Adapting Epigenetic Clocks for Cell-Free DNA High-Throughput Sequencing Data

Guangyu Li [1,6], Weilai Huang [2,6], Xuechen Zhao [2,6], Jun Wu [2], Yuanfang Guo [2], Liying Chen [3], Xingqi Cao [3], Zipeng Yang [2], Siqi Jiang [2], Bin Hu [2], Yijun Wang [4], Dinson Tan [2], Vicky Tong [2], Changdong Tang [2], Xiaowen Feng [2], Xiaomin Hu [5], Chuan Ouyang [2,*], Guangyu Zhou [1,2,*]

[1] Liangzhu Laboratory, Zhejiang University School of Medicine, Hangzhou, 310000, China

[2] Regenerative Bio Inc., Hangzhou, Zhejiang, China

[3] Department of General Practice, Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, Hangzhou, Zhejiang, China

[4] School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, United States

[5] Department of Medical Research Center, Peking Union Medical College Hospital, Chinese Academy of Medical Science & Peking Union Medical College, Beijing, China

[6] These authors contributed equally: Guangyu Li, Weilai Huang, Xuechen Zhao

[*] Corresponding authors

E-mail: gavin.zhou@regen-bio.com; orcharry@gmail.com

Abstract

Cell-free DNA (cfDNA) methylation sequencing holds promise for developing epigenetic aging clocks. However, current clocks—primarily trained on array-based data—do not readily generalize to high-throughput sequencing (HTS) cfDNA profiles. Using datasets with technical replicates encompassing HTS data from both cfDNA and gDNA, alongside gDNA methylation array data, we systematically assessed factors influencing clock accuracy and reproducibility. We identified key strategies to overcome HTS-specific challenges: maintaining ≥10× mean target depth, applying elastic net regression with strong L2 regularization, and imputing unreliable beta-values. Transfer learning further enhanced accuracy robustly across multiple independent

cohorts. Our findings demonstrate that array-derived epigenetic clocks can be effectively adapted to cfDNA sequencing data. This work offers critical methodological insights and practical guidelines, advancing the feasibility of minimally invasive aging assessment using cfDNA.

Main

Recent studies advocated `biological age` (BA) over `chronological age` (CA, the time elapsed since birth) for a more accurate indicator of an individual's physiological aging state. BA is estimated using `Biomarkers of Aging` (BoAs), which are quantified via mathematical models known as aging clocks [1–3]. Among diverse BoAs ranging from clinical parameters [4,5] and telomere length [6] to various "omic's" data and facial images [7–11], epigenetic clocks—based on the methylome (DNA methylation profiles)—have distinguished themselves by their outstanding predictive power [4,9,12–20], achieving high accuracy even with linear models [4,9,12,19,21,22]. Due to their superior performance, epigenetic clocks are now widely used in health monitoring, age-related disease assessment, and longevity interventions studies [2,17,23–25].

The circulating cell-free DNA (cfDNA or ccfDNA), primarily found in body fluids, enabled minimally invasive testing for early cancer detection and disease monitoring [26–29]. Its utility in epigenetic aging clocks is now being actively explored [30–32]. However, utilizing cfDNA for age prediction requires adapting epigenetic clocks, which are predominantly trained with methylation arrays (e.g. Illumina 450K and EPIC/EPICv2) data. Array platforms exhibit limited sensitivity to low-abundance methylation changes, narrow dynamic range, and inability to capture intra-sample methylation heterogeneity [33]. Furthermore, the low yield of cfDNA (typically 0.5–10 ng, compared to the recommended 200 ng input for arrays [34]) and its small fragment size present significant challenges for array-based platforms [35]. Even though methods such as preamplification can reduce the required input amount [36], the resulting potential distortion of beta-values still hampers array-based application. High-throughput sequencing (HTS), particularly targeted sequencing with hybridization probe capture (e.g., the Twist Human Methylome Panel), overcomes these limitations [33,37–43]. HTS offers a technically robust alternative solution for cfDNA methylation profiling [44,45], thereby providing a viable path toward cfDNA-based epigenetic aging and disease detection.

However, current epigenetic clocks have not been fully adapted to HTS data, particularly for cfDNA samples. Current clocks are still predominantly trained on array-based genomic DNA (gDNA) methylation data, with only a few exceptions trained on HTS gDNA data (such as TIME-seq) [18,30,33,46,47]. Previous studies illustrated that direct application of array-based gDNA clocks on cfDNA data resulted in impaired performance [13,33,42,46]. Datasets that allow comprehensive analysis of such performance differences are still lacking especially for cfDNA, although similar resource exists for buccal gDNA [42]. This gap is critical because cfDNA exhibits unique fragmentation patterns and multi-tissue origins [30,32,48]. Furthermore, methylation datasets with technical replicates are still scarce. This hinders assessment of clock reproducibility and reliability, particularly as no such evaluation has been conducted for cfDNA-based epigenetic clock applications. Finally, as cfDNA remains a novel biomarker for epigenetic clock applications, to our best knowledge, practical guidelines for experimental design, data processing, model selection, and benchmarking are still lacking.

Adaptation of epigenetic clocks across different data types has been explored under various rationales; however, none have specifically addressed cfDNA applications. Batch correction and harmonization methods can mitigate beta-value discrepancies between methylation arrays [49–51], though often at the cost of introducing false positives or artifacts [52]. Domain adaptation approaches enabled the transfer of array data across tissues and cohorts [53,54], but their reliance on empirical estimation of beta-value distributions is incompatible with the dynamic nature of cfDNA. Some epigenetic clocks have been designed with data adaptability in mind by selecting features robust to tissue-specific variance [55], but they are unlikely to generalize well for pan-tissue age prediction. Deep learning-based clocks [15,56] may achieve similar adaptability when trained on sufficiently comprehensive datasets, but cfDNA methylation data generated for age prediction remain limited. Principal component analysis (PCA) preprocessing has been widely applied and recognized for improving clock reproducibility in both array- and HTS-based datasets [22,42,57], albeit with a trade-off of reduced interpretability [55]. In summary, a comprehensive and robust adaptation of array-based epigenetic clocks for HTS-based data, particularly for cfDNA samples, is urgently needed.

Motivated by these knowledge gaps, we generated a dataset specifically designed for comparison between cfDNA/gDNA sample types and HTS/methylation array technologies. This dataset

enabled systematic benchmark of reproducibility and accuracy of array-based epigenetic clocks on HTS-based cfDNA methylation data. Furthermore, we delved into exploration of experimental and data aspects that can affect epigenetic accuracy and reproducibility. Finally, we developed and tested multiple improvement techniques according to these experimental and computation aspects, seeking experimental guidelines and data machine learning strategies to adapt epigenetic clocks for cfDNA-based applications. Overall, this study contributes to the adaptation of epigenetic clocks for cfDNA methylation data, paves the road for future minimally invasive aging evaluation.

Results

Methylation profiling technologies and epigenetic clocks exhibited heterogeneous CpG coverage

To evaluate the adaptability of array-trained epigenetic clocks for HTS-based methylation data, we compared four profiling methods: two array-based methods—the Illumina Human Methylation Screening Array (MSA) and the Illumina Methylation EPIC v2.0 BeadArray (EPICv2)—and two targeted sequencing-based methods—the iGeneTech BisCap CpG Galaxy Panel (Galaxy) and the Twist Human Methylome Panel (Twist). These methods were selected to represent both established and emerging methylation profiling technologies (Methods). For comparative analysis, 24 out of the 165 participants in the SRRSH were randomly selected (the SRRSH-24 cohort) and profiled using all four methods on both plasma cfDNA and buffy coat gDNA, yielding six datasets: gDNA_MSA, gDNA_EPICv2, gDNA_Galaxy, gDNA_Twist, cfDNA_Galaxy, and cfDNA_Twist. Two technical replicates were included in each of the six datasets for assessment of reproducibility. The remaining 141 participants (the SRRSH-141 cohort) were reserved for validation. Fig. 1a shows the overall study design and workflow.

The four methods (MSA, EPICv2, Galaxy and Twist) cover 0.28 to 4 million CpGs, among which 136,437 CpGs are common to all (Fig. S1a). Various degrees of feature coverage (Fig. S1b) of the 53 epigenetic clocks were observed (Methods, Table S1). Twist showed the highest overall CpG coverage. Although Galaxy and EPICv2 share similar theoretical reference coverage due to overlapping CpG sets by design, practical EPICv2 coverage was reduced by quality

control filtering, while Galaxy's coverage was slightly elevated due to off-target reads. Notably, MSA exhibited limited coverage (< 70 %) in 25 of 53 clocks, potentially limiting its utility.

After filtering for first- or second-generation clocks applicable to human adults, we identified two high-coverage clock categories: Hi-Cov-All comprising 20 clocks with ≥ 80 % coverage across all methods, and Hi-Cov-nonMSA comprising another 12 clocks with ≥ 80 % coverage across methods excluding MSA (Fig. S1b, Table S1). Six clocks (pchannum, pchorvath2013, pcphenoage, pcskinandblood, pcgrimage and zhangblup) were excluded from Hi-Cov-nonMSA due to their distinct age prediction performance as discussed later.

We next focused on two sets of shared CpG loci: 136,437 Method-shared sites, interrogated by all four profiling protocols, and 2,727 Clock-shared sites included in all Hi-Cov-All clock models. PCA of both CpG sets revealed distinct clustering among the six SRRSH-24 datasets, indicating both DNA type and profiling methods contributed to dataset-level variations, while minimally to replicate variations (Fig. 1b-c, Fig. S1c-d). Hierarchical clustering analysis (HCA) indicated that Clock-shared CpGs captured DNA-type information more effectively than Method-shared CpGs (Fig. S1e-f). implying that proper selection of CpGs contributes to clock robustness against method-related variances.

Together, these results highlight that CpG coverage is a critical factor for epigenetic clock adaptation across DNA types and methylation profiling methods.

Reduced CpG consistency and ICC posed challenges in HTS-based age prediction consistency and reproducibility

An ideal methylation profiling method required both reproducibility (agreement between technical replicates) and consistency (agreement across other methods). This motivated us to assess the 6 SRRSH-24 datasets to compare reproducibility and consistency across DNA types and methylation profiling methods.

For reproducibility, we calculated per-CpG intraclass correlation coefficients (ICCs) [58] across datasets, which quantify and compare intra- versus inter-group variance. As expected, HTS-based datasets (Galaxy and Twist) exhibited greater replicate differences (RD) (Fig. S1g) and lower ICCs compared to array-based datasets (MSA and EPICv2) for both Method-shared CpGs (Fig. 1d) and Clock-shared CpGs (Fig. S1g-h). Consistent with a previous study [22], CpGs with

low ICCs were enriched near extreme beta-values (near 0 or 1) regardless of methylation profiling method (Fig. S1g, S2a&c), and lower ICC values correlated with lower standard deviations (SD) of beta-values (Fig. S2b&d). In addition, the reduced ICCs in HTS-based datasets were likely independent to sequencing depth (Fig. 1e), despite the marginal effects of depth filtering on beta-value medians and SD (Fig. S2e-f).

For consistency, we calculated pairwise Pearson's correlation coefficients between datasets using the beta-value means of Method-shared CpGs (Fig. 1f). Array-based datasets exhibited very high correlation (R = 0.9947), whereas the correlations between HTS-based datasets were notably lower (R = 0.9125 - 0.9781). A similar pattern was also observed in Clock-shared CpGs (Fig. S1i).

These findings suggested that HTS-based datasets in general exhibited lower consistency and reproducibility compared with array-based methods.

CpG reproducibility and consistency influenced clock reproducibility and consistency

As shown in Fig. 1g-h, lower CpG ICCs in HTS-based datasets corresponded to reduced ICCs in predicted ages (0.469 - 0.999, median 0.950 for array-based datasets vs. -0.188 - 0.997, median 0.721 for HTS-based datasets), consistent with prior studies linking ICC to CpG reliability [22,42]. Six first- and second-generation clocks, including five PCA-preprocessed clocks [22] and zhangblup [21], showed markedly higher reproducibility (ICC ≥ 0.9) and were therefore designated as Hi-Repro (Fig. 1g, Fig. S3). Compared with Hi-Cov-All and Hi-Cov-nonMSA clocks, Hi-Repro clocks also exhibited significantly lower age prediction differences between replicates (Fig. S4). Besides reproducibility, age prediction consistency also associated with CpG consistency across datasets (Fig. 1f&i). Surprisingly, high age prediction reproducibility between gDNA datasets was only observed with Hi-Repro clocks while not with other clock categories (Fig. 1i), indicating that Hi-Repro clocks are more consistent and resilient against method-related methylation variations. Notably, 17 out of the 53 clocks (including all Hi-Repro clocks) exhibited Pearson's R ≥ 0.7 (Fig. S5), with pcgrimage achieving the highest.

Collectively, these results confirm that CpG reproducibility and consistency critically impacts clock prediction reproducibility and consistency. Amongst, Hi-Repro clocks are preferable than others for applications with HTS-based datasets for their superior reproducibility and consistency.

HTS-based dataset proposed challenges in age prediction accuracy

The Hi-Repro, Hi-Cov-All, and Hi-Cov-nonMSA clocks were comprehensively evaluated for age prediction accuracy across the six SRRSH-24 datasets. The gDNA_EPICv2 dataset generally exhibited the lowest mean absolute error (MAE) (Fig. 2a-b, S6a-b), consistent with these clocks being primarily trained on gDNA data obtained via methylation arrays. CfDNA samples consistently exhibited higher MAEs than gDNA samples, particularly with Hi-Cov-All clocks (Fig. 2c-d), emphasizing the necessity of clock adaptation between dataset derived from different DNA types or profiling methods. This consisted with a previous study which reported reduced reliability and accuracy when applying array-trained clocks on HTS-based data [42], though it only analyzed a few clocks. Among all clocks, zhangblup consistently demonstrated the lowest (MAE ≤ 5.2 years) across datasets except for gDNA_MSA (Fig. 2a, S6b), likely because of its poor CpG coverage (25%) on the MSA platform (Fig. S1b).

To test the reverse scenario, i.e. if HTS-trained clocks also suffer reduced accuracy on array-based data, we incorporated two independent whole-blood gDNA datasets: GSE245628 (N = 24, Illumina MethylationEPIC array) and GSE232346 (N = 24, HTS-based) [18]. As shown in Fig. 2e-f, the 4 array-trained clocks (hannum, horvath2013, skinandblood, dnamphenoage) and one HTS-trained clock (timeseq) displayed opposite accuracy patterns. This observation indicates that accuracy loss occurs both when applying array-trained clocks on HTS data, and also when applying HTS-trained clocks on array data [30,33,46].

In summary, these comparisons highlight the importance of matching data type and clock design, particularly regarding DNA source and measurement technology (array- vs. HTS-based). These findings further underscore the need to adapt epigenetic clocks for use across different DNA types and measurement platforms.

Epigenetic clock reproducibility indicates performance

Given the superior performance of Hi-Repro clocks compared with other clocks (Fig. 2a-d, Fig. S6a), we next investigated the relationship between clock reproducibility (ICC) and accuracy (MAE), or correlation with age (Pearson's R). Significant correlations between ICC and both metrics suggested that clock reproducibility serves as a practical indicator of epigenetic clock performance (Fig. 2g-h).

Age prediction RD is negatively associated with number of CpGs in linear regression-based epigenetic clocks

Noticing that all clocks in the Hi-Repro category had high numbers of input CpGs (≥ 70,000), we investigated 20 linear-architecture clocks from the Hi-Cov-All and Hi-Repro categories (collectively termed Hi-Cov/Repro-Linear, or HCRL clocks for short) to explore the relationship between CpG count and reproducibility. Detailed analysis revealed a negative association between the number of CpGs and age prediction RD among HCRL clocks, consistent across all 11 tested datasets, including the six SRRSH-24 datasets and five publicly available ones (GSE83944, N = 48; GSE247193, N = 22; GSE247195, N = 22; GSE247197, N = 46; GSE55763, N = 76, replicates only) (Table 1, Fig. S7).

These results align with earlier findings showing that clocks with more CpGs exhibit greater robustness to replicate variance, consistent with previous studies promoting PCA-preprocessed clocks [22,42]. These findings motivated in-depth investigation of how epigenetic clock model complexity contributes to clock reproducibility, which can inform the development and selection of more reliable epigenetic clocks.

L2 regularization strength contributed to low RD in independent testing

We trained a series of elastic net models with varying CpG counts to further examine the relationship between CpG count and RD. The elastic net framework was selected for its widespread use in epigenetic age modeling. The variation in CpG count was achieved by tuning the two hyperparameters: $\alpha \in [0, +\infty)$ (the overall regularization strength) and $\lambda \in [0,1]$ (the mixing parameter that balances L1 and L2 penalties). All models were trained on GSE55763 (N = 2711, all samples) and tested on the 6 SRRSH-24 datasets.

As shown in Fig. 3a, α exhibited a threshold effect around 0.01, beyond which the MAE increased sharply. We therefore focused on $\alpha \leq 0.01$ in subsequent analyses to better capture model optimization behavior. The negative correlation between CpG count and RD was consistently reproduced in all test datasets (Fig. 3b), confirming that this trend is intrinsic to elastic nets rather than specific to the Hi-Repro clocks [22].

Results revealed that λ had a stronger and more significant association (Pearson's R^2 = 0.67 - 0.78, $p = 1.6\times10^{-47}$ - 7.3×10^{-33}) than α (Pearson's R^2 = 0.05 - 0.21, $p = 9.7\times10^{-9}$ - 8.2×10^{-3}) (Fig.

3c-d). Notably, the positive correlation between λ and RD indicates that skewing towards L2 penalty (relative to L1) reduces RD. This observation partially explains the negative association between CpG count and RD, as L2 penalty (ridge regression) tends to retain a large number of small coefficients, while L1 (LASSO) favors sparse clocks.

Conclusively, these results suggest that elastic net models emphasizing L2 regularization (approaching the limiting case of ridge regression) enhance epigenetic clock reproducibility against replicate-level variability.

HTS-based methylation datasets are more prone to unreliable beta-values than methylation array datasets

Previous studies and our earlier findings both suggested that CpGs with beta-values near 0 or 1 are unreliable, as reflected by their low ICCs. We further found that HTS-based methylation datasets are particularly susceptible to their occurrence. Across eight tested datasets in Fig. S8, 18.4 - 61.3 % of CpGs (median 32.9 %) exhibited beta-values exactly equal to 0 or 1 in at least one sample. Approximately 1/3 to 3/4 of them were associated with low sequencing depth (or missing as indicated by 0 depth), except for the Buccal (Shokhirev and Johnson) dataset where low-depth (≤10× coverage) CpGs were explicitly filtered [42].

Further analyses revealed that these extreme beta-values are likely non-biological artifacts, as they displayed distinct depth distributions compared with CpGs having non-extreme values (Fig. S9-S10). The observation that a significant proportion of these CpGs were dataset-wise missing further supported this hypothesis (Fig. S11). Importantly, these CpGs overlapped with 2.0 - 34.7 % of the features in the 20 HCRL clocks (Fig. 4a, Table S3), although their impact on clock accuracy remains unclear. These observations highlight the widespread presence of artifactual beta-values in HTS-based data and motivated us to explore strategies for evaluating and improving model performance in the light of these beta-values.

Imputation on unreliable beta-values improved model prediction reproducibility

Multiple imputation strategies to correct unreliable beta-values were tested at varying cut-off depths ranging from 0 to 20×. The cut-off depth defines the depth threshold below which beta-values are treated as artifacts subject to imputation.

We first tested the K-nearest neighbor (KNN) method, one of the most widely used imputation approaches. Its effects of imputation on HCRL clocks were assessed by MAE and RD changes, where positive values indicate deterioration and negative values improvement. Results in Fig. 4b showed distinct imputation responses between cfDNA and gDNA. For cfDNA, MAE significantly increased by approximately 1 - 5 years, whereas no significant change was observed in gDNA, regardless of the capture panel (Twist or Galaxy). In contrast, RD improved consistently in both cfDNA and gDNA datasets across all cut-off depths (Fig. 4c), with larger reductions observed in cfDNA (0.7 - 2.1 years) compared to gDNA (0.7 - 1.1 years). These RD improvements likely stem from correcting CpGs with unreliable beta-values in one replicate but not in the other.

We further compared other common imputation techniques: mean, median, and 0-filling at 20× cut-off (Fig. 4d-e). Mean, median, and KNN imputation produced comparable outcomes, resulting in significant RD improvements, significant MAE increases in cfDNA, but only minimal, statistically non-significant MAE change in gDNA. This suggests these three techniques modify data in similar ways. In contrast, 0-filling caused no significant change in either metric, coincide with a previous report [42]. While performance change varied across clocks (Fig. S12), zhangblup again was identified as the best performer, showing improvements across four datasets.

Collectively, these results indicate that KNN, mean, and median imputations serve as conditional strategies. They can reliably improve RD with a potential cost of MAE deterioration in cfDNA samples. This approach works best with gDNA samples without compromising MAE.

Transfer learning significantly reduced age prediction MAE

To assess the adaptation of epigenetic clocks to HTS-based methylation data, we applied distillation as a transfer learning technique. Distillation was chosen for its accessibility, as it can be implemented for nearly any existing or future epigenetic clock without requiring access to the original algorithm, parameters, or training data. In this study, the original model (teacher) outputs were incorporated as an additional feature to train an elastic net model (student), which was then used for age prediction and validation. 6 classical clocks were tested as teachers, covering first- and second-generation clocks, PCA-preprocessed models, and the overall top performer identified in Fig. 2a.

Results in Fig. 5a demonstrated substantial MAE reduction for all six teacher models after applying distillation. The effect on RD, however, varied by teacher model (Fig. 5b). Nonetheless, the varying range of mean RD across clocks decreased, suggesting that distillation tends to normalize RD performance, improving clocks with high initial RD (e.g., dnamphenoage) while reducing those with low initial RD (e.g., zhangblup).

Overall, distillation consistently improved MAE, but its effect on RD was heterogeneous—some clocks showed improved RD, while others exhibited slight deterioration. Therefore, distillation is a highly effective strategy for enhancing accuracy, but its implementation should be carefully considered based on the specific reproducibility requirements of the application.

Ensuring epigenetic clock accuracy and reproducibility requires 10× minimum, 20× recommended mean target depth

Compared with array-based data, sequencing depth introduces an additional source of variance in HTS-based methylation data, making it essential to evaluate its effects on clock accuracy and reproducibility. Two mechanisms were examined: (1) The stochasticity in observing different compositions of methylated/unmethylated states even the intrinsic beta-value is fixed, using a binomial model; and (2) the resolution limit of beta-values constrained by finite sequencing depth. The latter illustrated the maximum precision to represent the beta-value of a specific CpG under finite sequencing depth even no stochastic effect is considered. Two sequencing depth thresholds at approximately 20× and 10× were identified respectively, above which MAE and RD stabilized around their baseline value in a given dataset (Fig. S13).

To validate these findings under more practical conditions, we conducted randomized downsampling tests on five high-depth samples from the SRRSH-141 cohort (with ≥ 50 Gb). MAE convergence was achieved at approximately 10-20× coverage (Fig. 6a) across six representative clocks (selected with the same criteria as teachers in the previous section) as well as 32 others (Fig. S14). Notably, results in Fig. 6b and Fig. S14 demonstrated that fraction of CpGs with ≥ 20× coverage increased rapidly under 20 - 30× downsampling depth, the increasing slowed thereafter. Taken together, these results suggest a minimum mean target depth of 10×, optimally ≥ 20×, for HTS-based epigenetic age prediction. All our 4 SRRSH-24 HTS-based

datasets and the SRRSH-141 dataset generated in this study comply with this standard (Fig. S15). In addition, this standard is comparable with a previous recommendation for WGBS, recommending 2 replicates with combined 30× coverage [59].

Validation of transfer learning as a promising tool for epigenetic clocks adaptation

Finally, we evaluated all eight possible combinations of three proposed strategies for adapting epigenetic clocks to HTS-based data: depth-based filtering (DF), unreliable beta-value imputation (IM), and transfer learning (TL), applied sequentially in DF, IM, TL order. These combinations were tested across three cohorts, SRRSH-141 (N = 141), GSE186458 (N = 23), and the Buccal cohort (N = 100). TL-along and IM + TL consistently outperformed other approaches across the 20 HCRL clocks (Fig. 6d). In the six representative classic clocks, this approach improved MAE by 1.0 - 12.0 years and Pearson's R by 0.03 - 0.78 (Fig. 6c). Similar trends were observed in GSE186458 (Fig. S16) with 20 HCRL clocks. In both datasets, all TL-based strategies significantly reduced MAE. In the Buccal cohort, which included technical replicates, TL markedly lowered MAE but slightly increasing RD (< 0.5 years on average; Fig. 6e, Fig. S17). Collectively, these findings underscore transfer learning as an promising tool for adapting classic epigenetic clocks to HTS-based data.

Discussion

In this study, we generated a unique, comprehensive dataset encompassing both array- and HTS-based methylation data. Public datasets including both replicate measurements and cross-platform profiling are currently scarce. While Shokhirev et al. [42] focused on buccal gDNA, our dataset extended assessments regarding different DNA types and methylation profiling technologies. The six SRRSH-24 datasets generated in this study expand the limited resources available for systematic comparisons between array- and HTS-based methylation technologies. Moreover, they contribute valuable replicate data critical for evaluating the reproducibility of methylation profiling methods and epigenetic clocks.

Imperfect reproducibility of epigenetic measurements inherently limits the accuracy of age prediction by epigenetic clocks. Conventional accuracy metrics such as MAE or RMSE

aggregate two sources of variation: (1) the intrinsic quantitative relationship between methylation features and biological/chronological age, and (2) the stochastic effects specific to sample measurements, Consequently, these metrics already encode reproducibility to some extent. Although these two components can be mathematically modeled, the precise biological disentanglement and quantification of their respective contributions remain challenging. Coupled with finite sample size, this modeling imperfection imposes an upper bound on the practical reproducibility of a clock, and thereby, an upper bound on its achievable predictive accuracy.

Our results show that HTS-based data pose additional challenges for epigenetic clocks as the suboptimal reliability and reproducibility of the HTS input signal can attenuate both the accuracy and reproducibility of the age predictions. In addition, we found that L2 regularization in elastic net clocks empirically enhances reproducibility. Related approaches, including PCA-preprocessed clocks and zhangblup (see Supplementary Discussion), achieved mean RD as low as 1.6 years across HTS-based datasets (Fig. S3, Fig. S4a). We therefore recommend L2 regularization as a practical and promising strategy for improving reproducibility of linear-architecture clocks.

HTS technologies utilize distinct signal processing principles compared to methylation arrays [37,60,61], which can affect CpG-level reliability and reproducibility. Among tested adaptation strategies, imputing unreliable beta-values consistently improved clock reproducibility, but causing accuracy loss particularly in cfDNA samples. In contrast, gDNA samples were minimally affected by this accuracy loss, likely due to their higher cellular homogeneity, a factor known to enhance imputation accuracy [41]. Although not tested, these findings suggest that performing cell-type deconvolution prior to imputation could mitigate the MAE trade-offs observed in cfDNA samples. Additionally, statistical modeling and downsampling experiments suggest $\geq 10\times$ mean HTS target depth (ideally $\geq 20\times$) for epigenetic clock applications. These recommendations, comparable to previous WGBS guidelines [59,62], aim to support more robust and standardized implementations of epigenetic clocks.

Finally, our systematic evaluation of multiple adaptation strategies (DF, IM and TL) demonstrated that TL offered the most significant improvements of epigenetic clock accuracy. This effect may stem from the correction of systematic pattern differences observed between DNA types and methylation profiling methods (Fig. 1b-c, Fig. S1c-d). While previous studies

have focused on comparability between methylation arrays [63–65], our results extend these observations to different capture panels and gDNA/cfDNA types. Similar efficacy of TL has been demonstrated in a recent study [54]. Distillation, the TL method tested in this study, shares characteristics with error correction and model recalibration techniques. A recent study demonstrated that model retraining and recalibration can compensate for the missing probes in array data [49], and our study effectively extends this utility to HTS-based data. We also observed that DNA types or its origin—specifically cfDNA, buccal, and buffy coat gDNA—influences clock predictions regardless of training platform [9,33,66]. These findings underscore future development and refinement of epigenetic clock adaptation methods for cfDNA applications.

Our study has several limitations. The type and size of our datasets prevented assessment of aging biomarkers beyond methylation and limited exploration of novel epigenetic clocks [14,16,67]. Due to the limited sample size and lack of data, no sex- or ethnicity/race-related effects was evaluated. Besides, all datasets were cross-sectional without longitudinal and intervention-oriented analysis, precluding insights into temporal methylation dynamics. Lacking these analyses may hamper the generalization of results found this study, although our dataset sizes are comparable among studies with similar goals. Primarily, we focused on linear relationships between methylation data and clock predictions to balance result interpretability and model complexity. While linear models can achieve high accuracy [9,12,19,21,22], the true biological relationships between methylation and aging may be complex. Dataset size also limited our ability to investigate tissue heterogeneity in cfDNA, which is important for validating our hypotheses about imputation accuracy. For transfer learning, we only tested distillation under the consideration of minimal requirements of prior knowledge about model structure, parameters, and original training data; consequently, alternative methods such as regularization sharing [68], transfer component analysis [69], or deep learning-based approaches [70] remain unexplored. In addition, PCA preprocessing prior to TL was applied to improve training efficiency and reduce overfitting, though it may independently influence TL performance [22,42,71]. Finally, the effects are not adjusted by confounders such as lifestyle factors due to the lack of related information, which may limit the explanation of mechanistic insights.

Overall, our study systematically compared the characteristics of HTS-based and array-based methylation data using dedicated datasets. We tested, and validated three machine learning

techniques to adapt epigenetic clocks for HTS data, thereby demonstrating these clocks' broad applicability with appropriate tuning. We also provided crucial technical recommendations and experimental guidelines, significantly advancing the field for future HTS-based and cfDNA-oriented epigenetic clock applications.

Methods

The SRRSH cohorts and ethics declaration

The ethics committee of the Sir Run Run Shaw Hospital (SRRSH, affiliated with Zhejiang University School of Medicine) approved this study (approval number 2024-0547). Written informed consent was obtained from all participants.

The recruitment and data collection for all volunteers in the cohort were conducted under strict protocols. The following criteria were applied for recruiting eligible volunteers: (1) healthy adults aged 20 - 84 years; (2) Within the past 5 years the subject has completed at least two routine health evaluations that included complete blood count, urinalysis, fasting plasma glucose, lipid profile, liver-function tests, renal-function tests, and at least one set of imaging studies: abdominal ultrasound (liver, gall-bladder, pancreas, spleen, both kidneys), chest radiograph, thyroid ultrasound, prostate ultrasound (men), uterus and adnexa ultrasound (women). All reported results must be within normal limits and free of any diagnosed disease; (3) no smoking history within 5 years before screening and agreement to remain abstinent from tobacco for 1 year after consent; (4) body-mass index 18.5-28.0 kg/m^2; (5) without acute infection or active inflammation within 14 days before screening; (6) without cancer-related signs or symptoms (e.g., hemoptysis, rectal bleeding, altered bowel habit, hematuria, dysphagia, palpable mass) within 30 days before screening; (7) without previous organ transplantation or any non-autologous bone-marrow or stem-cell transplantation; (8) without blood transfusion within 7 days before screening; (9) without use of therapeutic drugs for any illness (nutritional supplements excluded) within 30 days before screening; (10) without any surgical procedure within 5 years before screening; (11) pregnant or breastfeeding women were excluded.

Methylation array and capture panel selection

Four methylation profiling methods were included in this study, including two array-based methods and two HTS-based methods with different hybridization probe capture panels: Illumina Human Methylation Screening Array (referred to as MSA), Illumina Methylation EPIC v2.0 BeadArray (referred to as EPICv2), targeted sequencing with iGeneTech BisCap CpG Galaxy Panel for capturing (referred to as Galaxy) and targeted sequencing with Twist Human Methylome Panel for capturing (referred to as Twist). EPICv2 was included because it encompasses the largest CpG set among currently available methylation arrays, which serve as the basis for training most existing epigenetic clocks. MSA is a newer methylation array with a smaller CpG set than EPICv2, yet is claimed by the manufacturer to be suitable for epigenetic clock applications. Galaxy was included because it is designed as an alternative to EPICv2, covering over 95% of its CpG sites. Lastly, Twist was included as a mature solution for HTS-based methylation profiling with appropriate CpG coverage in the current marketplace. Overall, these methods were selected to cover mature, widely used, and emerging methods in both methylation array and HTS categories.

Experimental design

In total, 165 participants (100 % East Asian / Pacific Islander; aged 20 - 77 years old; 39.7 % female) were recruited as the SRRSH cohort. Among which, 24 subjects were randomly selected (22 - 77 years old; 54.2 % female) for investigating and evaluating the adaptation of epigenetic clocks for cfDNA-based methylation data, referred to as the SRRSH-24 cohort. Though no sample size calculations were conducted *a priori*, a previous study suggested a similar sample size is sufficient to identify 80% of reliable probes [64]. Their cfDNA was extracted for targeted DNA methylation sequencing, and buffy coat gDNA was extracted for both methylation profiling. This resulted in six datasets: gDNA_MSA, gDNA_EPICv2, gDNA_Galaxy, gDNA_Twist, cfDNA_Galaxy, and cfDNA_Twist. Two technical replicates per sample were generated via independent library preparations from the same DNA aliquot. At last, the SRRSH-141 dataset contained Twist-profiled buffy coat gDNA methylation data for the remaining 141 participants.

Epigenetic clock selection

A total of 53 epigenetic clocks were included in this study (Table S1). Although we comprehensively surveyed available epigenetic clocks, some clocks were excluded for their

incompatibility with our data. We used 47 clocks integrated into the `pyaging` package (version 0.1.27) [72] for convenience, which greatly simplified multiple clock predictions due to a uniform interface. Six GP-age clocks (https://github.com/mirivar/GP-age, version 1.0) were also included. In addition, TIME-seq (https://github.com/patricktgriffin/TIME-Seq, Commitment 140622f), was included as a representative of clocks trained by HTS data [18].

Biological sample pretreatment and storage

An 8-10 mL whole-blood sample was collected from each volunteer via intravenous collection , after an overnight fast. One anticoagulant tube (Streck, Cell-Free DNA BCT 10ml) was immediately transported to the laboratory for subsequent treatment. The blood was centrifuged at 4 ˚C (1900× g, 10 min) to obtain plasma and buffy coat. All biological samples were stored at -80 ˚C for long-term preservation.

DNA extraction

Cell-free DNA was extracted from 1.0 - 1.5 mL plasma using the QIAamp MinElute ccfDNA Mini Kit (Qiagen, Cat# 55204) following the manufacturer's instructions. Genomic DNA was isolated from approximately 500 μL of thawed buffy coat using TIANamp Genomic DNA Kit (TIANGEN, Cat# DP304) following the manufacture's protocol. The concentration of purified DNA was quantified using a Thermo Fisher Qubit 4.0 fluorometer (Qubit™ dsDNA HS Assay Kit, Cat# Q32854).

Library preparation and sequencing

For genomic DNA, 0.5% (by mass) of unmethylated lambda DNA and 0.05% (by mass) of CpG-methylated pUC19 DNA were spiked into 400-800 nanograms of buffy coat genomic DNA, followed by dilution to 130 μL of 10 mM Tris 0.1 mM EDTA, pH 8.0, following the manufacturer's recommendations. Internal control DNA underwent the same fragmentation procedure as gDNA samples. The DNA mixture was then transferred to a Covaris microTUBE and sheared to an average size of 100-200 bp (peak = 150bp) using a Covaris™ S220 focused ultrasonicator (Covaris, MA, USA).

For cfDNA, unmethylated lambda DNA and CpG methylated pUC19 DNA were respectively sheared to an average size of 100-200 bp (peak = 150bp) in advance. Next, 0.5% (by mass) of

fragmented unmethylated lambda DNA and 0.05% (by mass) of fragmented CpG methylated pUC19 DNA were added into 1-10 ng cfDNA samples.

Both cfDNA and sonicated gDNA samples, along with spike-in internal control DNA, were both used to construct EM-Seq libraries with the NEBNext Enzymatic Methyl-seq Kit (New England Biolabs, Cat# E7120) following the manufacturer's instructions. Libraries were quantified using a Thermo Fisher Qubit 4.0 fluorometer. Then capture-based human methylome enrichment was performed according to Twist's (South San Francisco, USA) protocols for their Twist Human Methylome Panel (Cat# 105521), or according to iGeneTech's (Beijing, China) protocols for their BisCap CpG Galaxy Panel (Cat# PE4000175). The final libraries were analyzed and quantified using TapeStation (Agilent Technologies). The whole-genome libraries were 150 bp paired-end sequenced using the Illumina NovaSeq X Plus sequencer on a 10B or 25B flow cell (Illumina) with 12.5 % PhiX spiked-in, to an average yield of 36.3Gbp raw base per library (Fig S15).

EPICv2 and MSA

Buffy coat genomic DNA samples were sent to Novogene (Tianjin, China) to perform Infinium MethylationEPIC v2.0 array (EPICv2), or sent to WeGene (Shenzhen, China) to perform Infinium Methylation Screening Array (MSA). For each technical replicate sample, 500 nanograms (for EPICv2) or 200 nanograms (for MSA) of genomic DNA were bisulfite treated using the EZ DNA Methylation Kit (Zymo Research, Cat# D5001) following the manufacturer's recommendations for Infinium assays. To assess the methylation status of the CpG sites, the Illumina Infinium MethylationEPIC v2.0 BeadChip array (937,691 probes) and Infinium Methylation Screening array (284,318 probes) was used.

As the EPICv2 and MSA arrays have a capacity of 8 and 48 samples per array, respectively, we included two replicates (2 replicates × 24 gDNA samples = 48 positions).

Methylation array data processing

Raw IDAT files from both Infinium MethylationEPIC v2.0 array and Infinium Methylation Screening Array (MSA) were processed using the `minfi` R package (version 1.54.1) [73] according to established protocols. A rigorous quality control and filtering pipeline was applied: probes located on sex chromosomes were excluded, as well as those with a detection p-value

greater than 0.01 regarding as unreliable signals. Furthermore, probes known to harbor single nucleotide polymorphisms (SNPs) at or near the CpG site, or those that did not uniquely map to the reference genome, were also removed. Following these filtering steps, the raw intensity data underwent intra-array normalization using the `Noob` [74] method. Finally, beta-values were calculated as the primary methylation metric and prepared for downstream quantitative analyses.

High-throughput sequence (HTS) methylation data processing

Prior to downstream processing, raw paired-end FASTQ files for all samples were downsampled to standardize sequencing depth at approximately 30 Gbp (1 Gbp = 10^9 base pairs). This step was performed to mitigate potential biases stemming from varying sequencing depths and to ensure fair comparisons among samples. Subsequently, these trimmed files underwent adapter removal using `fastp` [75] (version 0.23.4) and were aligned to a comprehensive reference genome, comprising GRCh38, pUC19, and phage lambda sequences, using `bwa-meth` [76] (version 0.2.7). Alignment files were then sorted and indexed using `Samtools` [77] (version 1.21), followed by the identification and marking of duplicate reads using `Samblaster` [78] (version 0.1.26). Methylation levels across CpG contexts were determined and extracted using the mbias and extract modules of `MethylDackel` (version 0.6.1).

Genomic coordinate alignment and probe annotation

Manifest files for the MSA, HM450K, EPICv1, and EPICv2 arrays were obtained from the manufacturers' official websites. Since our analysis was based on the hg38 reference genome, genomic coordinates for the HM450K probes were converted from hg19 to hg38 using the UCSC `LiftOver` tool. The four manifest files were then merged to create a comprehensive probe-to-coordinate reference file.

For sequencing data (Twist and Galaxy), the per-sample output files from `MethylDackel` (https://github.com/dpryan79/MethylDackel, version 0.6.1) were processed. Genomic coordinates and beta-values were extracted for each CpG sites and then merged using `mergeContext` function in `MethyDackel`. These files were subsequently merged into a matrix using `bedtools unionbedg` (version 2.31.1). Finally, the `bedtools intersect` function was employed to annotate the sequencing-derived CpG sites with corresponding probe IDs from the reference file, based on overlapping genomic coordinates.

Principal component, correlation, and consistency analyses

All analyses were performed using R (version 4.5.1).

Principal Component Analysis (PCA): PCA was performed on beta-value matrices comprising either all intersecting CpGs across platforms or the 2,727 CpG sites of the clock models. CpGs containing missing values (`NaN`) were removed. The `prcomp` function was used for PCA calculation, and the first two principal components were visualized using the `ggplot2` package (version 4.0.0), with samples colored by dataset or replicate status.

Sample-to-sample correlation: Pairwise Pearson correlation coefficients were calculated between samples using the `cor` function on the beta-value matrices (either all intersecting CpGs or clock model CpGs). Correlation matrices were visualized as heatmaps using the `Heatmap` function from the `ComplexHeatmap` package (version 2.22.0).

CpGs-level consistency: For each CpG, the mean methylation value was calculated across all samples within a dataset. Pairwise Pearson correlations of these mean values between datasets were then computed and visualized using the `qcorrplot` function from the `linkET` package (version 0.0.7.4).

Clock prediction consistency: For clock models grouped by category, the mean predicted age was calculated for each sample within the group. Pairwise Pearson correlations of these mean predictions between datasets were computed. For individual clocks, the pairwise correlation was calculated directly from their prediction outputs. Results were visualized using `qcorrplot`.

Reliability assessment

The reliability of measurements was assessed using the Intraclass Correlation Coefficient (ICC) [58]. ICC estimates were calculated using the `icc` function from the `irrR` package (version 0.84.1), based on a single-rater, absolute-agreement, two-way random-effects model.

Sequencing depth filtering analysis

To assess the impact of sequencing depth on data quality, we applied a series of five minimum depth thresholds. For each threshold, we identified CpG sites where the sequencing depth was greater than or equal to the threshold in all samples. We then examined the distribution of mean

beta-values, standard deviations (SD), and ICC values for the CpG sites retained at each depth filter.

Analysis of CpG site overlap across platforms

Manifest files detailing the targeted CpG sites were downloaded from the respective manufacturers. For MSA and EPICv2 arrays, only CpGs with defined genomic coordinates were retained. The overlap of CpG sites across the four profiling methods (MSA, EPICv2, Galaxy, Twist) was visualized using an UpSet plot generated with the `UpSet` function from the `ComplexHeatmap` package (version 2.22.0).

Clock model CpG coverage analysis

The CpG sites comprising the GPage models were obtained from the model repository (https://github.com/mirivar/GP-age). The CpG sites for all other clocks were extracted from the model metadata provided by the `pyaging` package [72]. The percentage of covered CpG sites by each platform for each clock model was calculated and visualized as a heatmap using the `pheatmap` function.

Data visualization

Scatter plots presented in Figure S3, Figure 2g, and Figure S6b were generated using the ggplot2 (version 4.0.0) package in R. Correlation analysis and statistics displayed on the plots were computed using the `stat_cor()` function.

Statistical analysis

The comparisons between means and medians of two groups of data were conducted using permutation test (implemented by custom script) by randomly permuting observations between the two groups and comparing the mean/median after permutation with the observed value. The p-values are calculated as the fraction of permutation rounds with its mean/median after permutation being greater than the observed value. Permutation tests make no assumptions of distributions and can be applied to both small and large number of observations.

To account for multiple comparisons, the p-values were adjusted for false discovery rate (FDR) control using the Benjamini-Hochberg (BH) method

(`statsmodels.stats.multitest.multipletests(method="fdr_bh")` in the `statsmodels` package, version 0.14.5).

Public datasets

To ensure reproducibility and external validation, we included 10 publicly available datasets. GSE55763 (N = 2711, Illumina HumanMethylation 450K) [79] was downloaded from the Gene Expression Omnibus (GEO) database for model and PCA training for its large size, and was also used to assess age prediction RD for containing 38 subjects with replicates (76 samples in total). GSE83944 (N = 48, Illumina HumanMethylation 450K) [80], GSE247193 (N = 23 including one subject without replicate, Illumina Infinium EPIC Human Methylation Beadchip) [81], GSE247195 (N = 23 including one subject without replicate, Illumina Infinium EPIC Human Methylation Beadchip) [81] and GSE247197 (N = 46, Illumina HumanMethylation 450K) [81] were downloaded from the GEO database to assess age prediction RD as they all contain technical replicates. GSE144691 (N = 4) [82] was downloaded from the GEO database, but for assessment of unreliable CpGs in HTS-based datasets only due to its small sample size. In addition, 3 additional datasets from the GEO database were downloaded for assessment of age prediction MAE and RD, including GSE186458 (N = 23, WGBS) [83], GSE232346 (N = 24, Twist) [18] and GSE245628 (N = 24, Illumina Infinium EPIC Human Methylation Beadchip) [18]. Note that only 23 cfDNA samples in the GSE186458 dataset were accessed in this study. Finally, we also downloaded the Buccal cohort (Shokhirev & Johnson, 2025, N = 100, Twist and Illumina Infinium EPIC Human Methylation Beadchip v2) dataset [42] for various validation steps. The ethnicity and sex information can be found in the source dataset but was not considered in this study.

Correlation between clock number of CpGs and RD

20 HCRL clocks were selected to predict biological age with `pyaging` (version 0.1.27) [72] using the standard protocol as described on https://pyaging.readthedocs.io/en/latest/tutorials/tutorial_dnam_illumina_human_array.html. No additional data preprocessing was performed except for filling missing values (`NaN`) with KNN imputation. RD were calculated as the mean absolute difference between the predicted ages between replicates. The log-transformed RD was then regressed against log-transformed number of CpGs in each clock using `scipy.stats.linregress` in the `scipy` package (version

1.15.3). Result p-values were corrected for multiple tests via the Benjamini-Hochberg procedure using custom scripts.

Elastic net model training for independent verification of the correlation between number of CpGs and and RD

A series of elastic net models were trained on the dataset GSE55763 (N = 2711) with `sklearn.linear_model.ElasticNet` implemented in the `scikit-learn` package (version 1.6.1). The CpGs shared between GSE55763 and the 6 datasets (gDNA_EPICv2, gDNA_MSA, gDNA_Galaxy, gDNA_Twist, cfDNA_Galaxy and cfDNA_Twist) were extracted (d=90394), and missing values (`NaN`) were filled with KNN imputation implemented in `pyaging`. No additional data cleaning was performed. The elastic net models were trained with 30 logarithmically spaced α values ranging from 10^{-5} to 10^{2}, and 15 linearly spaced `l1_ratio` (λ) values ranging from 0.0-1.0, in a total of 450 trained models. All other parameters were kept default. Linear regression and the resulting R and p-values were calculated with `scipy.stats.linregress`.

Unreliable beta-value imputation at different cut-off depths

For each combination of selected cut-off depth and imputation method, the beta matrix in steps in the following order:

(1) Missing value imputation (first pass): Missing values (`NaN`) were imputed using the chosen method implemented in `pyaging.preprocess`. Briefly, the KNN imputation method replaces missing values with the mean of k nearest neighbors, where distances between samples are computed based on features with non-missing beta-values. The mean and median methods replace missing entries with the mean or median value of the respective feature, calculated from non-missing observations. The 0-filling method assigns 0 to all missing entries. Notably, among these methods, KNN is the only one that considers information across all features rather than only the imputed one.

(2) Unreliable beta-value removal: Selected beta-values were set to `NaN` when both of the following conditions were met: (a) the observation was 0 or 1, and (b) the sequencing depth at the corresponding CpG site was below the defined cut-off.

(3) Missing value imputation (second pass): The selected imputation method was re-applied to impute these newly removed unreliable CpG sites.

This process was repeated for cut-off depths ranging from 1 to 20 across all four HTS-based datasets generated in this study (gDNA_Galaxy, gDNA_Twist, cfDNA_Galaxy, and cfDNA_Twist).

Transfer learning of classic epigenetic models

We used distillation to adopt classic epigenetic models for HTS methylation data. The classic epigenetic models are teacher models. Each HTS dataset were first PCA-transformed using PCA models trained from public dataset GSE55763 (N = 2711) with an output dimensionality of 2500. The transformed dataset was then combined with the output of a teacher model to train and test a student elastic net model using a 5-fold cross-validation (5-fold CV) implemented in the `scikit-learn` package (version 1.6.1). The final evaluation of age prediction MAE and RD were summarized from all testing splits. The hyperparameters (`alpha` and `l1_ratio`) was optimized using an independent 5-fold cross-validation (`sklearn.linear_models.ElasticNetCV(cv=5)`).

Minimal sequencing investigation based on binomial stochasticity.

This investigation evaluates the uncertainty in epigenetic clock prediction caused by the stochastic effect in observed beta-values. This simulation assumes that the true underlying beta-value is fixed and identical across technical replicates, isolating the contribution of sequencing stochasticity and resolution to prediction variability. To simulate this mechanism, randomized datasets were generated via multi-dimensional binomial model, where the observed methylation counts at each CpG follows independent binomial distribution. For each binomial distribution, the probability parameter was set to the observed beta-value in the original dataset, and the count parameter set to the sequencing depth, given by:

$$beta \sim \frac{1}{depth}\, Binomial(depth, beta_{obs})$$

Where:

depth denotes the simulated sequencing depth under investigation.

$beta_{obs}$ is the empirically estimated true beta-value, calculated as the mean in a replicate pair. Value of $beta_{obs}$ differs across subjects and datasets.

To reduce randomness, 1,000 random datasets were generated at each simulated sequencing depth. Finally, clock prediction metrics (MAE and RD) were calculated with respect to each simulated sequencing depth.

Minimal sequencing investigation based on beta resolution.

This investigation evaluates the impact of uncertainty related to the finite sequencing depths, when beta-values cannot be represented at arbitrary precision. For example, a sequencing depth of 4 can only represent beta-values among 0, 0.25, 0.5, 0.75 and 1.0. To simulate this mechanism, the beta-values were rounded under a given sequencing depth using the following formula before age prediction:

$$beta = \frac{\left\lfloor beta_{obs} \times depth + \frac{1}{2} \right\rfloor}{depth}$$

Where:

depth denotes the simulated sequencing depth under investigation.

$beta_{obs}$ is the empirically estimated true beta-value, calculated as the mean in a replicate pair.

$\lfloor \cdot \rfloor$ is the floor function that maps a real number to the greatest integer less than or equal to that number.

This calculation rounds the beta-value to the closest rational number with depth being the denominator. Clock prediction metrics (MAE and RD) were then calculated with respect to each simulated sequencing depth. No repeated tests were necessary since this process is deterministic. By this analysis, we can evaluate the maximal epigenetic clock precision possible under each sequencing depth even if no stochastic effect is subject to consider.

Minimal sequencing investigation based on downsampling.

A downsampling analysis was conducted on five samples, each having an initial sequencing depth of 50Gb, to assess the effect of varying data volumes on biological age estimation. Raw sequencing data were downsampled at proportions ranging from 0.025 to 0.9 (namely 0.025,

0.05, 0.1, 0.2, 0.3, 0.4, 0.5, ..., 0.9). At each downsampling point, biological age was estimated using various established models.

Acknowledgement

Special thanks to the high-performance computing cluster provided by Liangzhu Laboratory, Zhejiang University for computation support.

Data availability

The de-identified data generated in this study, including processed methylation matrices, sequencing depth matrices and non-sensitive subject metadata are available at zendoro.org under accession 17453844 (DOI: 10.5281/zenodo.17453844). Raw data available at GSA for Human (https://ngdc.cncb.ac.cn/gsa-human/) under project accession PRJCA050670.

Code availability

The code related to this project are publicly accessible in at https://github.com/regen-bio/cfDNA-HTS-Clock-Analysis.git.

Author contributions

C.O. and G.Z. conceived and supervised the study. G.L., C.O., and G.Z. managed the overall project. Cohort management: L.C., X.C., Z.Y., D.T., X.F., X.H., and V.T. organized the cohort, collected human blood samples, and coordinated resources. X.H. and V.T. also provided essential medical and technical support. Data generation: C.O., Y.G., S.J., B.H., and C.T. performed sample collection, DNA extraction, and library preparation. Bioinformatics analysis: J.W. and W.H. developed the bioinformatics pipeline. J.W. preprocessed HTS and methylation array data and performed downsampling analysis. W.H. conducted reproducibility, consistency, and

accuracy analyses. G.L. investigated clock complexity, unreliable beta filtering, transfer learning, and combination strategies. X.Z. contributed to statistical and machine learning analyses, with advisory support from Y.W. Manuscript preparation: G.L., W.H., X.Z., J.W., and C.O. wrote the manuscript. All authors reviewed and approved the final manuscript.

Competing interests

W.H., X.Z., J.W., Y.G., Z.Y., S.J., B.H., D.T., V.T., C.T., X.F., C.O. and G.Z. are affiliated with Regenerative Bio Inc., a global biotech company powered by AI, providing organ-specific aging assessment models and longevity-focused health solutions. Regenerative Bio Inc. was the study sponsor. The other authors declare no competing interests.

Funding

This research received no grants from any funding agency in the public, commercial, or not-for-profit sectors.

References

1. Horvath, S. & Raj, K. DNA methylation-based biomarkers and the epigenetic clock theory of ageing. *Nat. Rev. Genet.* **19**, 371–384 (2018).
2. Moqri, M. *et al.* Biomarkers of aging for the identification and evaluation of longevity interventions. *Cell* **186**, 3758–3775 (2023).

3. Biomarkers of Aging Consortium *et al.* Challenges and recommendations for the translation of biomarkers of aging. *Nat. Aging* **4**, 1372–1383 (2024).

4. Levine, M. E. *et al.* An epigenetic biomarker of aging for lifespan and healthspan. *Aging* **10**, 573–591 (2018).

5. Sagers, L., Melas-Kyriazi, L., Patel, C. J. & Manrai, A. K. Prediction of chronological and biological age from laboratory data. *Aging* **12**, 7626–7638 (2020).

6. Blasco, M. A. Telomere length, stem cells and aging. *Nat. Chem. Biol.* **3**, 640–649 (2007).

7. Bobrov, E. *et al.* PhotoAgeClock: deep learning algorithms for development of non-invasive visual biomarkers of aging. *Aging* **10**, 3249–3259 (2018).

8. Li, J. *et al.* Determining a multimodal aging clock in a cohort of Chinese women. *Med* **4**, 825-848.e13 (2023).

9. Horvath, S. DNA methylation age of human tissues and cell types. *Genome Biol.* **14**, 3156 (2013).

10. Peters, M. J. *et al.* The transcriptional landscape of age in human peripheral blood. *Nat. Commun.* **6**, 8570 (2015).

11. Álvez, M. B. *et al.* A human pan-disease blood atlas of the circulating proteome. *Science* eadx2678 (2025) doi:10.1126/science.adx2678.

12. Teschendorff, A. E. & Horvath, S. Epigenetic ageing clocks: statistical methods and emerging computational challenges. *Nat. Rev. Genet.* **26**, 350–368 (2025).

13. Varshavsky, M. *et al.* Accurate age prediction from blood using a small set of DNA methylation sites and a cohort-based machine learning algorithm. *Cell Rep. Methods* **3**, 100567 (2023).

14. Ying, K. *et al.* MethylGPT: a foundation model for the DNA methylome. Preprint at https://doi.org/10.1101/2024.10.30.621013 (2024).
15. De Lima Camillo, L. P., Lapierre, L. R. & Singh, R. A pan-tissue DNA-methylation epigenetic clock based on deep learning. *Npj Aging* **8**, 4 (2022).
16. Galkin, F. *et al.* Precious3GPT: Multimodal Multi-Species Multi-Omics Multi-Tissue Transformer for Aging Research and Drug Discovery. Preprint at https://doi.org/10.1101/2024.07.25.605062 (2024).
17. Sehgal, R. *et al.* Systems Age: a single blood methylation test to quantify aging heterogeneity across 11 physiological systems. *Nat. Aging* **5**, 1880–1896 (2025).
18. Griffin, P. T. *et al.* TIME-seq reduces time and cost of DNA methylation measurement for epigenetic clock construction. *Nat. Aging* **4**, 261–274 (2024).
19. Tong, H. *et al.* Quantifying the stochastic component of epigenetic aging. *Nat. Aging* **4**, 886–901 (2024).
20. Haghani, A. *et al.* EnsembleAge: enhancing epigenetic age assessment with a multi-clock framework. *GeroScience* https://doi.org/10.1007/s11357-025-01808-1 (2025) doi:10.1007/s11357-025-01808-1.
21. Zhang, Q. *et al.* Improved precision of epigenetic clock estimates across tissues and its implication for biological ageing. *Genome Med.* **11**, 54 (2019).
22. Higgins-Chen, A. T. *et al.* A computational solution for bolstering reliability of epigenetic clocks: implications for clinical trials and longitudinal tracking. *Nat. Aging* **2**, 644–661 (2022).

23. Conole, E. L. S., Robertson, J. A., Smith, H. M., Cox, S. R. & Marioni, R. E. Epigenetic clocks and DNA methylation biomarkers of brain health and disease. *Nat. Rev. Neurol.* **21**, 411–421 (2025).
24. Klein, E. A. *et al.* Clinical validation of a targeted methylation-based multi-cancer early detection test using an independent validation set. *Ann. Oncol.* **32**, 1167–1177 (2021).
25. Moqri, M. *et al.* Validation of biomarkers of aging. *Nat. Med.* **30**, 360–372 (2024).
26. Gao, Q. *et al.* Circulating cell-free DNA for cancer early detection. *The Innovation* **3**, 100259 (2022).
27. Medina, J. E. *et al.* Cell-free DNA approaches for cancer early detection and interception. *J. Immunother. Cancer* **11**, e006013 (2023).
28. Zhang, K., Fu, R., Liu, R. & Su, Z. Circulating cell-free DNA-based multi-cancer early detection. *Trends Cancer* **10**, 161–174 (2024).
29. Bruhm, D. C. *et al.* Genomic and fragmentomic landscapes of cell-free DNA for early cancer detection. *Nat. Rev. Cancer* **25**, 341–358 (2025).
30. Li, S.-J. *et al.* Cell-free DNA methylation patterns in aging and their association with inflamm-aging. *Epigenomics* **16**, 715–731 (2024).
31. Sergeev, A. V., Kisil, O. V., Eremin, A. A., Petrov, A. S. & Zvereva, M. E. “Aging Clocks” Based on Cell-Free DNA. *Biochem. Mosc.* **90**, S342–S355 (2025).
32. Shtumpf, M. *et al.* Aging clock based on nucleosome reorganisation derived from cell-free DNA. *Aging Cell* **23**, e14100 (2024).
33. Hu, C., Li, Y., Li, L., Zhang, N. & Zheng, X. BS-clock, advancing epigenetic age prediction with high-resolution DNA methylation bisulfite sequencing data. *Bioinformatics* **40**, btae656 (2024).

34. Infinium MethylationEPIC v2.0 BeadChip. https://www.illumina.com/content/dam/illumina/gcs/assembled-assets/marketing-literature/infinium-methylation-epic-data-sheet-m-gl-01156/infinium-methylation-epic-data-sheet-m-gl-01156.pdf.

35. Poggiali, B. *et al*. Evaluation of the HumanMethylationEPIC v2.0 Bead Chip Using Low Quality and Quantity DNA Samples. *Biol. Proced. Online* **27**, 34 (2025).

36. Lee, S. M. *et al.* Low-input and single-cell methods for Infinium DNA methylation BeadChips. *Nucleic Acids Res.* **52**, e38 (2024).

37. Shu, C., Zhang, X., Aouizerat, B. E. & Xu, K. Comparison of methylation capture sequencing and Infinium MethylationEPIC array in peripheral blood mononuclear cells. *Epigenetics Chromatin* **13**, 51 (2020).

38. Heiss, J. A. *et al.* Battle of epigenetic proportions: comparing Illumina's EPIC methylation microarrays and TruSeq targeted bisulfite sequencing. *Epigenetics* **15**, 174–182 (2020).

39. Zhao, L.-Y., Song, J., Liu, Y., Song, C.-X. & Yi, C. Mapping the epigenetic modifications of DNA and RNA. *Protein Cell* **11**, 792–808 (2020).

40. De Ridder, K., Che, H., Leroy, K. & Thienpont, B. Benchmarking of methods for DNA methylome deconvolution. *Nat. Commun.* **15**, 4134 (2024).

41. Zhu, T. *et al.* A pan-tissue DNA methylation atlas enables in silico decomposition of human tissue methylomes at cell-type resolution. *Nat. Methods* **19**, 296–306 (2022).

42. Shokhirev, M. N. & Johnson, A. A. Analysis of variability and epigenetic age prediction across microarray and methylation sequencing technologies. *GeroScience* https://doi.org/10.1007/s11357-025-01824-1 (2025) doi:10.1007/s11357-025-01824-1.

43. De Abreu, A. R. *et al*. Comparison of current methods for genome-wide DNA methylation profiling. *Epigenetics Chromatin* **18**, 57 (2025).

44. Liu, L. *et al.* Targeted methylation sequencing of plasma cell-free DNA for cancer detection and classification. *Ann. Oncol.* **29**, 1445–1453 (2018).

45. Buckley, D. N. *et al.* OvaPrint—A Cell-free DNA Methylation Liquid Biopsy for the Risk Assessment of High-grade Serous Ovarian Cancer. *Clin. Cancer Res.* **29**, 5196–5206 (2023).

46. Komaki, S. *et al.* Epigenetic profile of Japanese supercentenarians: a cross-sectional study. *Lancet Healthy Longev.* **4**, e83–e90 (2023).

47. Bl, O. *et al*. Time is encoded by methylation changes at clustered CpG sites. *Cell Rep.* https://doi.org/10.1016/j.celrep.2025.115958 (2025) doi:10.1016/j.celrep.2025.115958.

48. Jin, Y. *et al.* Whole-genome bisulfite sequencing of cell-free DNA unveils age-dependent and ALS-associated methylation alterations. *Cell Biosci.* **15**, 26 (2025).

49. Garma, L. D. & Quintela-Fandino, M. Applicability of epigenetic age models to next-generation methylation arrays. *Genome Med.* **16**, 116 (2024).

50. Wang, J. ComBat-met: Adjusting Batch Effects in DNA Methylation Data. Preprint at https://doi.org/10.1101/2024.08.13.607838 (2024).

51. Ying, K. *et al.* A unified framework for systematic curation and evaluation of aging biomarkers. *Nat. Aging* https://doi.org/10.1038/s43587-025-00987-y (2025) doi:10.1038/s43587-025-00987-y.

52. 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* **21**, 271 (2020).

53. Handl, L., Jalali, A., Scherer, M. & Pfeifer, N. Partially blind domain adaptation for age prediction from DNA methylation data. Preprint at https://doi.org/10.48550/ARXIV.1612.06650 (2016).

54. Luo, L., Shang, L., Goodrich, J. M., Peterson, K. E. & Song, P. X.-K. Bridging the gap: Enhancing the generalizability of epigenetic clocks through transfer learning. Preprint at https://doi.org/10.1101/2025.02.14.25322080 (2025).

55. Tomusiak, A. *et al.* Development of an epigenetic clock resistant to changes in immune cell composition. *Commun. Biol.* **7**, 934 (2024).

56. Galkin, F., Mamoshina, P., Kochetov, K., Sidorenko, D. & Zhavoronkov, A. DeepMAge: A Methylation Aging Clock Developed with Deep Learning. *Aging Dis.* **12**, 1252 (2021).

57. Fong, S. *et al.* Principal component-based clinical aging clocks identify signatures of healthy aging and targets for clinical intervention. *Nat. Aging* **4**, 1137–1152 (2024).

58. Koo, T. K. & Li, M. Y. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J. Chiropr. Med.* **15**, 155–163 (2016).

59. Ziller, M. J., Hansen, K. D., Meissner, A. & Aryee, M. J. Coverage recommendations for methylation analysis by whole-genome bisulfite sequencing. *Nat. Methods* **12**, 230–232 (2015).

60. Wang, T. *et al.* A systematic study of normalization methods for Infinium 450K methylation data using whole-genome bisulfite sequencing data. *Epigenetics* **10**, 662–669 (2015).

61. Rauluseviciute, I., Drabløs, F. & Rye, M. B. DNA methylation data by sequencing: experimental approaches and recommendations for tools and pipelines for data analysis. *Clin. Epigenetics* **11**, 193 (2019).

62. Zhou, L. *et al.* Systematic evaluation of library preparation methods and sequencing platforms for high-throughput whole genome bisulfite sequencing. *Sci. Rep.* **9**, 10383 (2019).
63. Pidsley, R. *et al.* Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. *Genome Biol.* **17**, 208 (2016).
64. Sugden, K. *et al.* Patterns of Reliability: Assessing the Reproducibility and Integrity of DNA Methylation Measurement. *Patterns* **1**, 100014 (2020).
65. Tay, J. H. *et al.* DNAm age differences between infinium methylationEPICv1 vs EPICv2 in buffy coat, PBMC, and saliva samples. *Commun. Biol.* **8**, 654 (2025).
66. Apsley, A. T. *et al.* Cross-tissue comparison of epigenetic aging clocks in humans. *Aging Cell* **24**, e14451 (2025).
67. De Lima Camillo, L. P. *et al.* CpGPT: a Foundation Model for DNA Methylation. Preprint at https://doi.org/10.1101/2024.10.24.619766 (2024).
68. Evgeniou, T. & Pontil, M. Regularized multi--task learning. in *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining* 109–117 (ACM, Seattle WA USA, 2004). doi:10.1145/1014052.1014067.
69. Pan, S. J., Tsang, I. W., Kwok, J. T. & Yang, Q. Domain Adaptation via Transfer Component Analysis. *IEEE Trans. Neural Netw.* **22**, 199–210 (2011).
70. Ganin, Y. *et al.* Domain-adversarial training of neural networks. *J Mach Learn Res* **17**, 2096–2030 (2016).
71. Faul, J. D. *et al.* Epigenetic-based age acceleration in a representative sample of older Americans: Associations with aging-related morbidity and mortality. *Proc. Natl. Acad. Sci.* **120**, e2215840120 (2023).

72. De Lima Camillo, L. P. pyaging□: a Python-based compendium of GPU-optimized aging clocks. *Bioinformatics* **40**, btae200 (2024).

73. Aryee, M. J. *et al.* Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* **30**, 1363–1369 (2014).

74. Fortin, J.-P., Triche, T. J. & Hansen, K. D. Preprocessing, normalization and integration of the Illumina HumanMethylationEPIC array with minfi. *Bioinformatics* **33**, 558–560 (2017).

75. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).

76. Pedersen, B. S., Eyring, K., De, S., Yang, I. V. & Schwartz, D. A. Fast and accurate alignment of long bisulfite-seq reads. Preprint at https://doi.org/10.48550/ARXIV.1401.1129 (2014).

77. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

78. Faust, G. G. & Hall, I. M. *SAMBLASTER*□: fast duplicate marking and structural variant read extraction. *Bioinformatics* **30**, 2503–2505 (2014).

79. Lehne, B. *et al.* A coherent approach for analysis of the Illumina HumanMethylation450 BeadChip improves data quality and performance in epigenome-wide association studies. *Genome Biol.* **16**, 37 (2015).

80. Oh, G. *et al.* Circadian oscillations of cytosine modification in humans contribute to epigenetic variability, aging, and complex disease. *Genome Biol.* **20**, 2 (2019).

81. Koncevičius, K. *et al.* Epigenetic age oscillates during the day. *Aging Cell* **23**, e14170 (2024).

82. Abdelsamed, H. A. *et al.* Beta cell-specific CD8+ T cells maintain stem cell memory-associated epigenetic programs during type 1 diabetes. *Nat. Immunol.* **21**, 578–587 (2020).

83. Loyfer, N. *et al*. A DNA methylation atlas of normal human cell types. *Nature* **613**, 355–364 (2023).

Tables

Table 1| Linear regression results between the age prediction replicate difference (RD) and the number of CpGs in 20 HCRL clocks. HCRL denotes 20 linear-architecture clocks classified as Hi-Cov-All or High-Repro. Negative correlations indicate larger number of CpGs is associated with lower age prediction discrepancies between technical replicates. P-values were adjusted for multiple testing using the Benjamini-Hochberg procedure. Full regression outputs for each dataset are provided in Fig. S7.

Dataset	Slope	Intercept	Pearson's R	p
gDNA_MSA	-0.36	2.34	-0.94	**3.65E-10**
gDNA_EPICv2	-0.24	1.30	-0.87	**4.67E-07**
gDNA_Twist	-0.28	2.87	-0.90	**4.91E-08**
gDNA_Galaxy	-0.31	3.01	-0.93	**3.49E-09**
cfDNA_Twist	-0.29	3.04	-0.91	**2.69E-08**
cfDNA_Galaxy	-0.28	3.01	-0.90	**5.86E-08**
GSE83944	-0.18	0.75	-0.83	**7.78E-05**
GSE247193	-0.19	1.60	-0.84	**5.65E-05**
GSE247195	-0.22	1.14	-0.85	**2.99E-05**
GSE247197	-0.14	0.55	-0.67	**4.62E-03**
GSE55763	-0.18	0.93	-0.81	**1.65E-04**

Figure Legend

Fig. 1| Study design and evaluation of methylation data reproducibility and consistency across six SRRSH-24 datasets. a, The overall design of the study. **b-c**, Principal component analysis (PCA) of all samples colored by datasets (**b**) or replicates (**c**). Beta-values of Method-shared CpGs (N = 136,437) were used for PCA calculation. **d**, Intraclass Correlation Coefficients (ICCs) of Method-shared CpGs across six datasets. **e**, ICC values of CpGs filtered by different depth across four HTS-based datasets. **f**, Pearson correlations among datasets computed using the mean beta-values of Method-shared CpGs. **g**, ICCs of 53 epigenetic clocks benchmarked in the study. Data are presented as ICC estimates with 95 % confidence intervals and colored by grouped ICC values as follows: Excellent: $0.9 \leq$ ICC, Good: $0.75 \leq$ ICC < 0.9, Moderate: $0.5 \leq$ ICC < 0.75, Poor: ICC < 0.5. Clocks were classified in two different rationales: (1) By generations: First, Chronological age prediction; Second, Phenotypic age or mortality prediction; Third, Pace of aging prediction; Fourth, Causal clocks; Gestational, Gestational age prediction; Other: Telomere length or mitotic age prediction. (2) By CpG coverage across the four profiling methods (MSA, EPICv2, Galaxy, Twist): Hi-Cov-All, first- or second-generation clocks, applicable to adults and have > 80 % CpGs covered by all four methods; Hi-Repro, first- or second-generation clocks with excellent ICC across datasets; Hi-Cov-nonMSA, first- or second-generation clocks, applicable to adults, have > 80 % CpGs covered by three methods except MSA, and not classified as Hi-Repro; Other, all remaining clocks. **h**, Comparisons of per clock ICCs of all tested epigenetic clocks (N = 53). **i**, Pearson's correlations across six datasets calculated using mean of clock predictions for three clock categories: Hi-Repro, Hi-Cov-All, and Hi-Cov-nonMSA. P-values in (**d**) and (**h**) were calculated by two-tailed block permutation test with 10,000 permutations, then adjusted for multiple comparisons using the Benjamini-Hochberg procedure. ***, $p < 0.001$; **, $p < 0.01$; *, $p < 0.05$; N.S., not significant. The center line of each boxplot marks the median, the box body marks the 1st and 3rd quartile, the caps and/or the span of whisker mark the min and max of RD values respectively.

Fig. 2| Decline of epigenetic clock performance in 6 SRRSH-24 HTS-based datasets. a, Lollipop plot showing Pearson's R and MAE (between predicted age and chronological age) for 6 Hi-Repro clocks. Each dot represents a replicate, and color represents R. **b**, Lollipop plot showing Pearson's R and MAE (between predicted age and chronological age) for 20 Hi-Cov-All clocks. **c-d**, Boxplots showing comparisons and statistical significance of R (**c**) and MAE (**d**) for Hi-Cov-All clocks across datasets. The center line of each boxplot marks the median, the box body marks the 1st and 3rd quartile, and the caps mark the min and max of RD values respectively. **e-f**, Performance evaluation (R and MAE) of 4 array-trained clocks (hannum, horvath2013, skinandblood and dnamphenoage) and a HTS data-trained clock (timeseq) cross-applied to array-based dataset (GSE245628, labelled as EPICv1) and HTS-based dataset (GSE232346, labelled as TimeSeq). Only the 24 shared samples in both GSE245628 and GSE232346 were incorporated. **g**, Scatter plot showing the correlation between clock ICC and R. **h**, Scatter plot showing the correlation between clock ICC and MAE. Each dot represents one clock from Hi-Cov-All category, color represents replicate group. Lines represent linear regression fits for each replicate group respectively. The p-values in (**c**) and (**d**) were calculated by two-tailed block permutation test with 10,000 permutations, then adjusted for multiple comparisons using the Benjamini-Hochberg procedure. ***, $p < 0.001$; **, $p < 0.01$; *, $p < 0.05$; N.S., not significant; The R and p-values in (**g**) and (**h**) were calculated by Pearson correlation at replicate level.

Fig. 3| Investigation of the negative correlation between number of CpGs (log-transformed) and the age prediction replicate difference (RD) (log-transformed) shows that the association is largely driven by the λ hyperparameter (L1-regularisation mixing ratio). a, A series of elastic nets were trained on dataset GSE55763 (N = 2711) and tested with the 6 datasets from this study. The models were trained under different hyperparameter (α and λ) values to obtain models with different number of CpGs. Each point represents a model trained with a distinct α and λ pair. **b**, The negative correlation between number of CpGs and age prediction RD was validated in all testing datasets with high significance ($p \leq 4.7\times10^{-10}$). **c-d**, Compared with hyperparameter α, λ showed a stronger and more significant correlation with RD (p: 9.7×10^{-9} - 8.2×10^{-3} vs. 1.6×10^{-47} - 7.3×10^{-33}; Pearson's R^2: 0.05 - 0.21 vs. 0.67 - 0.78). R^2 and p-values were calculated by Pearson correlation.

Fig. 4| Investigating the impact of unreliable beta-values on 20 HCRL clocks and imputation-based correction techniques. HCRL denotes 20 linear-architecture clocks classified as Hi-Cov-All or High-Repro. **a**, HCRL clocks are prone to unreliable beta-values originating from high-throughput sequencing (HTS)-based methylation profiling technologies, with CpG coverage varying from 2.0% to 34.6% across the clocks examined (Details are shown in Table S3). The center line of each boxplot marks the median, the box body marks the 1st and 3rd quartile, and the caps mark the min and max values respectively. **b-c**, The K-nearest-neighbor (KNN) imputation technique was tested at varying depth cut-offs; only CpG sites below the cut-off were imputed. **d-e**, Similarly, Mean and median imputation significantly improved RD but simultaneously increased MAE in cfDNA datasets. The horizontal strokes mark the medians of each respective distribution. No significant increase of MAE was observed in gDNA datasets. 0-filling yielded neither significant improvement nor deterioration in MAE or RD. P-values were obtained by two-sided permutation tests (10,000 permutations, custom script) comparing pre- and post-imputation predictions. Multiple testing results were adjusted using the Benjamini-Hochberg procedure. ***, $p < 0.001$; **, $p < 0.01$; *, $p < 0.05$; NS, not significant.

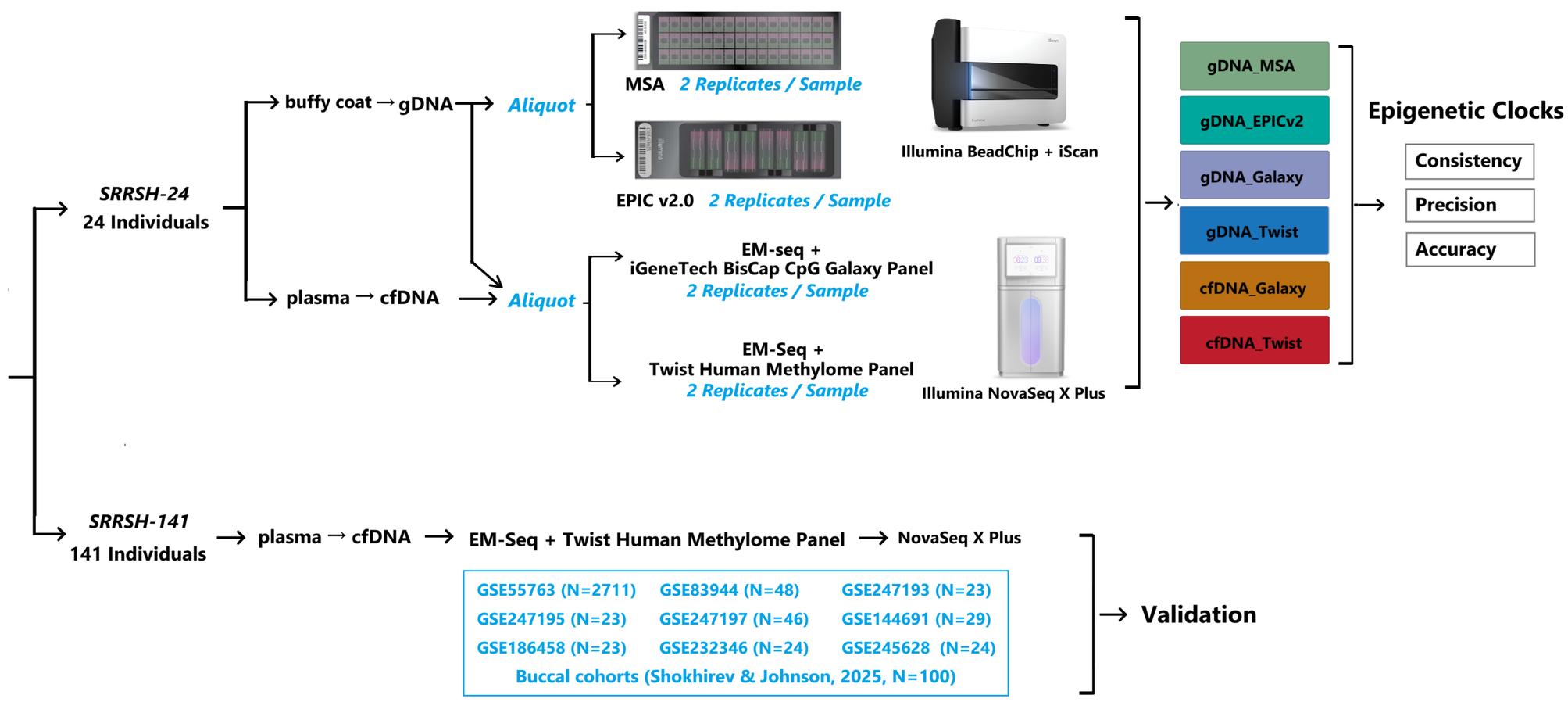
Fig. 5| Classic epigenetic clock prediction accuracy can be improved using transfer learning techniques. Distillation, selected for its minimal requirement for original parameters and training data, incorporates the original model (teacher model) output as an additional input feature to train an elastic net student model for prediction and validation. All models were trained and tested with 5-fold cross-validation. Prior to transfer learning, datasets were transformed by PCA (loadings derived from GSE55763; N = 2711) to reduce dimensionality and accelerate training. **a**, Distillation improved age prediction mean absolute error (MAE) with 6 selected epigenetic clocks as teacher models. MAE values for the teacher model and for the student model without transfer learning are shown for reference. MAE was assessed after summarizing all predicted splits after a full pass of 5-cold cross validation. **b**, Distillation reduced age prediction replicate difference (RD) between teacher clocks, however, the direction and magnitude of improvement varied across datasets and teacher model selections. RD values for the teacher model and the student model without transfer learning are shown for reference. RD was assessed after summarizing all predicted splits after a full pass of 5-cold cross validation. The center line of each boxplot marks the median, the box body marks the 1st and 3rd quartile, and the caps mark the min and max of RD values respectively.

Fig. 6| Investigating optimal sequencing depths for epigenetic clocks and strategies to improve accuracy and reproducibility. a-b, Investigation of the sequencing depth requirements with downsampling simulations with the SRRSH cohort (N = 165). a, Although only 20 - 40 % of CpGs in the 6 tested epigenetic clocks (the same as aforementioned teacher models in Fig. 5) achieved ≥ 20× depths, the age prediction MAE converged to the dataset baseline at 10-20× downsampling depths. b, The increasing of the fraction of CpGs with ≥ 20× slowed down above ≥ 20× downsampling depth, although it could continue climbing up to 50 - 80× downsampling depth. The markers and error bars stand for the means and standard deviations calculated from 5 qualified samples for downsampling experiments. Additional model simulations supporting similar range of minimum depth of 10 - 20× with different rationales were shown in Fig. S13. **c**, TL and IM+TL showed the greatest overall improvements in MAE, tested with the SRRSH-141 cohort and HCRL clocks. HCRL denotes 20 linear-architecture clocks classified as Hi-Cov-All or High-Repro. Method abbreviations: Direct, no improvement techniques used; DF, depth filtering with a minimal sequencing depth of 20×; IM, imputing unreliable (0 and 1) beta-values; TL, incorporated distillation as transfer learning technique. These techniques were sequentially applied in DF, IM, TL order. The center line of each boxplot marks the median, the box body marks the 1st and 3rd quartile, and the caps mark the min and max values respectively. **d-e**, The techniques combining unreliable beta imputation and distillation (IMputation+Transfer Learning, IM+TL) produced the best MAE improvements in the two validation datasets: SRRSH-141 cohort and the Buccal cohorts (N = 100). Comprehensive comparisons of all technique combination strategies are shown in Fig. S16-S17.

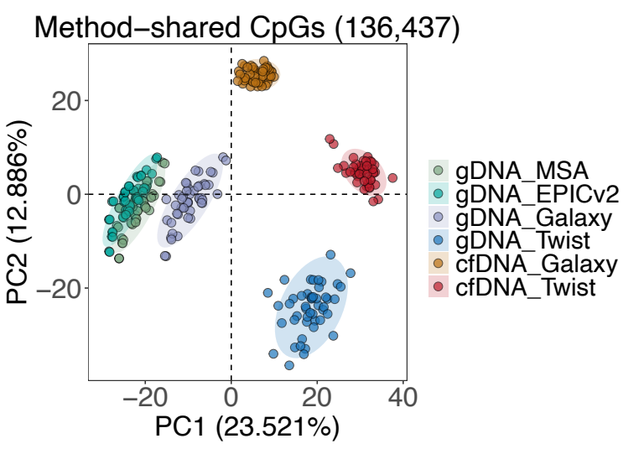
a

Age	Male	Female	Total
20~29	12	18	30 (18.2%)
30~39	21	21	42 (25.5%)
40-49	24	10	34 (20.6%)
50-59	26	12	38 (23%)
60-69	12	7	19 (11.5%)
>70	1	1	2 (1.2%)
Total	96 (58.2%)	69 (41.8%)	165

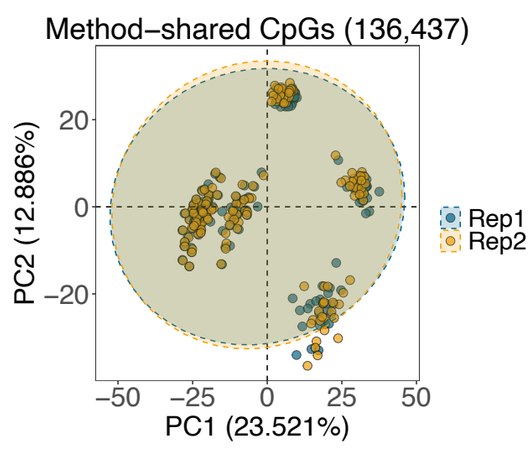
SRRSH Cohort (N=165)



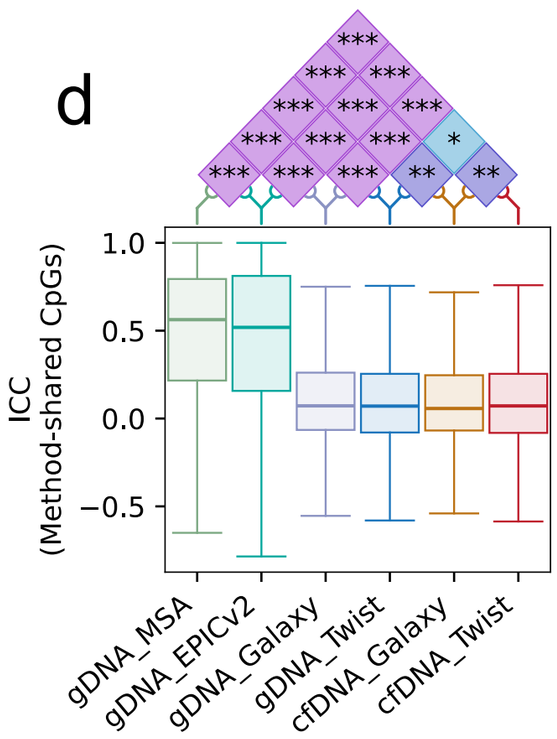
b



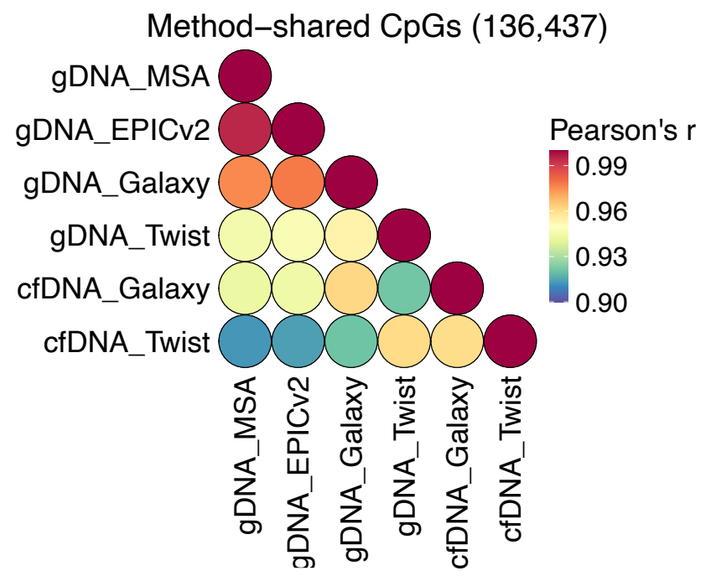
c



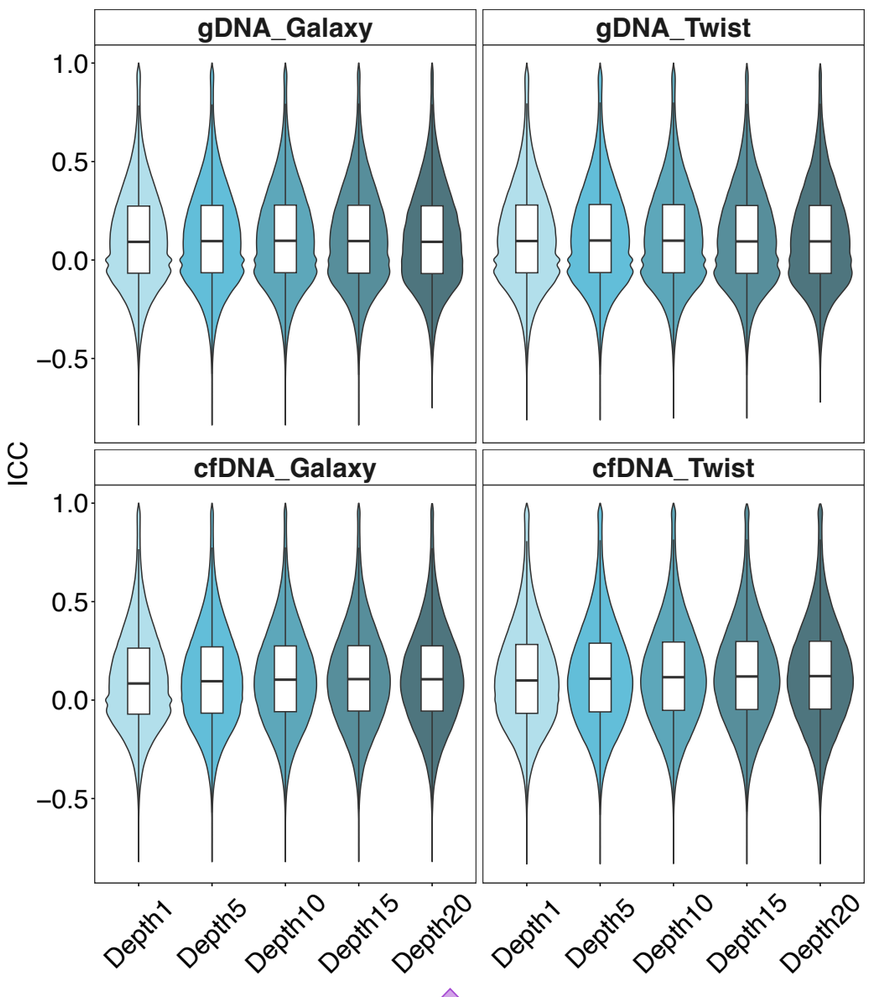
d



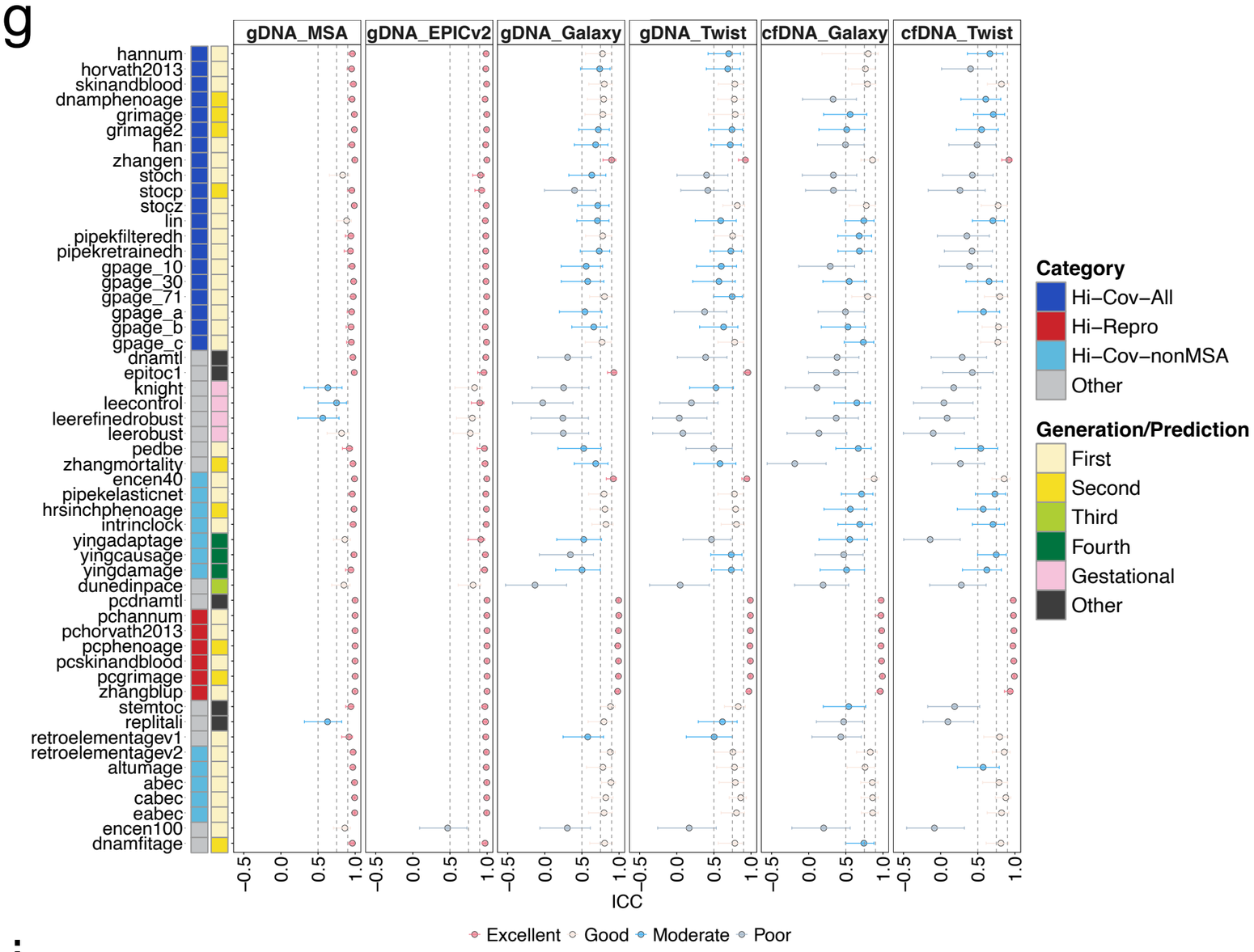
f



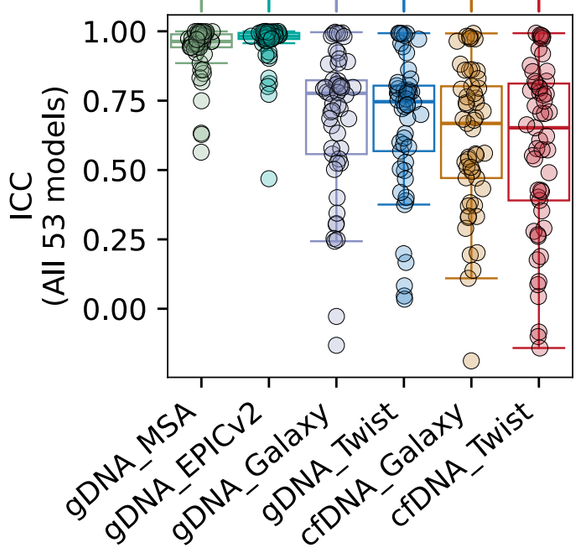
e



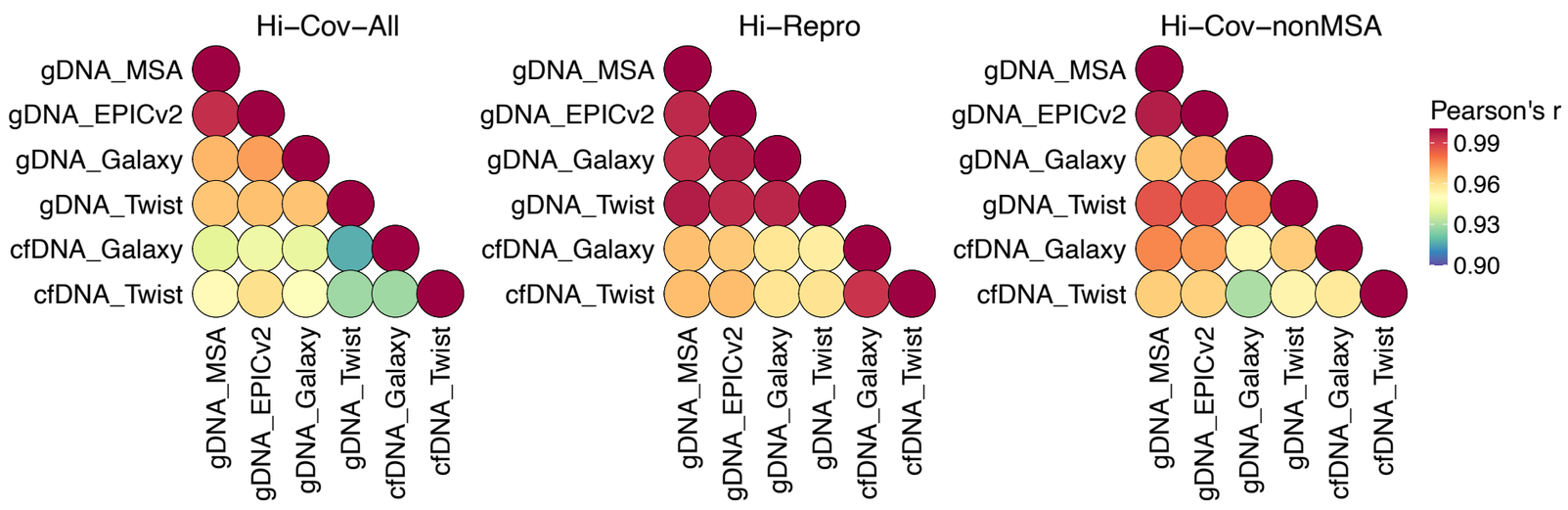
g

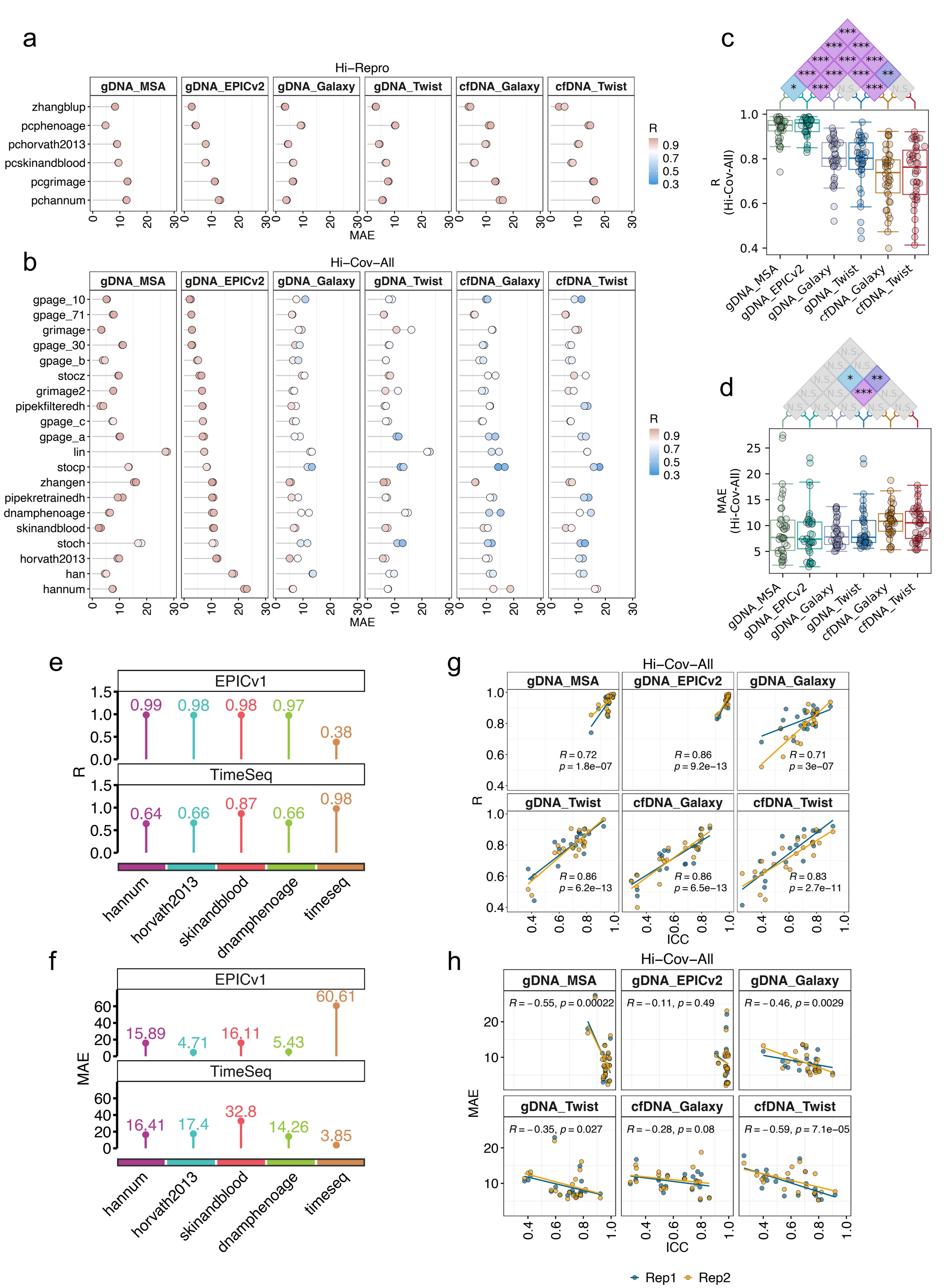


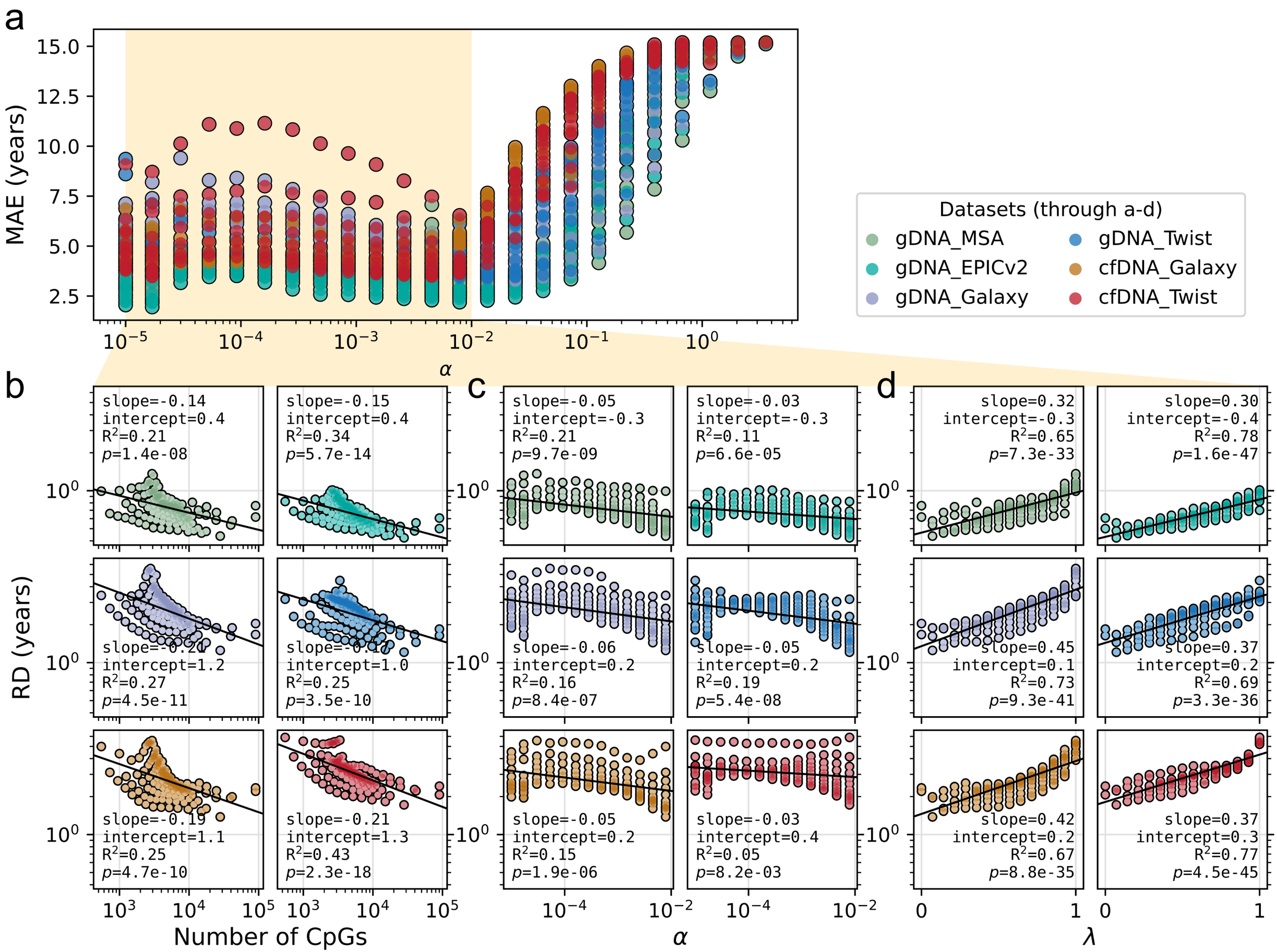
h

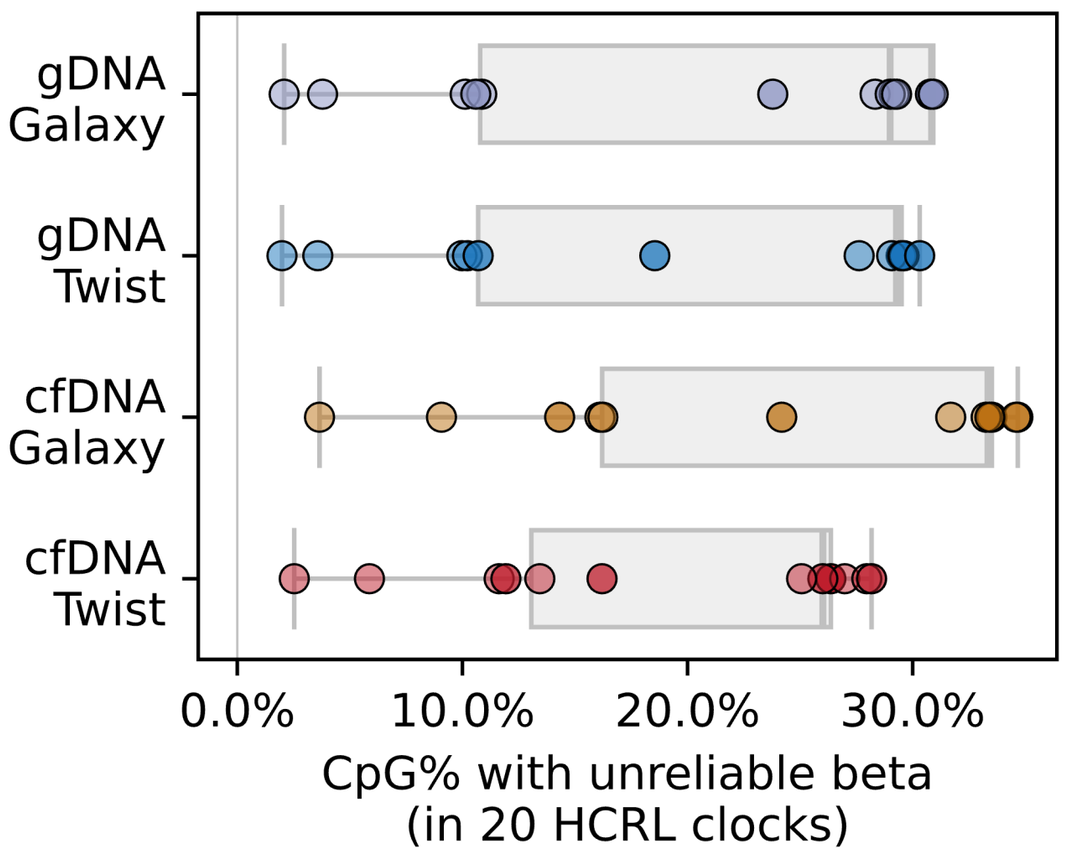


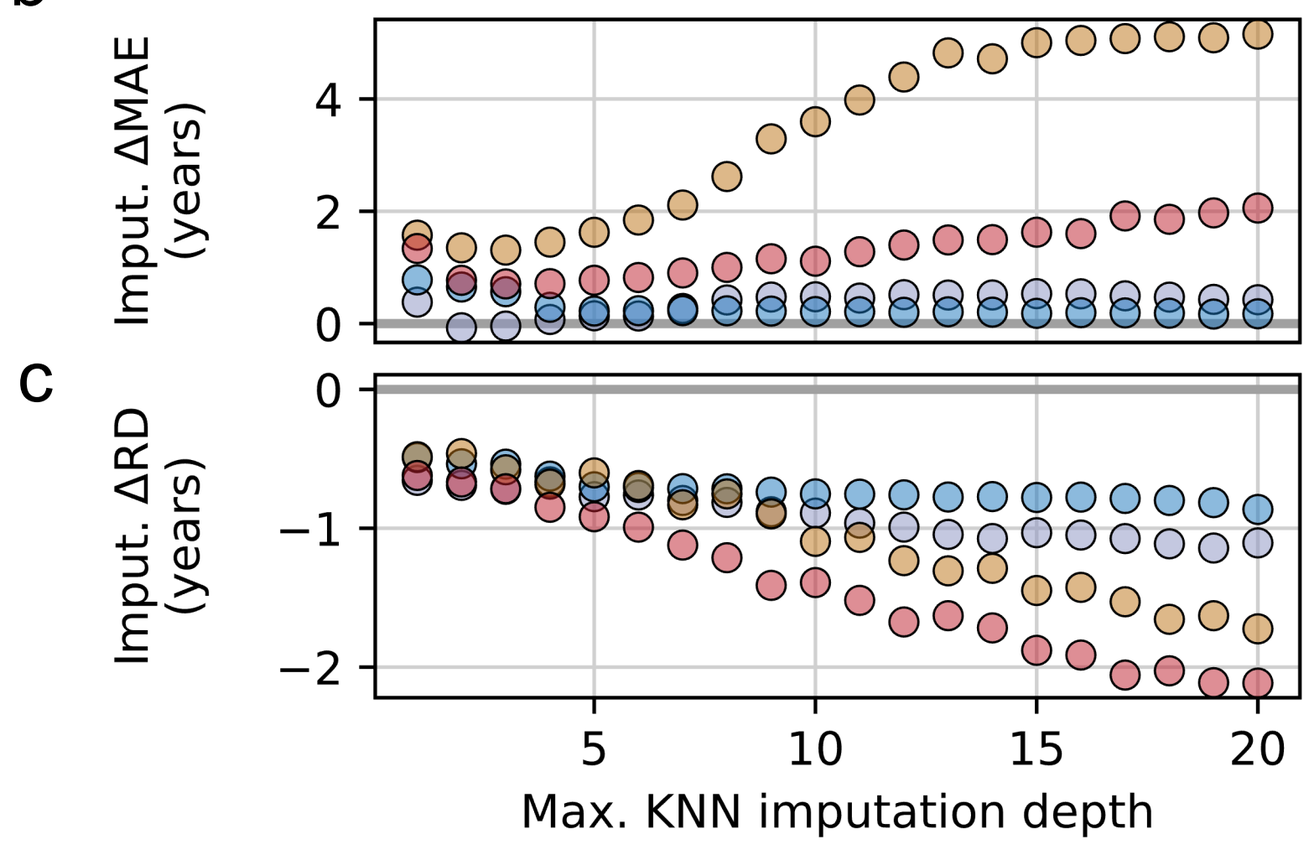
i

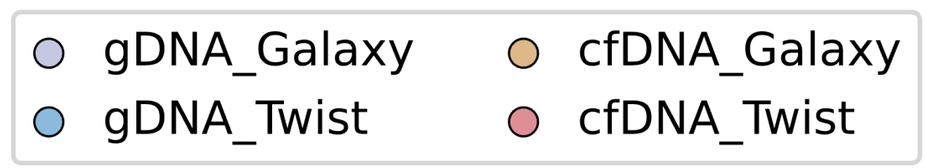


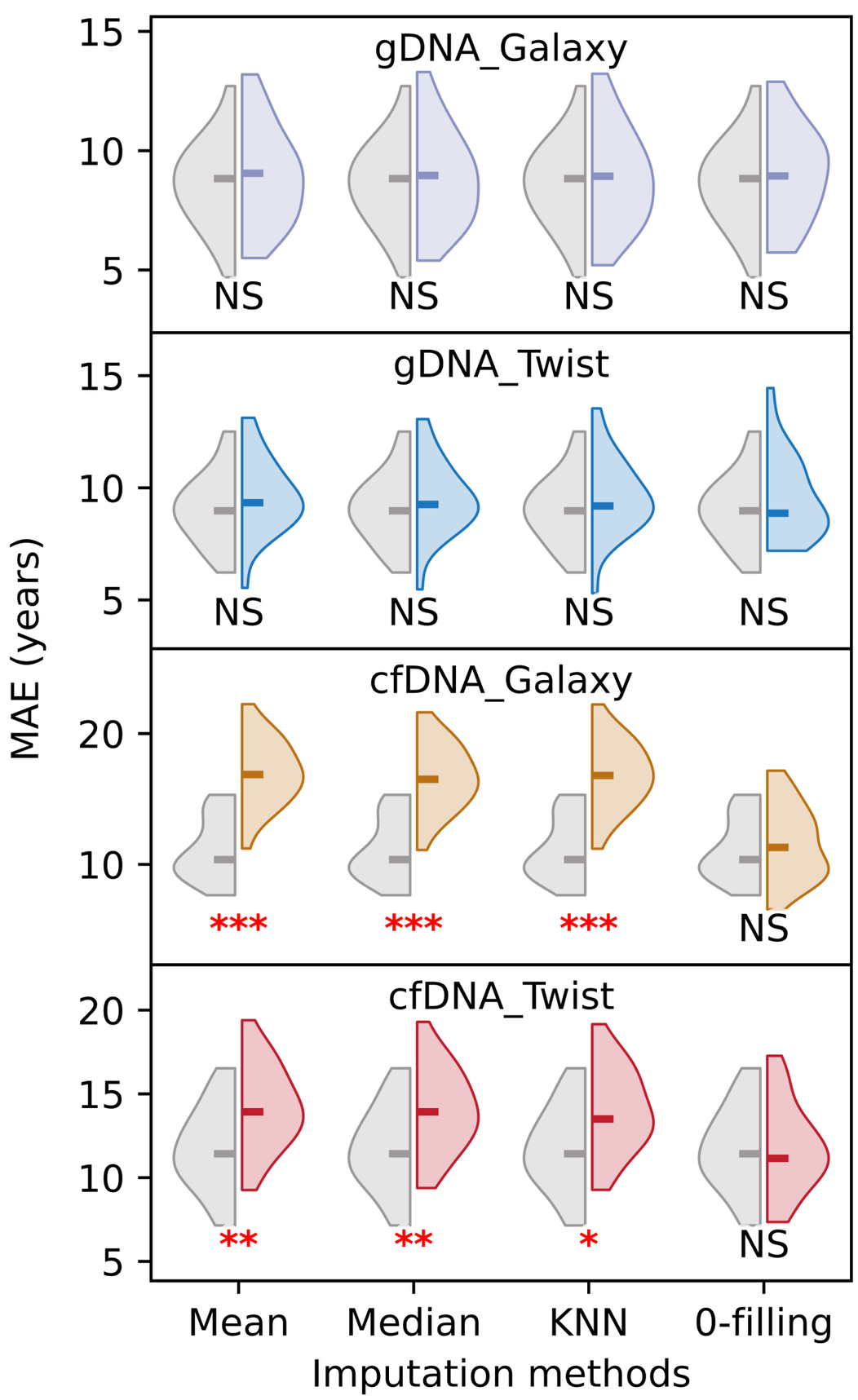

a
Hi–Repro
gDNA_MSA
gDNA_EPICv2
gDNA_Galaxy
gDNA_Twist
cfDNA_Galaxy
cfDNA_Twist
zhangblup
pcphenoage
pchorvath2013
pcskinandblood
pcgrimage
pchannum
MAE
R
0.9
0.7
0.5
0.3
b
Hi–Cov–All
gpage_10
gpage_71
grimage
gpage_30
gpage_b
stocz
grimage2
pipekfilteredh
gpage_c
gpage_a
lin
stocp
zhangen
pipekretrainedh
dnamphenoage
skinandblood
stoch
horvath2013
han
hannum
c
(Hi-Cov-All)
d
e
EPICv1
TimeSeq
0.99
0.98
0.98
0.97
0.38
0.64
0.66
0.87
0.66
0.98
hannum
horvath2013
skinandblood
dnamphenoage
timeseq
f
15.89
4.71
16.11
5.43
60.61
16.41
17.4
32.8
14.26
3.85
g
h
ICC
R = 0.72
p = 1.8e–07
R = 0.86
p = 9.2e–13
R = 0.71
p = 3e–07
R = 0.86
p = 6.2e–13
R = 0.86
p = 6.5e–13
R = 0.83
p = 2.7e–11
R = –0.55, p = 0.00022
R = –0.11, p = 0.49
R = –0.46, p = 0.0029
R = –0.35, p = 0.027
R = –0.28, p = 0.08
R = –0.59, p = 7.1e–05
Rep1
Rep2

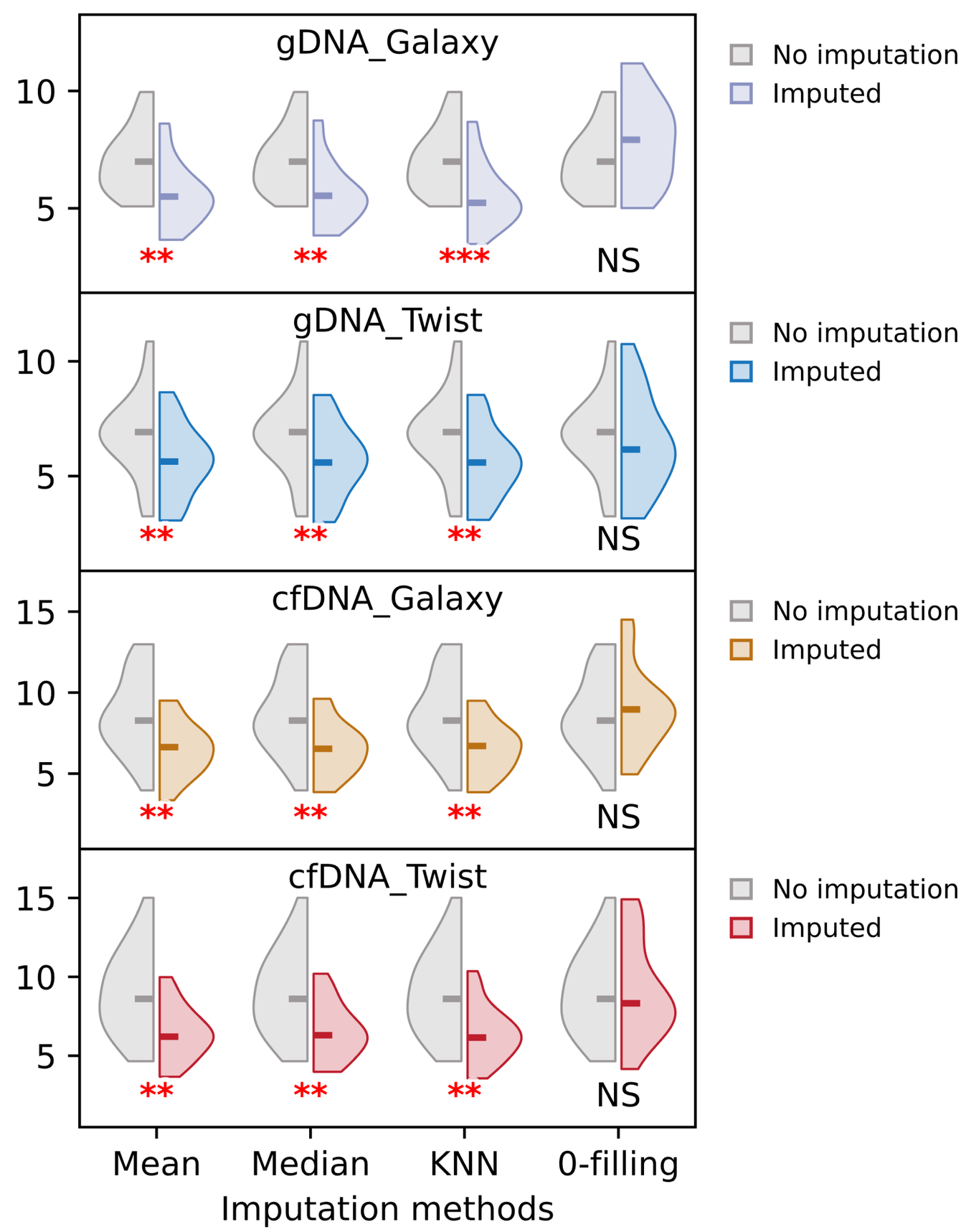

a
MAE (years)
15.0
12.5
10.0
7.5
5.0
2.5
10^{-5}
10^{-4}
10^{-3}
10^{-2}
10^{-1}
10^{0}
α
Datasets (through a-d)
gDNA_MSA
gDNA_EPICv2
gDNA_Galaxy
gDNA_Twist
cfDNA_Galaxy
cfDNA_Twist
b
c
d
RD (years)
slope=-0.14
intercept=0.4
R^2=0.21
p=1.4e-08
slope=-0.15
intercept=0.4
R^2=0.34
p=5.7e-14
slope=-0.05
intercept=-0.3
R^2=0.21
p=9.7e-09
slope=-0.03
intercept=-0.3
R^2=0.11
p=6.6e-05
slope=0.32
intercept=-0.3
R^2=0.65
p=7.3e-33
slope=0.30
intercept=-0.4
R^2=0.78
p=1.6e-47
slope=-0.20
intercept=1.2
R^2=0.27
p=4.5e-11
slope=-0.17
intercept=1.0
R^2=0.25
p=3.5e-10
slope=-0.06
intercept=0.2
R^2=0.16
p=8.4e-07
slope=-0.05
intercept=0.2
R^2=0.19
p=5.4e-08
slope=0.45
intercept=0.1
R^2=0.73
p=9.3e-41
slope=0.37
intercept=0.2
R^2=0.69
p=3.3e-36
slope=-0.19
intercept=1.1
R^2=0.25
p=4.7e-10
slope=-0.21
intercept=1.3
R^2=0.43
p=2.3e-18
slope=-0.05
intercept=0.2
R^2=0.15
p=1.9e-06
slope=-0.03
intercept=0.4
R^2=0.05
p=8.2e-03
slope=0.42
intercept=0.2
R^2=0.67
p=8.8e-35
slope=0.37
intercept=0.3
R^2=0.77
p=4.5e-45
10^{0}
10^{3}
10^{4}
10^{5}
Number of CpGs
10^{-4}
10^{-2}
α
0
1
λ

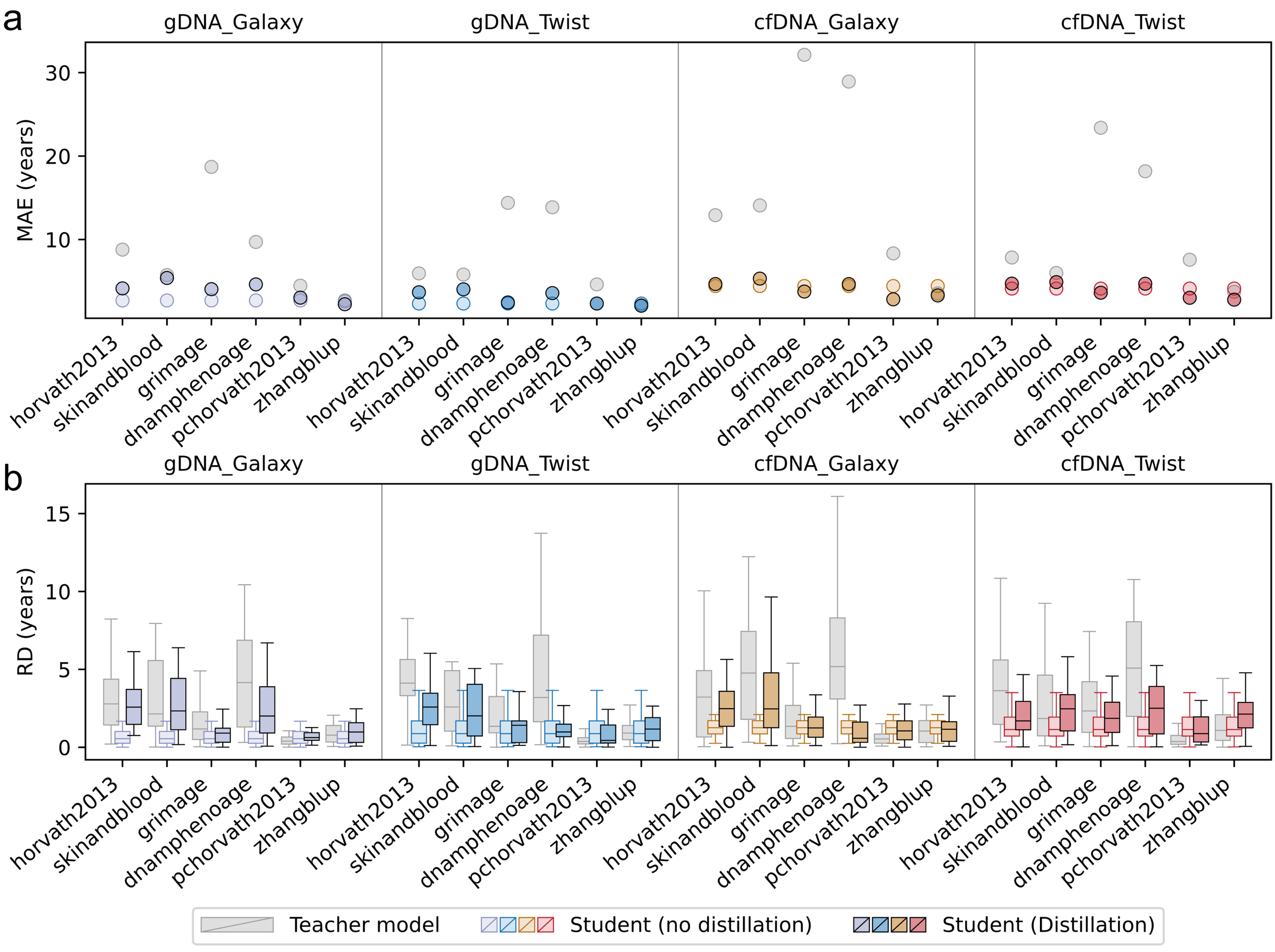

a
gDNA Galaxy
gDNA Twist
cfDNA Galaxy
cfDNA Twist
0.0%
10.0%
20.0%
30.0%
CpG% with unreliable beta
(in 20 HCRL clocks)

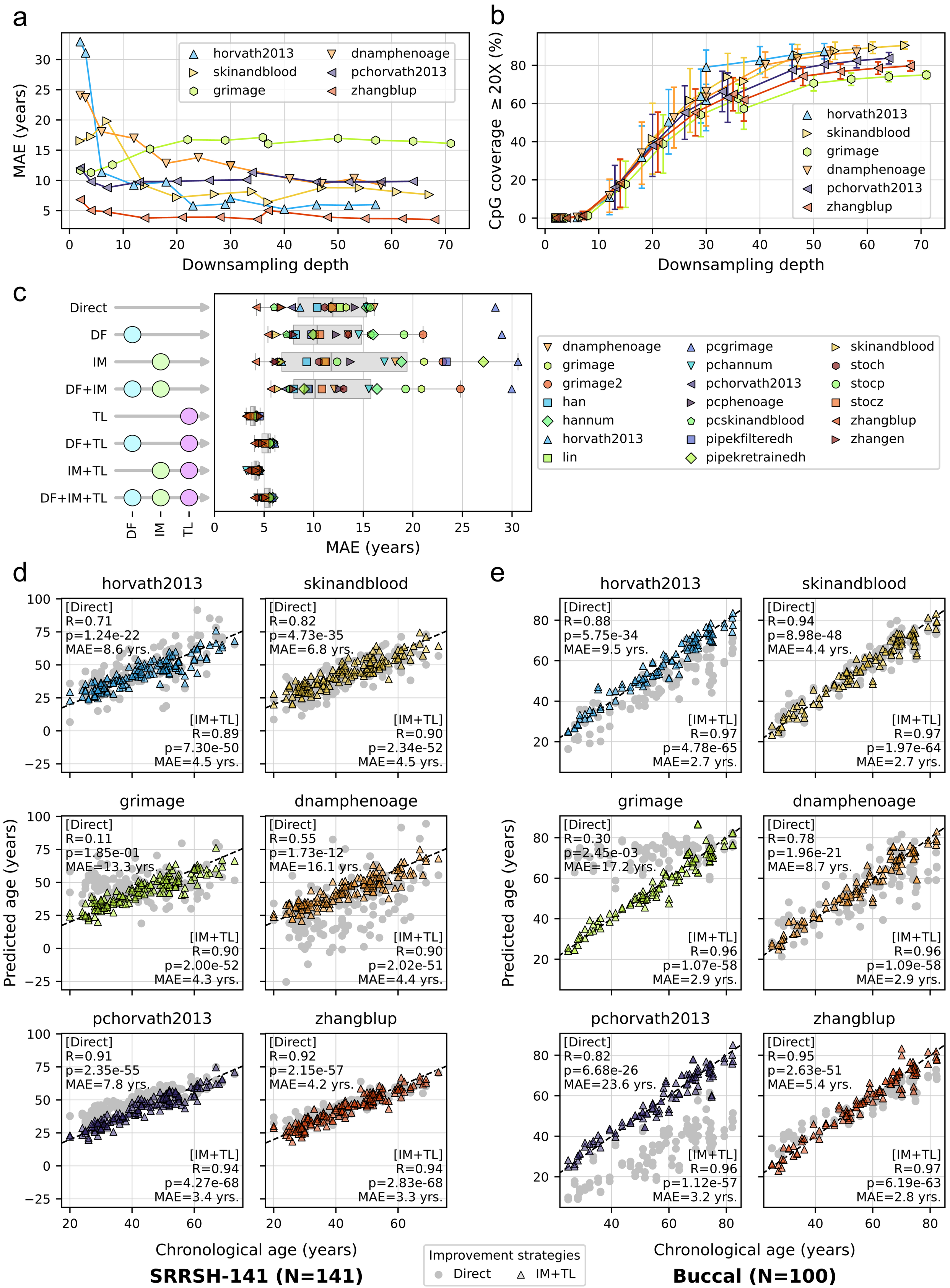

b
Imput. ΔMAE (years)
c
Imput. ΔRD (years)
Max. KNN imputation depth


gDNA_Galaxy
gDNA_Twist
cfDNA_Galaxy
cfDNA_Twist


d
gDNA_Galaxy
gDNA_Twist
cfDNA_Galaxy
cfDNA_Twist
MAE (years)
Mean
Median
KNN
0-filling
Imputation methods


e
gDNA_Galaxy
gDNA_Twist
cfDNA_Galaxy
cfDNA_Twist
RD (years)
No imputation
Imputed
Mean
Median
KNN
0-filling
Imputation methods


a
gDNA_Galaxy
gDNA_Twist
cfDNA_Galaxy
cfDNA_Twist
MAE (years)
30
20
10
horvath2013
skinandblood
grimage
dnamphenoage
pchorvath2013
zhangblup
b
RD (years)
15
10
5
0
Teacher model
Student (no distillation)
Student (Distillation)


a
MAE (years)
Downsampling depth
horvath2013
skinandblood
grimage
dnamphenoage
pchorvath2013
zhangblup
b
CpG coverage ≥ 20X (%)
Downsampling depth
horvath2013
skinandblood
grimage
dnamphenoage
pchorvath2013
zhangblup
c
Direct
DF
IM
DF+IM
TL
DF+TL
IM+TL
DF+IM+TL
MAE (years)
dnamphenoage
grimage
grimage2
han
hannum
horvath2013
lin
pcgrimage
pchannum
pchorvath2013
pcphenoage
pcskinandblood
pipekfilteredh
pipekretrainedh
skinandblood
stoch
stocp
stocz
zhangblup
zhangen
d
horvath2013
[Direct] R=0.71 p=1.24e-22 MAE=8.6 yrs.
[IM+TL] R=0.89 p=7.30e-50 MAE=4.5 yrs.
skinandblood
[Direct] R=0.82 p=4.73e-35 MAE=6.8 yrs.
[IM+TL] R=0.90 p=2.34e-52 MAE=4.5 yrs.
grimage
[Direct] R=0.11 p=1.85e-01 MAE=13.3 yrs.
[IM+TL] R=0.90 p=2.00e-52 MAE=4.3 yrs.
dnamphenoage
[Direct] R=0.55 p=1.73e-12 MAE=16.1 yrs.
[IM+TL] R=0.90 p=2.02e-51 MAE=4.4 yrs.
pchorvath2013
[Direct] R=0.91 p=2.35e-55 MAE=7.8 yrs.
[IM+TL] R=0.94 p=4.27e-68 MAE=3.4 yrs.
zhangblup
[Direct] R=0.92 p=2.15e-57 MAE=4.2 yrs.
[IM+TL] R=0.94 p=2.83e-68 MAE=3.3 yrs.
Predicted age (years)
Chronological age (years)
SRRSH-141 (N=141)
e
horvath2013
[Direct] R=0.88 p=5.75e-34 MAE=9.5 yrs.
[IM+TL] R=0.97 p=4.78e-65 MAE=2.7 yrs.
skinandblood
[Direct] R=0.94 p=8.98e-48 MAE=4.4 yrs.
[IM+TL] R=0.97 p=1.97e-64 MAE=2.7 yrs.
grimage
[Direct] R=0.30 p=2.45e-03 MAE=17.2 yrs.
[IM+TL] R=0.96 p=1.07e-58 MAE=2.9 yrs.
dnamphenoage
[Direct] R=0.78 p=1.96e-21 MAE=8.7 yrs.
[IM+TL] R=0.96 p=1.09e-58 MAE=2.9 yrs.
pchorvath2013
[Direct] R=0.82 p=6.68e-26 MAE=23.6 yrs.
[IM+TL] R=0.96 p=1.12e-57 MAE=3.2 yrs.
zhangblup
[Direct] R=0.95 p=2.63e-51 MAE=5.4 yrs.
[IM+TL] R=0.97 p=6.19e-63 MAE=2.8 yrs.
Predicted age (years)
Chronological age (years)
Buccal (N=100)
Improvement strategies
Direct
IM+TL