent analyses.

Unsupervised clustering analyses on DNA methylation data

To perform unsupervised hierarchical clustering on samples, we calculated array level Spearman correlations using the 721,378 probes shared between EPICv1 and EPICv2 in pair-wise manner for all the samples in the four cohorts. Array level Spearman correlations were then clustered by the complete linkage method with Euclidean distance using the hclust function implemented in the *stats* R package[52] and visualized by pheatmap in the *pheatmap* R package[53]. We first performed clustering on each of the four cohorts separately, then by combining the three cohorts processed in the in-house facility (VHAS, CLHNS and CALERIE), and finally by combining all four cohorts regardless of processing facility. Grouping cohorts once again in a similar manner, we performed principal component analyses (PCA) using the shared probes between the EPIC

versions. We tested the association between each of the top five PC loadings and EPIC version, cohort, and sex by employing one-way analysis of variance (ANOVA) or t-tests and applying a Bonferroni multiple test correction. The percentage of DNA methylation variation explained by each variable (EPIC version, cohort, and sex) was calculated as the ANOVA or t-test R-squared value. It should be noted that while performing the PCA on each cohort independently, we did not test the association between PC loadings and sex in the CLHNS cohort since this cohort comprises of only females; for the other three cohorts, associations with EPIC version, cohort, and sex were tested.

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[54,55] independently on EPICv1 and EPICv2 samples for each cohort. Cohort-specific technical replicate sample correlations were used to monitor preprocessing (technical replicates were not available in the BeCOME external validation cohort). 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). 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[40,56]. 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[57]. 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[39] 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[39]. 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. Technical

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

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[58,59]. 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, CALERIE, and the external validation cohort BeCOME. Epigenetic clocks analyses on capillary blood samples in the VHAS cohort were also performed (Supplementary material, Supplementary Figures 25-28). 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). Missing clock CpG β value imputation was performed by the clock algorithms. 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[36] 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.

DunedinPACE, a rate-based clock, was calculated using the DunedinPACE R package[17], and the two other clocks, epiTOC and DNA methylation-based estimator of telomere length (DNAmTL)[16] were obtained using the *getEpiTOC* function in the *cgageR* R package[18] and the online DNA Methylation Age Calculator (https://dnamage.genetics.ucla.edu/new), respectively. Missing clock CpG β value imputation was performed by the clock algorithms. Unlike EAA calculation, there were no secondary measures calculated from these three clock 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 and other clock estimates were calculated for EPICv1 and EPICv2 samples without EPIC version adjustment.
2. **EPIC versions combined and version adjusted:** rate-based and other 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 ≥ 0.8) based on recommended benchmarks[60].

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

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

References

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

24. Dhingra R, Kwee LC, Diaz-Sanchez D, Devlin RB, Cascio W, Hauser ER, et al. Evaluating DNA methylation age on the Illumina MethylationEPIC Bead Chip. PLOS ONE. 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. Brückl TM, Spoormaker VI, Sämann PG, Brem A-K, Henco L, Czamara D, et al. The biological classification of mental disorders (BeCOME) study: a protocol for an observational deep-phenotyping study for the identification of biological subtypes. BMC Psychiatry. 2020;20:213.

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. 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.

31. 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.

32. Chen BH, Zhou W. mLiftOver: harmonizing data across Infinium DNA methylation platforms. Bioinformatics. 2024;40:btae423.

33. 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.

34. 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.

35. 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.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

51. Illumina Inc. BeadArray Controls Reporter Software Guide. 2015;

52. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing. 2022; Available from: https://www.R-project.org/

53. Kolde R. pheatmap: Pretty Heatmaps. 2019;

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

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

56. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. J Chiropr Med. 2016;15:155–63.

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

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

59. 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.

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

Declarations

Ethics approval and consent to participate

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

Consent for publication

Not applicable.

Availability of data and materials

The datasets generated and/or analyzed during the current study are not publicly available but are available from the corresponding author or cohort owners/custodians upon reasonable request. The EPICv2 preprocessing pipeline and analysis scripts in R are available on GitHub https://github.com/kobor-lab/EPICv2_QC_preprocessing, and https://github.com/kobor-lab/EPICv1v2_comparison_manuscript.

Competing interests

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

Funding

This work received supported from the Canadian Institutes of Health 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.

Authors' contributions

Conceptualization: B.C.Z, M.S.J., C.K., K.Korthauer, M.S.K.; biological data collection/generation: 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., M.S.K.; DNA methylation data generation: J.L.M., K.D., M.S.K. (Kobor Lab facility), E.B.B. (MPIP); DNA methylation data curation and processing: B.C.Z., M.S.J., N.Y., H-R.E., A.A.H., D.C. J.W.; Phenotypic data curation: 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.; Data analysis: B.C.Z., M.S.J.; Interpretation of results: M.S.J., C.K., B.C.Z., K.Korthauer, M.S.K.; Manuscript writing: M.S.J., C.K., B.C.Z., M.S.K.; Figures/visualization: B.C.Z., M.S.J.; Manuscript review: 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.

Acknowledgements

We thank Dr. Sarah Merrill, Maggie Fu, and Dr. Kimberly Schmidt for their helpful feedback during manuscript preparation. We would also like to thank all cohort participants for their involvement and providing biological samples for research, and cohort study teams for their efforts. We wish to thank Monika Rex-Haffner and the whole BeCOME study team for their technical assistance. This research utilized the *FlowSorted.BloodExtended.EPIC* software packages developed at Dartmouth College, which are governed by the licensing terms provided by Dartmouth Technology Transfer. D.W.B. and M.S.K. are fellows of the CIFAR CBD Network.

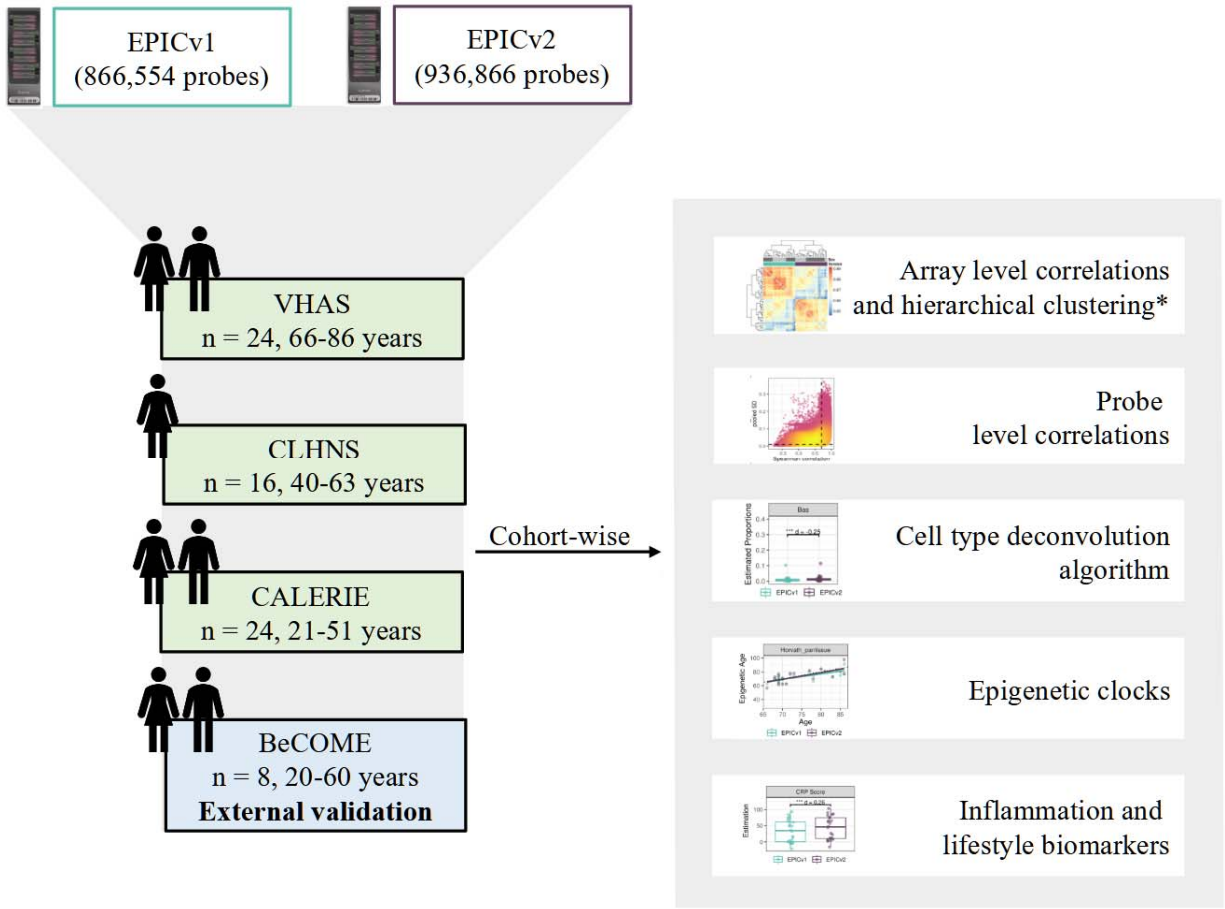

Figure 1. Overview of study design and analyses.

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

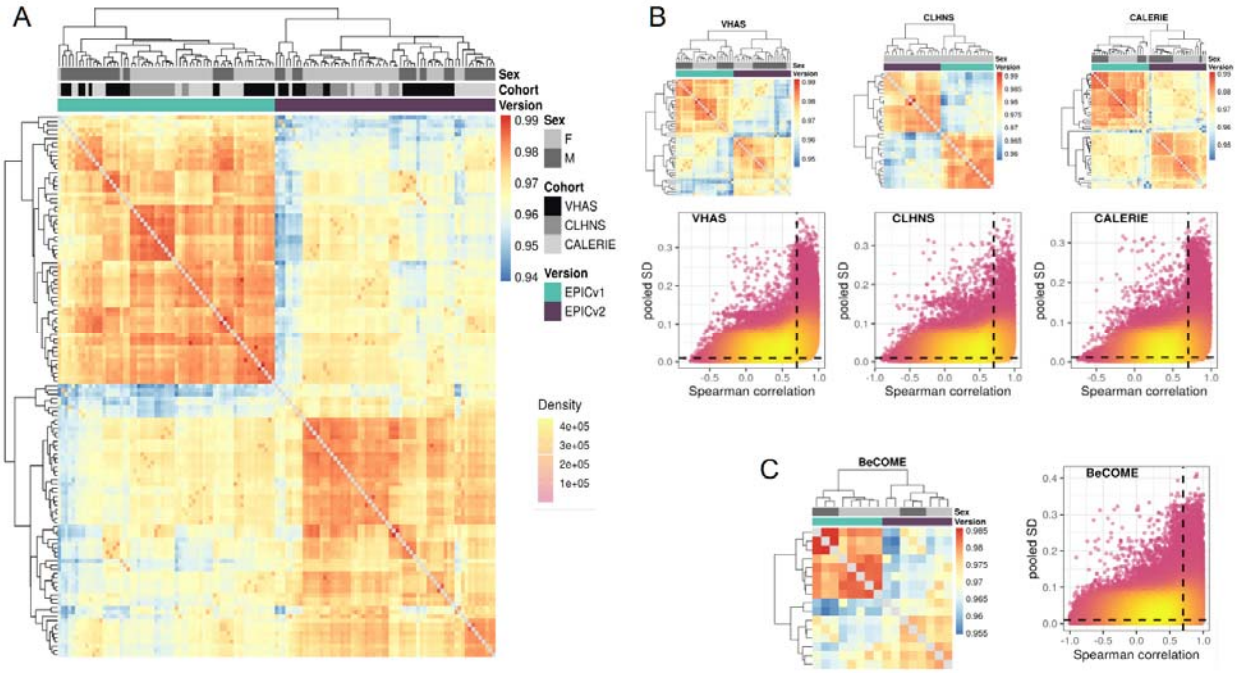

Figure 2. (A) Unsupervised hierarchical clustering of array level Spearman correlations between matched VHAS, CLHNS, and CALERIE samples on EPICv1 and EPICv2 across 721,378 shared probes; blue to red color range denotes Spearman correlation from 0.94-1.00. **(B)** Top: Cohort-wise hierarchical clustering of sample-to-sample Spearman correlations across 721,378 probes common between EPICv1 and EPICv2; Bottom: Cohort-wise Spearman correlation and pooled standard deviation (SD) of common probes. The X-axis represents the Spearman correlation and Y-axis represents the pooled SD of probes common to both arrays. Dashed horizontal line indicates the pooled SD threshold set at lower quartile pooled SD for each cohort, 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. **(C)** Hierarchical clustering of Spearman correlations between matched BeCOME (external validation cohort) samples on EPICv1 and EPICv2 across 721,378 shared probes and Spearman correlation and pooled SD of common probes.

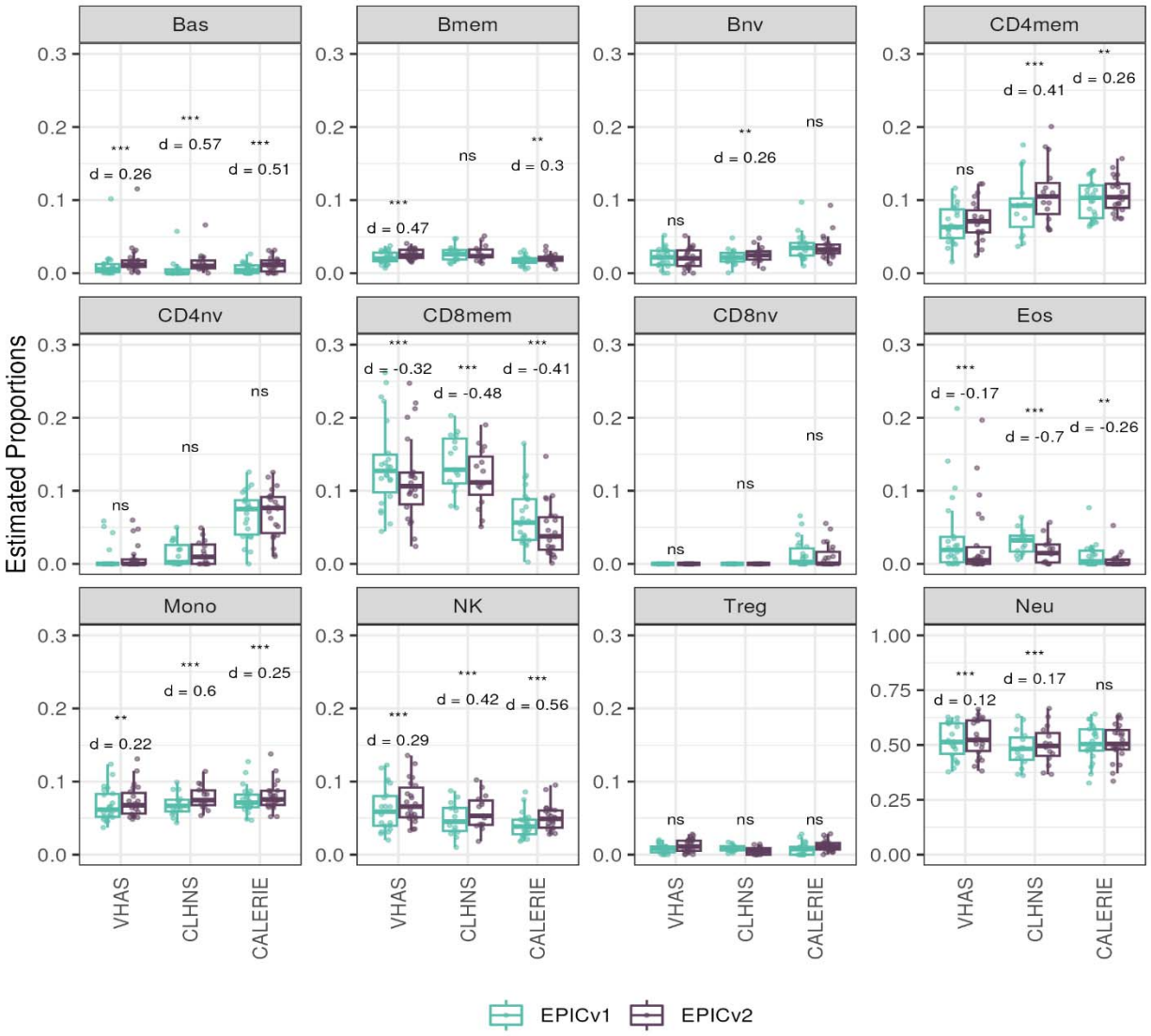

Figure 3. Differences in DNA methylation-based immune cell type proportions estimated using the IDOL reference on matched samples assessed on EPICv1 and EPICv2 in VHAS, CLHNS, and CALERIE. Statistical significance was defined as Bonferroni adjusted *p*-value <0.05. ** denotes Bonferroni *p* <0.05, ***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

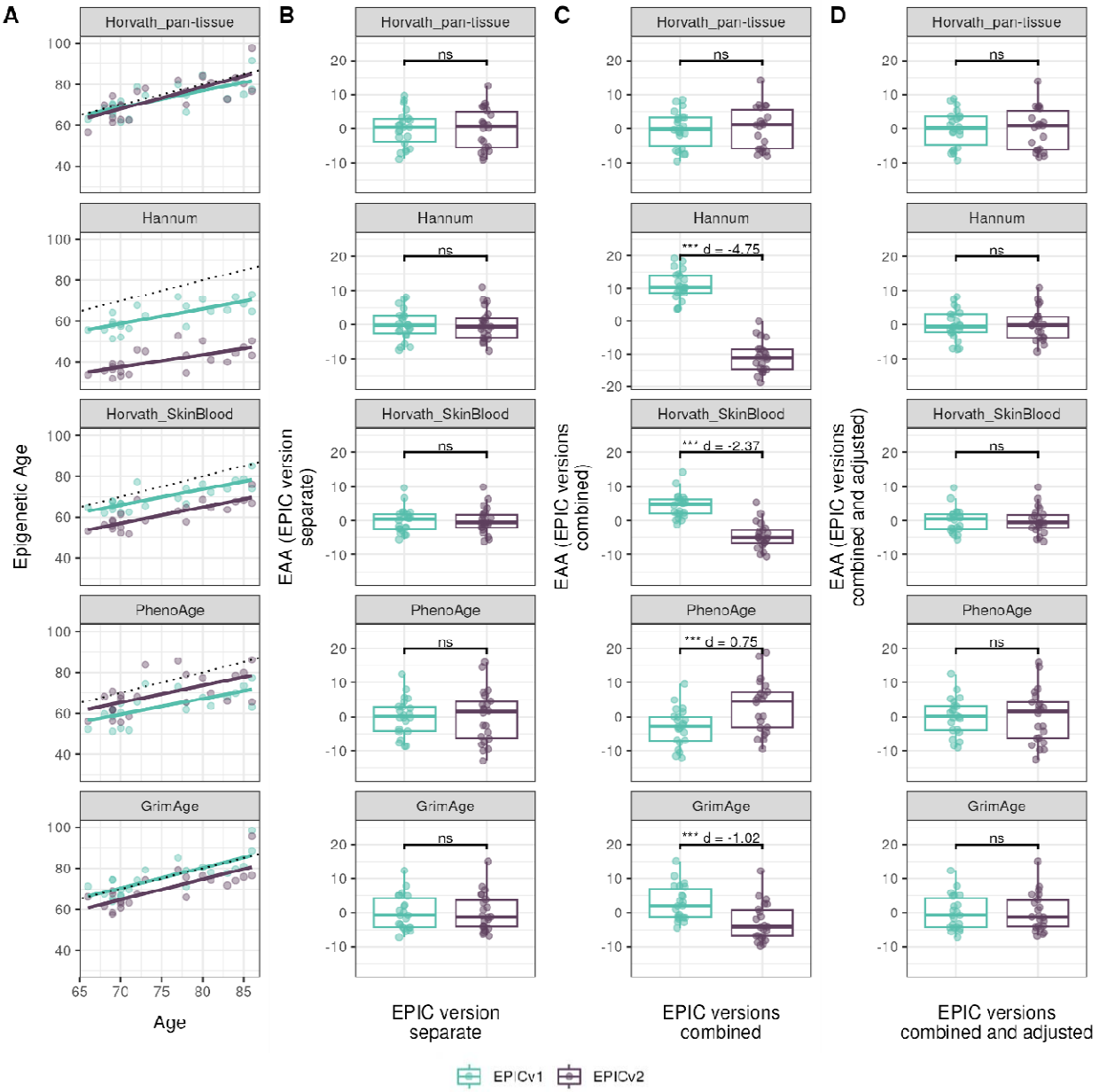

Figure 4. Epigenetic ages on matched samples assessed on EPICv1 and EPICv2 in VHAS. **(A)** Scatter plot of Horvath pan-tissue, Hannum, Horvath skin and blood, PhenoAge, and GrimAge clock ages (Y axis) and chronological age (X axis) with dotted line indicating x=y, coloured by EPIC version. **(B-D)**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

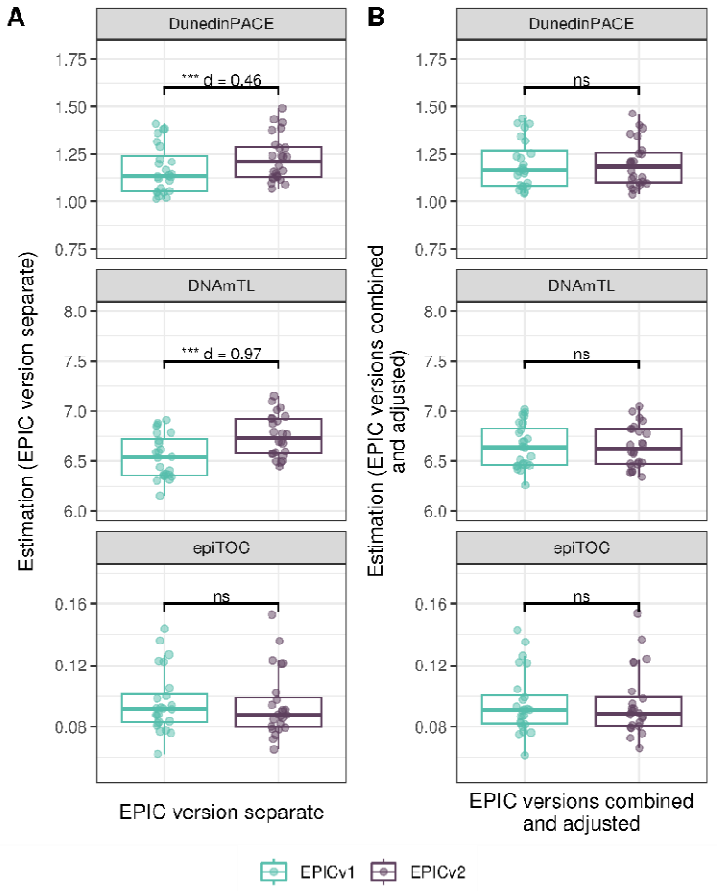

Figure 5. Rate-based and other clock estimates on matched samples assessed on EPICv1 and EPICv2 in VHAS. Boxplots comparing DunedinPACE, DNAmTL and epiTOC estimates calculated by considering EPIC versions separately **(A)** and combined and EPIC version adjusted **(B)**, between EPICv1 and EPICv2. Statistical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

	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

Table 2. Summary of predictive CpGs of DNA methylation-based clocks, biomarker predictors, and cell 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	0	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)

* There are CpGs commonly absent in both EPICv1 and EPICv2 in these clocks: Horvath pan-tissue: 3 CpGs; Hannum: 2 CpGs; epiTOC: 6 CpGs; CRP score: 20 CpGs. There are probes absent in EPICv1 and are reintroduced in EPICv2 in these clocks: Horvath pan-tissue: 14 CpGs; Hannum: 4 CpGs; epiTOC: 25 CpGs; CRP score: 83 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.