A robust and integrated framework for cross-platform adaptation of epigenetic clocks in cell-free DNA (cfDNA) sequencing

Guangyu Li [1,7], Weilai Huang [2,7], Xuechen Zhao [2,3,7], Jun Wu [2], Yuanfang Guo [2], Liying Chen [4], Xingqi Cao [4], Zipeng Yang [2], Siqi Jiang [2], Bin Hu [2], Yijun Wang [5], Dinson Tan [2,3], Vicky Tong [2,3], Changdong Tang [2], Xiaowen Feng [2,3], Xiaomin Hu [6], Chuan Ouyang [2,*], Guangyu Zhou [1,2,3,*]

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

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

[3] Regenerative Bio Limited, Hong Kong, Hong Kong SAR, China

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

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

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

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

[*] Corresponding authors

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

Abstract

Epigenetic clocks trained on methylation arrays generalize poorly to high-throughput sequencing (HTS) of cell-free DNA (cfDNA). Using paired array and HTS replicates, we systematically identified requirements to bridge this platform gap and developed a standardized, model-agnostic adaptation framework. Optimal performance requires maintaining at least 10× mean target depth, utilizing L2-weighted regularization, implementing targeted beta-value imputation and transfer learning. A combined framework using these strategies significantly enhanced legacy clock

performance across independent aging and disease cohorts, enabling robust, minimally invasive biological age assessment without compromising biological interpretability.

Keywords: Epigenetic Aging Clocks; cfDNA Methylation; High-throughput Sequencing; Machine Learning; Model Adaptation Framework.

Background

Recent studies increasingly advocate using biological age (BA), rather than chronological age (CA), as a more informative indicator of an individual's physiological aging state. BA is estimated from biomarkers of aging (BoAs) through mathematical models known as aging clocks [1–3]. Among diverse BoAs, epigenetic clocks—based on the methylome (DNA methylation profiles)—have emerged as leading tools due to their outstanding predictive accuracy [4–8], achievable even with simple models [7–12]. Consequently, epigenetic clocks are now widely used in health monitoring, age-related disease assessment, and longevity intervention studies [2,6,13–15].

Circulating cell-free DNA (cfDNA or ccfDNA), primarily found in body fluids, enables minimally invasive testing for early cancer detection and disease monitoring [16–19]. Its utility in epigenetic aging clocks is being actively explored [20–22]. Unlike array-based genomic DNA (gDNA) profiling (e.g., Illumina HumanMethylation450 BeadChip and Illumina MethylationEPIC BeadChip), cfDNA methylation analysis typically relies on high-throughput sequencing (HTS) technologies because cfDNA is characterized by high tissue-of-origin heterogeneity [19], short fragment sizes [23] and low input (0.1–10 ng, compared with 200 ng recommended for arrays) [24] —features that array platforms struggle to accommodate. Arrays further suffer from limited sensitivity to low-abundance methylation changes, narrow dynamic range, inability to resolve intra-sample methylation heterogeneity [25]. Input reduction methods such as preamplification may also distort beta-values [26]. HTS, particularly targeted sequencing based on hybridization probe capture, overcomes these constraints [27–30] and is therefore tightly coupled with cfDNA methylation profiling [31,32].

Despite their predictive utility, most extant epigenetic clocks are optimized for array-based gDNA profiling, resulting in limited cross-platform transferability when array-trained clocks are applied to sequencing data [25,28]. This incompatibility is primarily rooted in divergent data architectures: microarrays generate continuous beta-values based on fluorescent signal intensities, whereas HTS yields discrete, count-based methylation ratios. Consequently, while array-derived variances typically follow predictable distributions, HTS data are inherently heteroscedastic, with variance structures stochastically driven by local sequencing depth. Additionally, array-trained models are restricted to fixed, curated CpG subsets and cannot leverage the expansive genomic landscapes accessible via HTS platforms [25]. These limitations create a significant technological gap between HTS-based cfDNA applications and legacy epigenetic clocks, impairing performance when array-trained gDNA clocks are applied to cfDNA data [4,25,28].

To bridge this gap, two methodological paradigms have emerged. First, *de novo* HTS models, such as BS-clock and timeseqage [25,33], are trained exclusively on sequencing datasets to bypass platform-specific biases. However, they typically identify novel HTS-specific feature sets rather than retaining the established biological relevance of legacy array-trained clocks. Second, adaptation strategies aim to harmonize array-trained models with HTS data. Data-centric approaches — including batch correction [34–36] and domain adaptation [37,38] — can introduce technical artifacts [39] or rely on distributional assumptions that fail to capture the high stochasticity and depth-dependent variance inherent in cfDNA sequencing. Alternatively, model-transfer frameworks like Transfer Elastic Net [40] are fundamentally model-specific, requiring architectural transparency, explicit coefficient access, or even original training data, all of which are often unavailable in practice. Although recent robust feature-selection methods attempt to identify consensus features across platforms [41], their reliance on cell-type-specific tuning is largely incompatible with the intrinsic cell type heterogeneity of cfDNA. Furthermore, the scarcity of systematic benchmarks featuring paired array/HTS technical replicates has hindered the establishment of standardized experimental and bioinformatic protocols — a gap particularly acute for cfDNA, where platform-specific biases readily confound biological aging signals. Therefore, an urgent need remains for a systematic, model-agnostic framework capable of benchmarking and adapting legacy array-trained clocks for high-precision, HTS-based cfDNA analysis.

To systematically address the platform gap, we established a controlled benchmarking framework using a matched dataset of cfDNA and gDNA profiled by both HTS and array-based methylation technologies. This design enabled quantitative assessment of clock accuracy, precision, and reproducibility across DNA types and platforms, and allowed us to decompose and correct sources of variance, including coverage depth, CpG missingness, and platform biases. Through structured evaluation of these experimental and computational factors, we identified recurrent bias patterns and failure modes arising from cross-platform transfer of epigenetic clocks. Based on these insights, we developed and validated an end-to-end adaptation pipeline integrating experimental design guidance, data harmonization, feature recalibration, and machine-learning-based correction. Collectively, this work provides a systematic benchmarking resource and a validated methodological framework for robust application of epigenetic clocks to HTS-derived cfDNA data.

Results

Systematic benchmarking of epigenetic profiling across array and HTS platforms

We benchmarked four profiling technologies, Illumina MSA and EPICv2 (arrays) alongside iGeneTech Galaxy and Twist (HTS), using the SRRSH-24 cohort (24 participants; 6 datasets across gDNA and cfDNA with technical replicates; Fig. 1a). The remaining 141 participants (SRRSH-141) were reserved for independent validation. The above methods were selected to represent both established and emerging methylation profiling technologies (Methods).

Platform-induced feature coverage variance and coverage drift

While HTS capture panels provide significantly larger total marker sets (Galaxy: ~2.0M; Twist: ~3.98M) than arrays (MSA: ~270K; EPICv2: ~935K) (Fig. S1a), this genomic breadth did not translate to higher clock-specific coverage (Fig. S1b). Additionally, empirical assessment revealed a “coverage drift” between observed and expected CpG coverage of clocks (calculated from manufacturer’s specifications): HTS platforms exceeded design specifications potentially due to off-target reads (Galaxy: +4.8%; Twist: +1.06%), whereas arrays underperformed potentially due to stringent QC filtering (MSA: -2.5%; EPICv2: -3.0%; Fig. S1b). Despite these discrepancies, HTS datasets maintained ≥ 70% feature coverage, suggesting that while HTS is

over-engineered for legacy clocks, it ensures headroom for future *de novo* model development. The Twist panel achieved the highest overall feature retention, demonstrating a more faithful preservation of the legacy methylation array (e.g., 450K) sites that constitute the training foundation for the majority of established epigenetic clocks.

CpG-level reproducibility and technical noise

Technical precision was evaluated using 136,437 "Method-shared CpGs" which are observed in all datasets. An ideal methylation profiling method requires both reproducibility (agreement between technical replicates) and consistency (agreement across other methods). However, PCA revealed a systematic platform bias separating HTS-based and array-based methods (Fig. 1b-c). Notably, HTS platforms exhibited significantly lower Intraclass Correlation Coefficients (ICCs) [42] compared to arrays ($P < 1\times10^{-5}$; Fig. 1d). Analysis of the joint distribution showed that low-stability sites (ICC ≤ 0.3) were enriched at beta-value extremes (0 or 1), and the ratio of these sites was significantly higher in HTS datasets (60.5-79.9% of total CpGs) than in array-based datasets (30.8-35.7%) (Fig. S2). Consistent with previous studies, beta-values near 0 or 1 are unreliable and associate with low ICCs [12,28]. This increased stochastic noise in HTS translated to reduced age-prediction reproducibility, with ICCs of -0.188 – 0.997 (median 0.721) compared to 0.469 – 0.999 (median 0.950) for arrays (Fig. 1e). Inter-dataset beta-value correlations further confirmed superior technical consistency within array-based methodologies, with Pearson's R reaching 0.9947 between the two array platforms, compared to a range of 0.9125 – 0.9781 for all other cross-platform or HTS-to-HTS comparisons (Fig. 1f).

Clock categorization and cross-platform robustness

Based on the feature coverage and predictive stability evaluated above, we identified 4 categories among the 53 tested array-based clocks (Fig. 1g, Table S1). Six models, including five PCA-preprocessed clocks and zhangblup, were designated "Hi-Repro" for superior reproducibility (ICC ≥ 0.9; median replicate difference (RD) 0.17-3.81; Fig. S3, S4). Additionally, 20 models were classified as "Hi-Cov-All" (≥ 80% coverage across all platforms) and 12 as "Hi-Cov-nonMSA" (same criteria but excluding MSA). Notably, Hi-Repro clocks demonstrated the highest cross-platform correlation in gDNA, indicating their superior consistency and resilience against platform-related methylation variations. However, these clocks still suffered performance declines in cfDNA (Fig. 1h, S5), likely due to high fragmentation and multi-organ origins

reducing the signal-to-noise ratio. Notably, only 17 clocks, including all Hi-Repro clocks, achieved Pearson's R ≥ 0.7 across all platforms, with pcgrimage demonstrating the highest consistency (Fig. S5).

Impact of feature selection and platform divergence on predictive stability

To determine if biological feature selection confers cross-platform technical resilience, we benchmarked 2,727 "Clock-shared CpGs" shared across the Hi-Cov-All and Hi-Repro subsets. While this curated subset achieved higher global consistency (mean Pearson's R = 0.97; Fig. S1h) than the broader Method-shared CpGs (mean Pearson's R = 0.95; Fig. 1f), it also retained platform-specific noise (Fig. S1e-f). Array-based platforms maintained superior consistency (Fig. S1c-d) and reproducibility (Fig. S1g) over HTS platforms. This suggests that while feature selection effectively filters out biological noise, it cannot mitigate the fundamental technical stochasticity inherent in HTS profiling, necessitating further computational adaptation.

Benchmarking the Hi-Repro, Hi-Cov-All, and Hi-Cov-nonMSA subsets across SRRSH-24 datasets revealed a significant accuracy penalty for sequencing-based cfDNA. The gDNA_EPICv2 array served as the technical reference, yielding the highest accuracy (MAE = 2.03 – 23.08, median 7.39; Fig. 2a-b, S6). Conversely, cfDNA profiles exhibited higher MAEs (5.26 - 18.75, median 10.83) and lower correlations with CA (0.40-0.92, median 0.76) than gDNA (MAE = 5.02 - 22.92, median 7.80; R = 0.44-0.96, median 0.80) (Fig. 2b-c). We identified a strong correlation between prediction ICCs and accuracy metrics (Fig. 2d-e), confirming that reproducibility and accuracy are jointly compromised by common technical noise in HTS environments. Among the evaluated models, zhangblup demonstrated the greatest cross-platform resilience (MAE ≤ 5.2 years), though its performance was substantially compromised in the gDNA_MSA dataset due to critical (25%) feature coverage deficits (Fig. 2a, S1b, S7). Furthermore, a reciprocal accuracy and reproducibility decline was observed with HTS-trained clocks even when applied to HTS-based datasets (Fig. 2f). More intriguingly, using original models, array-trained clocks (hannum, etc.) outperformed HTS-trained models (timeseqage) on an array dataset (GSE245628), while the pattern reversed on an HTS-based dataset (GSE232346) (Fig. 2g-h). These results indicate that clocks are inherently tuned to dataset- or platform-specific patterns [20,25,43], reinforcing the need for targeted adaptation strategies even for HTS-based clocks. In this study, further benchmarking and adaptation focused

on array-trained clocks because of their better standardized feature set and more versatile clinical applications beyond CA predictions.

Optimizing sequencing depth for HTS-based epigenetic clock applications

To investigate the factors underlying the diminished technical reproducibility in sequencing-based cfDNA datasets, we identified sequencing depth as a primary HTS-specific source of variation and systematically investigated its impact. We assessed its impact on beta-value stability by analyzing the stochasticity of observed read counts and the inherent errors in ratio approximation. Benchmarking and randomized downsampling of the SRRSH-141 cohort demonstrated that MAE and RD stabilize at a minimum mean target depth of 10×, with 20× required for optimal stability (Fig. 3a-d, S8). This threshold is consistent with existing guidelines for differential methylation region calling in whole-genome bisulfite sequencing [44]. Notably, all HTS datasets generated in the present study satisfied this 10× minimum requirement (Fig. S9).

Feature density and L2 regularization enhance reproducibility

To identify the primary drivers of clock reproducibility, we analyzed 20 High Coverage/Reproducibility Linear (HCRL) clocks, which are defined as linear models within the Hi-Cov and High-Repro categories, across 11 datasets (Table 1, Fig. S10). We observed a robust negative correlation between CpG feature count and RD, indicating that high feature density and PCA-preprocessing significantly enhance resilience to low technical reproducibility. These results reinforce prior evidence that broader feature sets and dimensionality reduction provide a structural buffer against platform-specific noise [12,45], suggesting that the aging-related signal may spread across a vast majority of CpGs.

To isolate the mechanisms underlying this stability, we systematically tuned the elastic net hyperparameters: regularization strength (α) and the L1/L2 mixing parameter (λ). A critical α threshold (~0.01) was identified, beyond which accuracy degraded sharply (Fig. 3e). Within the stable range ($\alpha \leq 0.01$), the correlation between feature count and RD was reaffirmed (Fig. 3f), and sensitivity analysis showed that λ is a significantly more potent determinant of reproducibility (R^2 = 0.67-0.78; p = 1.6×10^{-47} - 7.3×10^{-33}) than α (R^2 = 0.05-0.21; p = 9.7×10^{-9} - 8.2×10^{-3}; Fig. 3g-h). The strong positive correlation between λ and RD demonstrates that shifting models toward an L2 (Ridge) penalty—which retains a broad distribution of small-effect coefficients—effectively minimizes the impact of stochastic variation in HTS data. Consequently,

L2-heavy architectures are essential for achieving reproducible predictions in sequencing-based applications.

Clock accuracy improvement with the DF-IM-TL adaptation pipeline

To mitigate CpG-level stochasticity, we developed a three-stage HTS data processing pipeline: depth filtering (DF), imputation (IM), and transfer learning (TL).

Stochastic noise mitigation via DF and IM. The DF stage screens out beta-values below specified depth thresholds, which are subsequently handled by IM. Benchmarking eight HTS datasets revealed that technical artifacts at distribution extremes (0 or 1) account for 18.4%-61.3% of total CpGs, overlapping with 2.0%-34.7% of features in 20 HCRL clocks (Fig. 4a, S11-S12, Table S3). 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. S13-S14). The IM step significantly enhanced model reproducibility despite these unreliable beta-value extrema. K-nearest neighbor (KNN) imputation [46] consistently reduced RD across all platforms, yielding improvements of 0.7-2.1 years in cfDNA and 0.7-1.1 years in gDNA (Fig. 4b). Meanwhile, other common imputation methods such as mean, median exhibited RD improvements and MAE trade-offs comparable to KNN, except for 0-filling (Fig. S15), potentially because heuristic imputation methods may amplify the stochastic effects related to cell type heterogeneity as observed in cfDNA samples. Notably, zhangblup emerged as the most resilient architecture, showing consistent improvements across all datasets (Fig. S16).

TL for cross-platform alignment. To align array-derived clocks with HTS data, we utilized model distillation (Teacher-Student TL). This model-agnostic approach uses an array-trained "teacher model" to guide an HTS-specific "student model". To reduce overfitting, we also incorporated PCA-based dimensionality reduction. Five-fold cross-validation confirmed that distillation significantly reduces prediction error across six first- and second-generation representative clocks and stabilizes reproducibility (Fig. 4d-e). Distillation effectively normalized RD across platforms, stabilizing high-variance models like dnamphenoage. Crucially, this recalibration selectively mitigated platform-specific variance while remaining sensitive to DNA source; performance gains were maximized when adaptation and testing datasets shared the same DNA type (Fig. 4f).

Systematic validation and functional benchmarking of the DF-IM-TL pipeline

We evaluated eight serial combinations of the DF-IM-TL pipeline, with the three steps applied in DF → IM → TL order. In the Buccal cohort (N=100), TL-inclusive strategies universally outperformed others, with the TL-only strategy performing best, yielding MAE reductions of 1.5 - 42.0 (median 10.3) years and CA correlation improvement of 0.02 - 0.78 (median 0.15) (Fig. 5a-b, S17). Because CA is an imperfect proxy for biological aging, we further assessed adaptation performance against PhenoAge [8] in the SRRSH-141 cohort, which confirmed the pipeline's biological sensitivity. PhenoAge provides a clinically oriented measure of aging by integrating CA with nine plasma biomarkers. The IM+TL strategy achieved the highest alignment with clinical aging metrics, reducing MAE by 1.0 - 23.4 (median 6.6) years and improving Pearson's R by 0.03 - 0.77 (median 0.17) (Fig. 5c, S18-S20). Notably, the DF-IM-TL pipeline outperformed the existing MAPLE framework in MAE while maintaining comparable reproducibility (Fig. 5a, S18), regardless of whether MAPLE was used with its pre-trained model or trained from scratch. MAPLE is specifically designed to handle cross-platform biases [47]. The DF-IM-TL pipeline also outperformed three other commonly used adaptation algorithms—quantile mapping, ComBat, and correlation alignment (CORAL)—each attempting to align HTS datasets to gDNA_EPICv2 before prediction with original HCRL clocks. The results in Fig. 5d suggest DF-IM-TL's highly specialized efficacy in correcting HTS platform biases in epigenetic age prediction.

As proof-of-concept for disease sensitivity, we also benchmarked the IM+TL framework's utility for disease detection using an amyotrophic lateral sclerosis (ALS) cfDNA dataset (N=24; GSE164600), where epigenetic delta-age (difference between BA and CA) has proven to be a discriminative biomarker in recent studies [48,49]. Adaptation significantly increased the Jensen-Shannon Divergence (JSD) between ALS and control samples (Fig. 6a-b). This enhanced distributional separation consequently improved diagnostic performance. An SVM classifier utilizing delta-age from the top four adapted clocks achieved a maximum AUC improvement of 0.125 over non-adapted clocks (Fig. 6c-d). These results demonstrate that the pipeline successfully recovers biological signals from HTS technical noise, enabling high-performance disease-centric liquid biopsy applications.

Discussion

The operational integration of array-trained epigenetic clocks with HTS-based cfDNA methylation data demands a rigorous, quantitative reconciliation of platform-specific biases. To address the scarcity of comprehensive cross-platform benchmark resources, we generated the SRRSH-24 dataset, providing a comprehensive technical replicate matrix across two DNA types and profiling technologies. Our benchmarking exposes a fundamental challenge: HTS methodologies introduce significant technical stochasticity, manifested as a loss of CpG-level ICC compared to arrays as the state-of-the-art standard. Critically, we demonstrate that technical reproducibility is an essential yet insufficient prerequisite [45] for biological signal preservation. We therefore propose a shift in the evaluative framework: future HTS-ready biomarkers must also prioritize RD alongside traditional accuracy metrics like MAE to effectively decouple technological noise from genuine biological information.

A robust epigenetic clock should inherently buffer against platform-specific noise. Our analysis reveals that high-density feature sets provide a “collective stability” effect, where small-effect signals distributed across numerous CpGs mitigate the stochastic fluctuations of individual sites. For linear architectures, this necessitates a fundamental departure from the sparsity-oriented L1 (LASSO) regularization typically used in array-based models, to L2 (Ridge) regularization for HTS-based applications. This L2-heavy preference was verified across our elastic net benchmarks and is consistent with PCA-preprocessed [12] and BLUP-based models (e.g., zhangblup [10]; see Supplementary Discussion for their mathematical relationship with the L2 regularization), and achieved a mean RD as low as 1.6 years across HTS datasets (Fig. S3, Fig. S4a). We therefore recommend L2 regularization as the primary practical strategy for fortifying legacy linear clocks against sequencing-related noise.

To standardize the HTS-based epigenetic landscape, we propose a three-stage adaptation pipeline: Depth Filtering (DF), Imputation (IM), and Transfer Learning (TL).

- By formalizing a 10× mean target depth threshold (20× for optimal stability), we ensure that the age prediction RD and MAE related to technical reproducibility can be confined to a manageable range. These guidelines align HTS-based epigenetic clock applications with established standards for whole genome bisulfite sequencing [44,50].

- Imputation of unreliable beta-values can reduce RD and stabilize overall reproducibility. However, its application to cfDNA may introduce notable accuracy biases. This trade-off is likely driven by the heightened cellular heterogeneity of cfDNA compared to genomic DNA [30], which may restrict pan-tissue generalizability [25]. We hypothesize that incorporating cell-type deconvolution as a pre-processing step prior to imputation could mitigate these trade-offs by providing more biologically informed priors for missing values [30,51].
- Transfer learning via model distillation emerged as the most critical component of our framework. Unlike de novo training approaches such as timeseqage and BS-clock [25,33], distillation is a model-agnostic tool that retains the biological interpretability of the original epigenetic clocks. Notably, even clocks explicitly designed for cross-platform robustness (e.g., MAPLE [47]), when tested in this study, failed to guarantee cross-platform performance. Hence, model-agnostic adaptation methods such as distillation are invaluable for epigenetic clock adaptation without requiring retraining of the original clock.

A pivotal discovery of this study is that HTS capture panels are interchangeable in age prediction performance under transfer learning. Once a student model is trained, it maintains reliability across disparate platforms (e.g., Twist vs. Galaxy), provided the DNA substrate remains consistent. This suggests that our pipeline corrects for generalized sequencing-related variation rather than idiosyncratic panel artifacts. Furthermore, validation against clinical PhenoAge and ALS datasets confirms that the DF-IM-TL pipeline preserves and enhances biological sensitivity. The framework does not merely align predictions with chronological age; it sharpens the biological signal necessary to distinguish disease states, as evidenced by superior JSD and AUC metrics in ALS classification.

While this study establishes a robust methodological foundation, certain limitations remain. Due to the limited sample size and lack of data, no sex- or ethnicity/race-related effects were evaluated. Furthermore, all datasets were cross-sectional without longitudinal and intervention-oriented analysis, precluding insights into temporal methylation dynamics. The absence of these analyses may limit the generalizability of our findings. The dataset size and cross-sectional design preclude assessments of sex- or ethnicity-related effects and temporal methylation

dynamics. Future work should extend to longitudinal cohorts to evaluate how adaptation influences sensitivity to age-related interventions. While linear models were prioritized for their interpretability and high performance in high-dimensional spaces, the true biological relationship between methylation and aging is not necessarily linear. Therefore, though linear epigenetic clocks can already achieve high accuracy [10,12,52], future research into advanced clocks and domain adaptation methods may further close the gap [36,53–55]. Additionally, the challenge of tissue-of-origin heterogeneity in cfDNA remains a critical variable that warrants further investigation to optimize imputation and recalibration strategies. Finally, a recently developed correction method for array platforms (e.g., ComBat-met [35]) remains to be tested and verified in the HTS environment.

Conclusion

This study provides a comprehensive methodological framework for the cross-platform application of epigenetic clocks. By contributing the SRRSH-24 benchmarking resource, formalizing depth requirements, and validating the DF-IM-TL pipeline, we offer a clear path towards HTS-based aging research. The proposed pipeline enables the extensive collection of array-derived epigenetic clocks to be reliably deployed in the emerging field of HTS-based epigenetic aging and liquid biopsy diagnostics.

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 accession HRA014554.

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.

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) the subject has completed at least two routine health evaluations within the past 5 years that included complete blood count, urinalysis, fasting plasma glucose, lipid profile, liver-function tests, renal-function tests, and at least one set of imaging examinations: 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 (referred to as Galaxy) and targeted sequencing with Twist Human Methylome Panel (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, which covers 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; 41.8 % female) were recruited as the SRRSH cohort. From this cohort, 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 was sufficient to identify 80% of reliable probes [56]. Their cfDNA was extracted for targeted DNA methylation sequencing, and buffy coat gDNA was extracted for array-based and HTS-based 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. Finally, the SRRSH-141 dataset contained Twist-profiled plasma cfDNA methylation data for the remaining 141 participants. In addition, nine phenoage-related [8] plasma biomarkers were collected from the SRRSH-141 cohort, namely albumin, creatinine, glucose, CRP, lymphocyte percent, mean cell volume, red blood cell distribution width, alkaline phosphatase and white blood cell count. The sample details are listed in Table S2.

Epigenetic clock selection

A total of 57 epigenetic clocks were included in this study, including 53 array-trained and 4 HTS-trained clocks (Table S1). The 53 array-trained clocks include 47 integrated in the pyaging package (version 0.1.27) and 6 GP-age clocks (https://github.com/mirivar/GP-age, version 1.0). The 4 HTS-trained clocks include emseqage [20], timeseqage (https://github.com/patricktgriffin/TIME-Seq, Commit 140622f), BS-clock (https://github.com/hucongcong97/BS-clock, Commit 097b800) and komakiage [43]. Two clocks, BS-clock and komakiage were not included in analyses beyond feature coverage assessment due to their overall low feature coverage in all six SRRSH-24 datasets.

Biological sample pretreatment and storage

An 8-10 mL whole-blood sample was collected from each volunteer via venipuncture, 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 manufacturer'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 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 bases per library (Fig S15).

EPICv2 and MSA

Buffy coat genomic DNA samples were sent to Novogene (Tianjin, China) to be analyzed using the Illumina Infinium MethylationEPIC v2.0 BeadChip array (EPICv2; 937,691 probes), or sent to WeGene (Shenzhen, China) for profiling on the Infinium Methylation Screening array (MSA; 284,318 probes). 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.

As the EPICv2 and MSA arrays can accommodate 8 and 48 samples per array, respectively, we included two replicates (2 replicates × 24 gDNA samples = 48 slots).

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) 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 regarded 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` method. Finally, beta-values were calculated as the primary methylation metric and prepared for downstream quantitative analyses.

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, the raw reads underwent adapter removal using `fastp` (version 0.23.4) and were aligned to a combined reference sequence set, comprising GRCh38, pUC19, and phage lambda sequences, using `bwa-meth` (version 0.2.7). Alignment files were then sorted and indexed using `Samtools` (version 1.21), followed by the identification and marking of duplicate reads using `Samblaster` (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 site and then merged using `mergeContext` function in `MethylDackel`. 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.

Dataset benchmarking calculations

Principal component analysis (PCA) was performed using `prcomp()` from R version 4.5.1 (see below). Pearson correlation analysis was performed using `cor()`. Intraclass correlation coefficient (ICC) was calculated using `icc()` from the `irrR` package (version 0.84.1) based on a single-rater, absolute agreement, two-way random-effects model. Missing values (`NaN`) were removed before these analyses. Benchmarking results were visualized using R packages `ggplot2` (version 4.0.0), `ComplexHeatmap` (version 2.22.0), and `linkET` (version 0.0.7.4) respectively.

CpG coverage analyses were conducted in two levels: (1) across platform; and (2) clock CpGs covered by each platform. Reference CpG sites for each platform were extracted from manufacturer-provided datasheets. In addition, for MSA and EPICv2 arrays, only CpGs with defined genomic coordinates were retained. Input CpGs for each of the 53 clocks were extracted from clock models respectively. Results were visualized using R functions `UpSet()` (`ComplexHeatmap` version 2.22.0) and `pheatmap()` (`pheatmap` version 1.0.13).

Statistical analysis

The comparisons between means and medians of two groups of data were conducted using permutation test (implemented via a 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 numbers 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 11 publicly available datasets. GSE55763 (N = 2711, Illumina HumanMethylation 450K) [58] 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 using the 72 replicated samples. GSE83944 (N = 48, Illumina HumanMethylation 450K) [59], GSE247193 (N = 23 including one subject without replicate, Illumina Infinium EPIC Human Methylation Beadchip) [60], GSE247195 (N = 23 including one subject without replicate, Illumina Infinium EPIC Human Methylation Beadchip) [60] and GSE247197 (N = 46, Illumina HumanMethylation 450K) [60] were downloaded from the GEO database to assess age prediction RD as they all contain technical replicates. HTS-based datasets GSE144691 (N = 29) [61] and GSE86832 (N = 4) [62] were downloaded from the GEO database for assessment of unreliable CpGs, because they lacked replicates. Two additional datasets from the GEO database were downloaded for assessment of age prediction MAE and RD, namely GSE232346 (N = 24, TimeSeq) [33] and GSE245628 (N = 24, Illumina Infinium EPIC Human Methylation Beadchip) [33]. Finally, we also downloaded the Buccal cohort (Shokhirev & Johnson, 2025, N = 100, Twist and Illumina Infinium EPIC Human Methylation Beadchip v2) [28] and GSE164600 (N = 24, cfDNA WGBS) [51] datasets 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 CpG number and RD

20 HCRL clocks were selected to predict biological age with `pyaging` (version 0.1.27). 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 linear regression was performed 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 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 90,394 CpGs shared between GSE55763 and the 6 datasets (gDNA_EPICv2, gDNA_MSA, gDNA_Galaxy, gDNA_Twist, cfDNA_Galaxy and cfDNA_Twist) were extracted, 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, resulting in a total of 450 trained models. All other parameters were kept default. Linear regression was performed using `scipy.stats.linregress`.

Minimal sequencing investigation based on binomial stochasticity.

Randomized samples were generated via multi-dimensional binomial model, where the observed methylation counts at each CpG follows independent binomial distributions. 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.

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.

Minimal sequencing investigation based on downsampling.

A downsampling analysis was conducted on five samples, each having an initial sequencing result size ≥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.

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 was 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 was summarized from all testing splits. The hyperparameters (`alpha` and `l1_ratio`) were optimized using an independent 5-fold cross-validation (`sklearn.linear_models.ElasticNetCV(cv=5)`).

Adaptation baseline methods comparison.

Three adaptation methods were evaluated as baselines for comparison with the DF-IM-TL pipeline: quantile mapping, ComBat, and correlation alignment (CORAL). In all cases, target datasets (gDNA_Galaxy, gDNA_Twist, cfDNA_Galaxy, cfDNA_Twist, and Buccal) were adapted to the gDNA_EPICv2 reference prior to age prediction with the original array-trained clocks.

Quantile mapping was applied feature-wise by transforming values in the target dataset to match the empirical distribution of the reference dataset via quantile-to-quantile alignment.

Batch correction using ComBat was performed under an empirical Bayes framework to remove additive and multiplicative batch effects while preserving biological variation.

CORAL was used to align second-order statistics between datasets by matching covariance structures. To accommodate the high dimensionality of methylation data, dimensionality reduction was first performed using truncated singular value decomposition (SVD). CORAL was then applied in the reduced space, followed by projection back to the original feature space.

Quantile mapping and CORAL were implemented using custom scripts using scipy and numpy, while ComBat was applied using the neuroCombat (version 0.2.12) Python package.

MAPLE training

For a head-to-head comparison with MAPLE, we also evaluated a MAPLE model trained from scratch in addition to using its pretrained age prediction model. The training was performed with the entire SRRSH dataset (6 SRRSH-24 datasets and the SRRSH-141 dataset) with a 5-fold cross-validation. The use of mixed array-based and HTS-based samples is to ensure maximum efficiency of MAPLE's pairwise learning architecture. For evaluation, only the predictions of the SRRSH-141 samples in the testing split were considered, ensuring comparability with other DF-IM-TL results. The input dataset contained 70,110 CpGs, the same as the original MAPLE model. Default neural network layer configuration was used, with 1024, 1024, 512, 512, 256, 256, 64, 64 encoder hidden layer dimensions and 32, 32, 16, 16 decoder hidden layer dimensions. The model was trained with encoder regularization on, learning rate of 0.0001, batch size of 256 and 6000 epochs.

PhenoAge calculation.

The PhenoAge of the samples from the SRRSH-141 cohort was calculated from collected clinical biomarkers using the phenoage clock [8] integrated in `pyaging`.

JSD calculation.

Jensen-Shannon Divergence (JSD) was calculated on delta-age predicted by each clock individually. The distribution of delta-age probability density functions (p.d.f.) were estimated using Gaussian kernel density estimation (KDE) (`scipy.stats.gaussian_kde`) independently for normal and diseased samples. JSD was then calculated between the normal and disease sample p.d.f.'s using `scipy.spatial.distance.jensonshannon`. The bandwidth parameter in KDE is automatically determined using the "scott" method.

ALS classification.

The ALS classification was performed via support vector machine (SVM) classifier with a linear kernel (`sklearn.svm.SVC(kernel="linear")`) using the delta-ages calculated before and after model adaptation respectively. The classification predicts if a sample is ALS (1) or normal (0). A paired leave-one-out procedure was repeated 12 times to evaluate classification performance while reducing potential overfitting. Briefly, 1 normal and 1 ALS sample were randomly chosen (without replacement) as testing samples, and the rest were for training. Classification performance was evaluated after 12 iterations, i.e. all samples have been predicted as testing cases using the area under receiver operating characteristic curve (AUC) metric (`sklearn.metrics.auc_score`).

References

1. Horvath S, Raj K. DNA methylation-based biomarkers and the epigenetic clock theory of ageing. Nat Rev Genet. 2018;19:371–84. https://doi.org/10.1038/s41576-018-0004-3

2. Moqri M, Herzog C, Poganik JR, Justice J, Belsky DW, Higgins-Chen A, et al. Biomarkers of aging for the identification and evaluation of longevity interventions. Cell. 2023;186:3758–75. https://doi.org/10.1016/j.cell.2023.08.003

3. Biomarkers of Aging Consortium, Herzog CMS, Goeminne LJE, Poganik JR, Barzilai N, Belsky DW, et al. Challenges and recommendations for the translation of biomarkers of aging. Nat Aging. 2024;4:1372–83. https://doi.org/10.1038/s43587-024-00683-3

4. Varshavsky M, Harari G, Glaser B, Dor Y, Shemer R, Kaplan T. Accurate age prediction from blood using a small set of DNA methylation sites and a cohort-based machine learning algorithm. Cell Rep Methods. 2023;3:100567. https://doi.org/10.1016/j.crmeth.2023.100567

5. De Lima Camillo LP, Lapierre LR, Singh R. A pan-tissue DNA-methylation epigenetic clock based on deep learning. Npj Aging. Nature Publishing Group; 2022;8:4. https://doi.org/10.1038/s41514-022-00085-y

6. Sehgal R, Markov Y, Qin C, Meer M, Hadley C, Shadyab AH, et al. Systems Age: a single blood methylation test to quantify aging heterogeneity across 11 physiological systems. Nat Aging. 2025;5:1880–96. https://doi.org/10.1038/s43587-025-00958-3

7. Horvath S. DNA methylation age of human tissues and cell types. Genome Biol. 2013;14:3156. https://doi.org/10.1186/gb-2013-14-10-r115

8. Levine ME, Lu AT, Quach A, Chen BH, Assimes TL, Bandinelli S, et al. An epigenetic biomarker of aging for lifespan and healthspan. Aging. 2018;10:573–91. https://doi.org/10.18632/aging.101414

9. Teschendorff AE, Horvath S. Epigenetic ageing clocks: statistical methods and emerging computational challenges. Nat Rev Genet. 2025;26:350–68. https://doi.org/10.1038/s41576-024-00807-w

10. Zhang Q, Vallerga CL, Walker RM, Lin T, Henders AK, Montgomery GW, et al. Improved precision of epigenetic clock estimates across tissues and its implication for biological ageing. Genome Med. 2019;11:54. https://doi.org/10.1186/s13073-019-0667-1

11. Tong H, Dwaraka VB, Chen Q, Luo Q, Lasky-Su JA, Smith R, et al. Quantifying the stochastic component of epigenetic aging. Nat Aging. 2024;4:886–901. https://doi.org/10.1038/s43587-024-00600-8

12. Higgins-Chen AT, Thrush KL, Wang Y, Minteer CJ, Kuo P-L, Wang M, et al. A computational solution for bolstering reliability of epigenetic clocks: implications for clinical trials and longitudinal tracking. Nat Aging. 2022;2:644–61. https://doi.org/10.1038/s43587-022-00248-2

13. Conole ELS, Robertson JA, Smith HM, Cox SR, Marioni RE. Epigenetic clocks and DNA methylation biomarkers of brain health and disease. Nat Rev Neurol. 2025;21:411–21. https://doi.org/10.1038/s41582-025-01105-7

14. Klein EA, Richards D, Cohn A, Tummala M, Lapham R, Cosgrove D, et al. Clinical validation of a targeted methylation-based multi-cancer early detection test using an independent validation set. Ann Oncol. 2021;32:1167–77. https://doi.org/10.1016/j.annonc.2021.05.806

15. Moqri M, Herzog C, Poganik JR, Ying K, Justice JN, Belsky DW, et al. Validation of biomarkers of aging. Nat Med. 2024;30:360–72. https://doi.org/10.1038/s41591-023-02784-9

16. Gao Q, Zeng Q, Wang Z, Li C, Xu Y, Cui P, et al. Circulating cell-free DNA for cancer early detection. The Innovation. 2022;3:100259. https://doi.org/10.1016/j.xinn.2022.100259

17. Medina JE, Dracopoli NC, Bach PB, Lau A, Scharpf RB, Meijer GA, et al. Cell-free DNA approaches for cancer early detection and interception. J Immunother Cancer. 2023;11:e006013. https://doi.org/10.1136/jitc-2022-006013

18. Zhang K, Fu R, Liu R, Su Z. Circulating cell-free DNA-based multi-cancer early detection. Trends Cancer. Elsevier; 2024;10:161–74. https://doi.org/10.1016/j.trecan.2023.08.010

19. Bruhm DC, Vulpescu NA, Foda ZH, Phallen J, Scharpf RB, Velculescu VE. Genomic and fragmentomic landscapes of cell-free DNA for early cancer detection. Nat Rev Cancer. Nature Publishing Group; 2025;25:341–58. https://doi.org/10.1038/s41568-025-00795-x

20. Li S-J, Gao X, Wang Z-H, Li J, Zeng L-T, Dang Y-M, et al. Cell-free DNA methylation patterns in aging and their association with inflamm-aging. Epigenomics. 2024;16:715–31. https://doi.org/10.1080/17501911.2024.2340958

21. Sergeev AV, Kisil OV, Eremin AA, Petrov AS, Zvereva ME. “Aging Clocks” Based on Cell-Free DNA. Biochem Mosc. Pleiades Publishing Ltd; 2025;90:S342–55. https://doi.org/10.1134/s0006297924604076

22. Shtumpf M, Jeong S, Bikova M, Mamayusupova H, Ruje L, Teif VB. Aging clock based on nucleosome reorganisation derived from cell-free DNA. Aging Cell. 2024;23:e14100. https://doi.org/10.1111/acel.14100

23. Liu Y. At the dawn: cell-free DNA fragmentomics and gene regulation. Br J Cancer. 2022;126:379–90. https://doi.org/10.1038/s41416-021-01635-z

24. Infinium MethylationEPIC v2.0 BeadChip [Internet]. [cited 2025 Oct 23]. 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. Accessed 23 Oct 2025

25. 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. 2024;40:btae656. https://doi.org/10.1093/bioinformatics/btae656

26. Lee SM, Loo CE, Prasasya RD, Bartolomei MS, Kohli RM, Zhou W. Low-input and single-cell methods for Infinium DNA methylation BeadChips. Nucleic Acids Res. 2024;52:e38. https://doi.org/10.1093/nar/gkae127

27. Heiss JA, Brennan KJ, Baccarelli AA, Téllez-Rojo MM, Estrada-Gutiérrez G, Wright RO, et al. Battle of epigenetic proportions: comparing Illumina’s EPIC methylation microarrays and

TruSeq targeted bisulfite sequencing. Epigenetics. 2020;15:174–82. https://doi.org/10.1080/15592294.2019.1656159

28. Shokhirev MN, Johnson AA. Analysis of variability and epigenetic age prediction across microarray and methylation sequencing technologies. GeroScience [Internet]. 2025 [cited 2025 Sept 17]; https://doi.org/10.1007/s11357-025-01824-1

29. Shu C, Zhang X, Aouizerat BE, Xu K. Comparison of methylation capture sequencing and Infinium MethylationEPIC array in peripheral blood mononuclear cells. Epigenetics Chromatin. 2020;13:51. https://doi.org/10.1186/s13072-020-00372-6

30. Zhu T, Liu J, Beck S, Pan S, Capper D, Lechner M, et al. A pan-tissue DNA methylation atlas enables in silico decomposition of human tissue methylomes at cell-type resolution. Nat Methods. 2022;19:296–306. https://doi.org/10.1038/s41592-022-01412-7

31. Liu L, Toung JM, Jassowicz AF, Vijayaraghavan R, Kang H, Zhang R, et al. Targeted methylation sequencing of plasma cell-free DNA for cancer detection and classification. Ann Oncol. 2018;29:1445–53. https://doi.org/10.1093/annonc/mdy119

32. Buckley DN, Lewinger JP, Gooden G, Spillman M, Neuman M, Guo XM, et al. OvaPrint—A Cell-free DNA Methylation Liquid Biopsy for the Risk Assessment of High-grade Serous Ovarian Cancer. Clin Cancer Res. 2023;29:5196–206. https://doi.org/10.1158/1078-0432.CCR-23-1197

33. Griffin PT, Kane AE, Trapp A, Li J, Arnold M, Poganik JR, et al. TIME-seq reduces time and cost of DNA methylation measurement for epigenetic clock construction. Nat Aging. 2024;4:261–74. https://doi.org/10.1038/s43587-023-00555-2

34. Garma LD, Quintela-Fandino M. Applicability of epigenetic age models to next-generation methylation arrays. Genome Med. 2024;16:116. https://doi.org/10.1186/s13073-024-01387-4

35. Wang J. ComBat-met: Adjusting Batch Effects in DNA Methylation Data [Internet]. Bioinformatics; 2024 [cited 2025 Oct 18]. https://doi.org/10.1101/2024.08.13.607838

36. Ying K, Paulson S, Eames A, Tyshkovskiy A, Li S, Eynon N, et al. A unified framework for systematic curation and evaluation of aging biomarkers. Nat Aging [Internet]. 2025 [cited 2025 Nov 5]; https://doi.org/10.1038/s43587-025-00987-y

37. Handl L, Jalali A, Scherer M, Pfeifer N. Partially blind domain adaptation for age prediction from DNA methylation data [Internet]. arXiv; 2016 [cited 2025 Oct 18]. https://doi.org/10.48550/ARXIV.1612.06650

38. Luo L, Shang L, Goodrich JM, Peterson KE, Song PX-K. Bridging the gap: Enhancing the generalizability of epigenetic clocks through transfer learning [Internet]. Epidemiology; 2025 [cited 2025 Oct 18]. https://doi.org/10.1101/2025.02.14.25322080

39. Zindler T, Frieling H, Neyazi A, Bleich S, Friedel E. Simulating ComBat: how batch correction can lead to the systematic introduction of false positive results in DNA methylation

microarray studies. BMC Bioinformatics. 2020;21:271. https://doi.org/10.1186/s12859-020-03559-6

40. Tomo Y, Nakaki R. Transfer Elastic Net for Developing Epigenetic Clocks for the Japanese Population. Mathematics. 2024;12:2716. https://doi.org/10.3390/math12172716

41. Tomusiak A, Floro A, Tiwari R, Riley R, Matsui H, Andrews N, et al. Development of an epigenetic clock resistant to changes in immune cell composition. Commun Biol. 2024;7:934. https://doi.org/10.1038/s42003-024-06609-4

42. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. J Chiropr Med. 2016;15:155–63. https://doi.org/10.1016/j.jcm.2016.02.012

43. Komaki S, Nagata M, Arai E, Otomo R, Ono K, Abe Y, et al. Epigenetic profile of Japanese supercentenarians: a cross-sectional study. Lancet Healthy Longev. 2023;4:e83–90. https://doi.org/10.1016/S2666-7568(23)00002-8

44. Ziller MJ, Hansen KD, Meissner A, Aryee MJ. Coverage recommendations for methylation analysis by whole-genome bisulfite sequencing. Nat Methods. 2015;12:230–2. https://doi.org/10.1038/nmeth.3152

45. Sehgal R, Borrus D, Gonzalez J, Markov Y, ADNI, Higgins-Chen A. Biological versus Technical Reliability of Epigenetic Clocks and Implications for Disease Prognosis and Intervention Response [Internet]. Bioinformatics; 2025 [cited 2026 Jan 23]. https://doi.org/10.1101/2025.10.13.682176

46. Lena PD, Sala C, Prodi A, Nardini C. Methylation data imputation performances under different representations and missingness patterns. BMC Bioinformatics. 2020;21:268. https://doi.org/10.1186/s12859-020-03592-5

47. Zhang Y, Yao Y, Tang Y, Cheng Y, Xu Y, He Y, et al. A robust computational framework for methylation age and disease-risk prediction based on pairwise learning. Nat Comput Sci [Internet]. 2026 [cited 2026 Jan 21]; https://doi.org/10.1038/s43588-025-00939-x

48. Zhao Y, Li X, Wang K, Iyer G, Sakowski SA, Zhao L, et al. Epigenetic age acceleration is associated with occupational exposures, sex, and survival in amyotrophic lateral sclerosis. eBioMedicine. 2024;109:105383. https://doi.org/10.1016/j.ebiom.2024.105383

49. Merbaum P, Zwamborn R, Hop P, Sequencing Consortium PMA, Van Rheenen W, Veldink J. Dissecting epigenetic age acceleration in amyotrophic lateral sclerosis. Epigenomics. 2025;17:1153–61. https://doi.org/10.1080/17501911.2025.2563502

50. Zhou L, Ng HK, Drautz-Moses DI, Schuster SC, Beck S, Kim C, et al. Systematic evaluation of library preparation methods and sequencing platforms for high-throughput whole genome bisulfite sequencing. Sci Rep. 2019;9:10383. https://doi.org/10.1038/s41598-019-46875-5

51. Caggiano C, Celona B, Garton F, Mefford J, Black BL, Henderson R, et al. Comprehensive cell type decomposition of circulating cell-free DNA with CelFiE. Nat Commun. 2021;12:2717. https://doi.org/10.1038/s41467-021-22901-x

52. Galkin F, Naumov V, Pushkov S, Sidorenko D, Urban A, Zagirova D, et al. Precious3GPT: Multimodal Multi-Species Multi-Omics Multi-Tissue Transformer for Aging Research and Drug Discovery [Internet]. Bioinformatics; 2024 [cited 2025 June 12]. https://doi.org/10.1101/2024.07.25.605062

53. Ganin Y, Ustinova E, Ajakan H, Germain P, Larochelle H, Laviolette F, et al. Domain-adversarial training of neural networks. J Mach Learn Res. 2016;17:2096–2030.

54. Evgeniou T, Pontil M. Regularized multi--task learning. Proc Tenth ACM SIGKDD Int Conf Knowl Discov Data Min [Internet]. Seattle WA USA: ACM; 2004 [cited 2025 Oct 11]. p. 109–17. https://doi.org/10.1145/1014052.1014067

55. Ying K, Song J, Cui H, Zhang Y, Li S, Chen X, et al. MethylGPT: a foundation model for the DNA methylome [Internet]. Systems Biology; 2024 [cited 2025 Apr 9]. https://doi.org/10.1101/2024.10.30.621013

56. Sugden K, Hannon EJ, Arseneault L, Belsky DW, Corcoran DL, Fisher HL, et al. Patterns of Reliability: Assessing the Reproducibility and Integrity of DNA Methylation Measurement. Patterns. 2020;1:100014. https://doi.org/10.1016/j.patter.2020.100014

57. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics. 2014;30:1363–9. https://doi.org/10.1093/bioinformatics/btu049

58. Lehne B, Drong AW, Loh M, Zhang W, Scott WR, Tan S-T, et al. A coherent approach for analysis of the Illumina HumanMethylation450 BeadChip improves data quality and performance in epigenome-wide association studies. Genome Biol. 2015;16:37. https://doi.org/10.1186/s13059-015-0600-x

59. Oh G, Koncevičius K, Ebrahimi S, Carlucci M, Groot DE, Nair A, et al. Circadian oscillations of cytosine modification in humans contribute to epigenetic variability, aging, and complex disease. Genome Biol. 2019;20:2. https://doi.org/10.1186/s13059-018-1608-9

60. Koncevičius K, Nair A, Šveikauskaitė A, Šeštokaitė A, Kazlauskaitė A, Dulskas A, et al. Epigenetic age oscillates during the day. Aging Cell. 2024;23:e14170. https://doi.org/10.1111/acel.14170

61. Abdelsamed HA, Zebley CC, Nguyen H, Rutishauser RL, Fan Y, Ghoneim HE, et al. Beta cell-specific CD8+ T cells maintain stem cell memory-associated epigenetic programs during type 1 diabetes. Nat Immunol. 2020;21:578–87. https://doi.org/10.1038/s41590-020-0633-5

62. Pidsley R, Zotenko E, Peters TJ, Lawrence MG, Risbridger GP, Molloy P, et al. Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. Genome Biol. 2016;17:208. https://doi.org/10.1186/s13059-016-1066-1

63. De Lima Camillo LP. pyaging□: a Python-based compendium of GPU-optimized aging clocks. Kelso J, editor. Bioinformatics. 2024;40:btae200. https://doi.org/10.1093/bioinformatics/btae200

Tables

Table 1| Linear regression results between the age prediction replicate difference (RD) and the number of CpGs in 20 HCRL clocks. 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. S10.

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**, PCA of all samples colored by datasets (**b**) or replicates (**c**), using beta-values of Method-shared CpGs (N = 136,437). **d**, Intraclass Correlation Coefficients (ICCs) of Method-shared CpGs across six datasets. **e**, Comparisons of per clock ICCs of all array-trained clocks (N = 53). **f**, Pearson correlations among datasets computed using the mean beta-values of Method-shared CpGs. **g**, ICCs of 55 epigenetic (53 array trained and 2 HTS-trained) 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 ≤ ICC, Good: 0.75 ≤ ICC < 0.9, Moderate: 0.5 ≤ ICC < 0.75, Poor: ICC < 0.5. Clocks were classified according to 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 methylation profiling methods. **h**, Pearson correlations across six datasets calculated using the mean of clock predictions for three clock performance categories. P-values in (**d**) and (**e**) 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 quartiles, the caps mark the min and max values.

Fig. 2| Decline of epigenetic clock performance in 6 SRRSH-24 HTS-based datasets. a, Lollipop plot showing Pearson's R and MAE for 6 Hi-Repro clocks. Each dot represents a replicate. **b-c**, Boxplots showing comparisons and statistical significance of (**b**) MAE and (**c**) Pearson's R for Hi-Cov-All clocks. The center line of each boxplot marks the median, the box body marks the 1st and 3rd quartiles, and the caps mark the min and max of RD values respectively. **d**, Scatter plot showing the correlation between clock ICC and MAE in the four SRRSH-24 HTS datasets. **e**, Scatter plot showing the correlation between clock ICC and R in the four SRRSH-24 HTS datasets. Lines represent linear regression fits for each replicate group respectively. **f**, Lollipop plot showing Pearson's R and MAE for the 2 HTS-trained clocks. Each dot represents a replicate. **g-h**, Performance evaluation of 4 array-trained clocks (hannum, horvath2013, skinandblood and dnamphenoage) and an HTS data-trained clock (timeseqage) cross-applied to the 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 analyzed. The p-values in (**b**) and (**c**) 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 (**d**) and (**e**) were calculated by Pearson correlation at the replicate level.

Fig. 3| Optimization of epigenetic clock applications for HTS-based methylation data via sequencing depth and model selection. a, Investigation of the sequencing depth threshold via beta-value stochasticity effect at individual CpG level on HCRL clocks. **b**, Investigation of sequencing depth threshold via beta-value approximation precision on HCRL clocks. **c-d**, Investigation of the sequencing depth requirements with downsampling simulations with 5 qualified samples from the SRRSH cohort with sequencing result size ≥ 50Gb. **c**, Age prediction MAE at different downsampling depths. **d**, The increasing of the fraction of CpGs with ≥ 20× slowed down above ≥ 20× downsampling depth. **e**, A series of elastic nets were trained on the dataset GSE55763 (N = 2711) and tested with the 6 datasets from this study. **f**, The negative correlation

observed between the number of CpGs and age prediction RD. **g-h**, Comparison of correlation between hyperparameter (α and λ) and RD. R^2 and p-values were calculated by Pearson correlation.

Fig. 4| Investigating depth filtering (DF), imputation (IM) and transfer learning (TL) to improve epigenetic clock performance across array and HTS platforms. a, HCRL clocks are prone to unreliable beta-values originating from HTS-based methylation profiling technologies (Details are shown in Table S3). **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**, Comparison of age prediction MAE (**d**) and RD (**e**) between teacher and student (distillation) clocks, assessed with 5-fold cross validation. **f**, Generalizability benchmarking of transfer learning across HTS capture panels and DNA types. The center line of each boxplot marks the median, the box body marks the 1st and 3rd quartiles, and the caps mark the min and max values.

Fig. 5| Testing and validating the DF-IM-TL epigenetic clock adaptation pipeline for biological age prediction on HTS platforms. a, Evaluation of all DF-IM-TL combinations for 20 HCRL clocks on the Buccal cohort (N = 100). **b**, Performance improvements via the best identified strategy (TL-only) with respect to chronological age (CA) on the Buccal cohort. **c**, Performance improvements via the best identified strategy (IM-TL) with respect to biomarker-derived PhenoAge on the SRRSH-141 cohort (N = 141). **d**, Adaptation efficacy comparison between DF-IM-TL and three other commonly used adaptation strategies.

Fig. 6| Testing and validating the DF-IM-TL epigenetic clock adaptation strategies for disease classification applications. Validation performed on a public dataset GSE164600 (N = 24) with healthy and amyotrophic lateral sclerosis (ALS) labels. **a**, The distributional change of delta-age predicted by 20 HCRL clocks before and after clock adaptation (IM+TL as the best strategy). The center line in each violin plot marks the median. **b**, The Jensen-Shannon Divergence (JSD) calculated to evaluate the distributional distances between ALS and normal samples. The models in (**a**) were sorted in descending order by ΔJSD with respect to clock adaptation. Each box marks the 1st -3rd quartile range, caps mark minimum and maximum values, and the center line marks the median. **c**, ALS classification performance collectively using delta-ages predicted by top-ΔJSD models. Classification was conducted using support vector machine (SVM) classifier with a linear kernel and was evaluated using area under the receiver operating characteristic curve (AUC). **d**, Maximum AUC achieved using as few as 4 top ΔJSD clocks, surpassing the non-adapted prediction by 0.125 AUC.

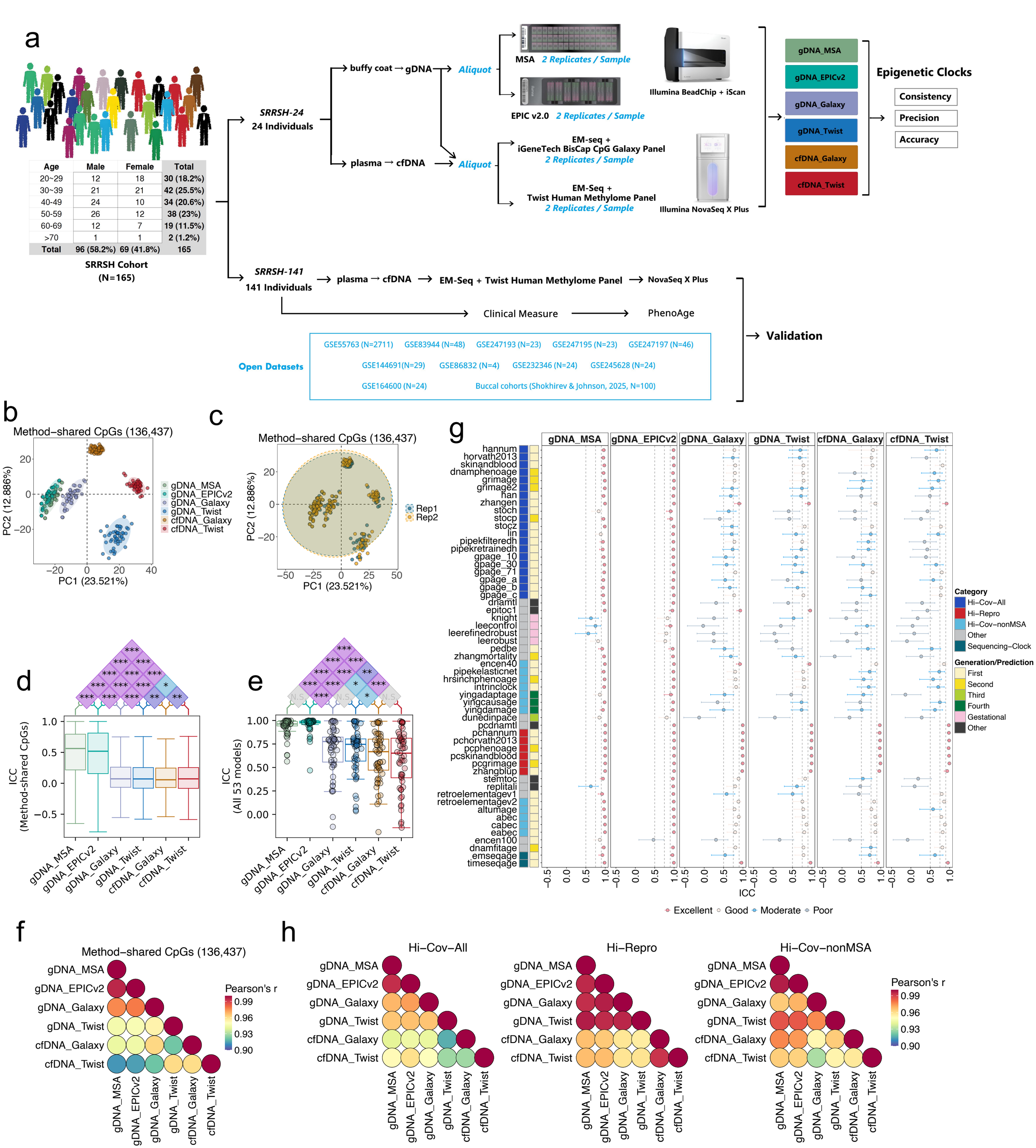
a
Age | Male | Female | Total
20~29 | 12 | 18 | 30 (18.2%)
30~39 | 21 | 21 | 42 (25.5%)
40-49 | 24 | 10 | 34 (20.6%)
50-59 | 26 | 12 | 38 (23%)
60-69 | 12 | 7 | 19 (11.5%)
>70 | 1 | 1 | 2 (1.2%)
Total | 96 (58.2%) | 69 (41.8%) | 165
SRRSH Cohort (N=165)
SRRSH-24
24 Individuals
buffy coat → gDNA
plasma → cfDNA
Aliquot
MSA 2 Replicates / Sample
EPIC v2.0 2 Replicates / Sample
Illumina BeadChip + iScan
EM-seq + iGeneTech BisCap CpG Galaxy Panel
2 Replicates / Sample
EM-Seq + Twist Human Methylome Panel
2 Replicates / Sample
Illumina NovaSeq X Plus
gDNA_MSA
gDNA_EPICv2
gDNA_Galaxy
gDNA_Twist
cfDNA_Galaxy
cfDNA_Twist
Epigenetic Clocks
Consistency
Precision
Accuracy
SRRSH-141
141 Individuals
plasma → cfDNA → EM-Seq + Twist Human Methylome Panel → NovaSeq X Plus
Clinical Measure → PhenoAge
Open Datasets
GSE55763 (N=2711) GSE83944 (N=48) GSE247193 (N=23) GSE247195 (N=23) GSE247197 (N=46)
GSE144691(N=29) GSE86832 (N=4) GSE232346 (N=24) GSE245628 (N=24)
GSE164600 (N=24) Buccal cohorts (Shokhirev & Johnson, 2025, N=100)
Validation
b
Method−shared CpGs (136,437)
PC2 (12.886%)
PC1 (23.521%)
gDNA_MSA
gDNA_EPICv2
gDNA_Galaxy
gDNA_Twist
cfDNA_Galaxy
cfDNA_Twist
c
Method−shared CpGs (136,437)
PC2 (12.886%)
PC1 (23.521%)
Rep1
Rep2
d
ICC (Method-shared CpGs)
e
ICC (All 53 models)
g
ICC
Category
Hi−Cov−All
Hi−Repro
Hi−Cov−nonMSA
Other
Sequencing−Clock
Generation/Prediction
First
Second
Third
Fourth
Gestational
Other
Excellent
Good
Moderate
Poor
f
Method−shared CpGs (136,437)
Pearson's r
h
Hi−Cov−All
Hi−Repro
Hi−Cov−nonMSA
Pearson's r

Fig. 1

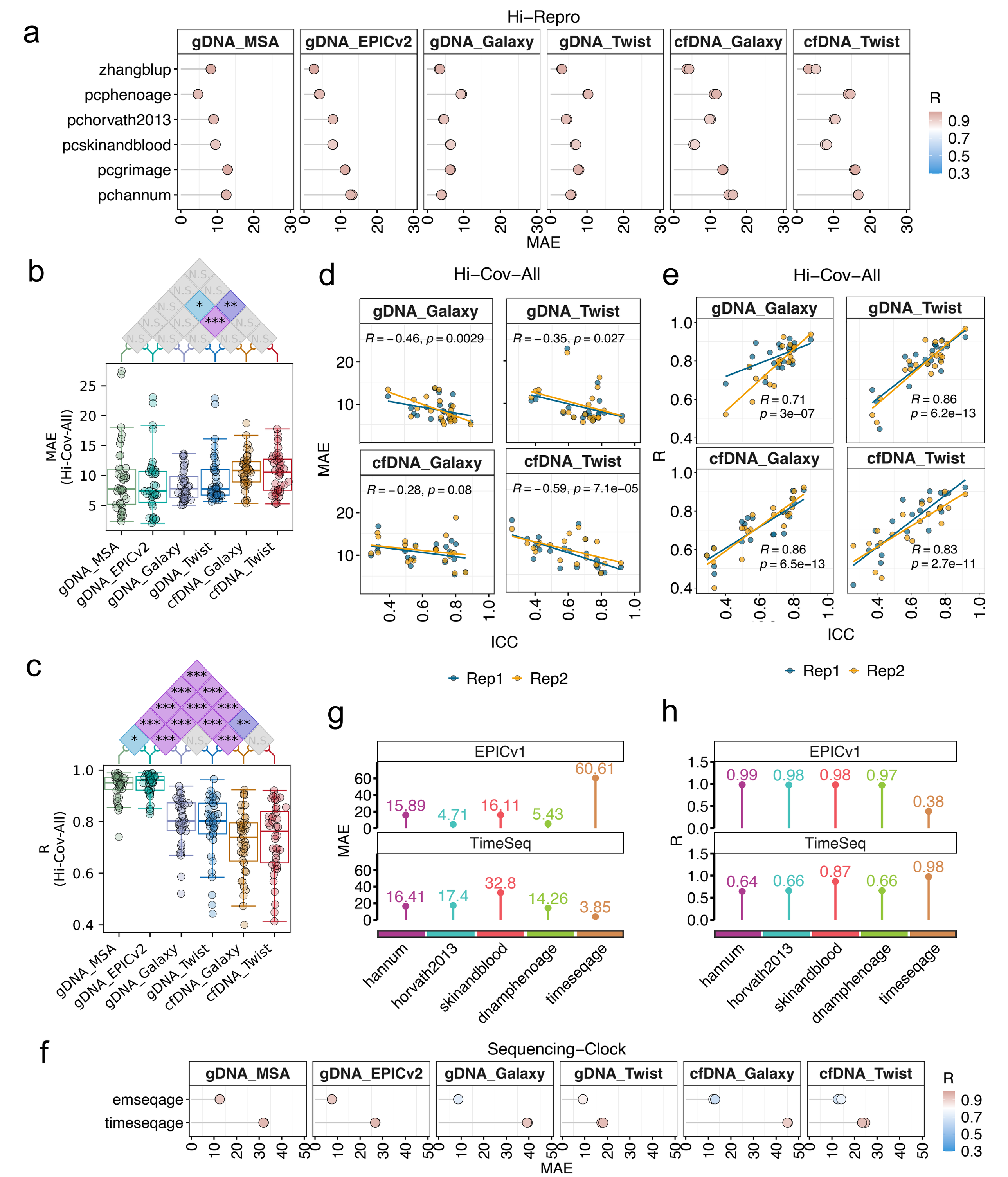
a
Hi−Repro
gDNA_MSA
gDNA_EPICv2
gDNA_Galaxy
gDNA_Twist
cfDNA_Galaxy
cfDNA_Twist
zhangblup
pcphenoage
pchorvath2013
pcskinandblood
pcgrimage
pchannum
MAE
R
b
MAE (Hi-Cov-All)
c
R (Hi-Cov-All)
d
Hi−Cov−All
R = −0.46, p = 0.0029
R = −0.35, p = 0.027
R = −0.28, p = 0.08
R = −0.59, p = 7.1e−05
ICC
e
Hi−Cov−All
R = 0.71
p = 3e−07
R = 0.86
p = 6.5e−13
R = 0.86
p = 6.2e−13
R = 0.83
p = 2.7e−11
Rep1
Rep2
g
EPICv1
TimeSeq
15.89
4.71
16.11
5.43
60.61
16.41
17.4
32.8
14.26
3.85
hannum
horvath2013
skinandblood
dnamphenoage
timeseqage
h
0.99
0.98
0.98
0.97
0.38
0.64
0.66
0.87
0.66
0.98
f
Sequencing−Clock
emseqage
timeseqage

Fig. 2

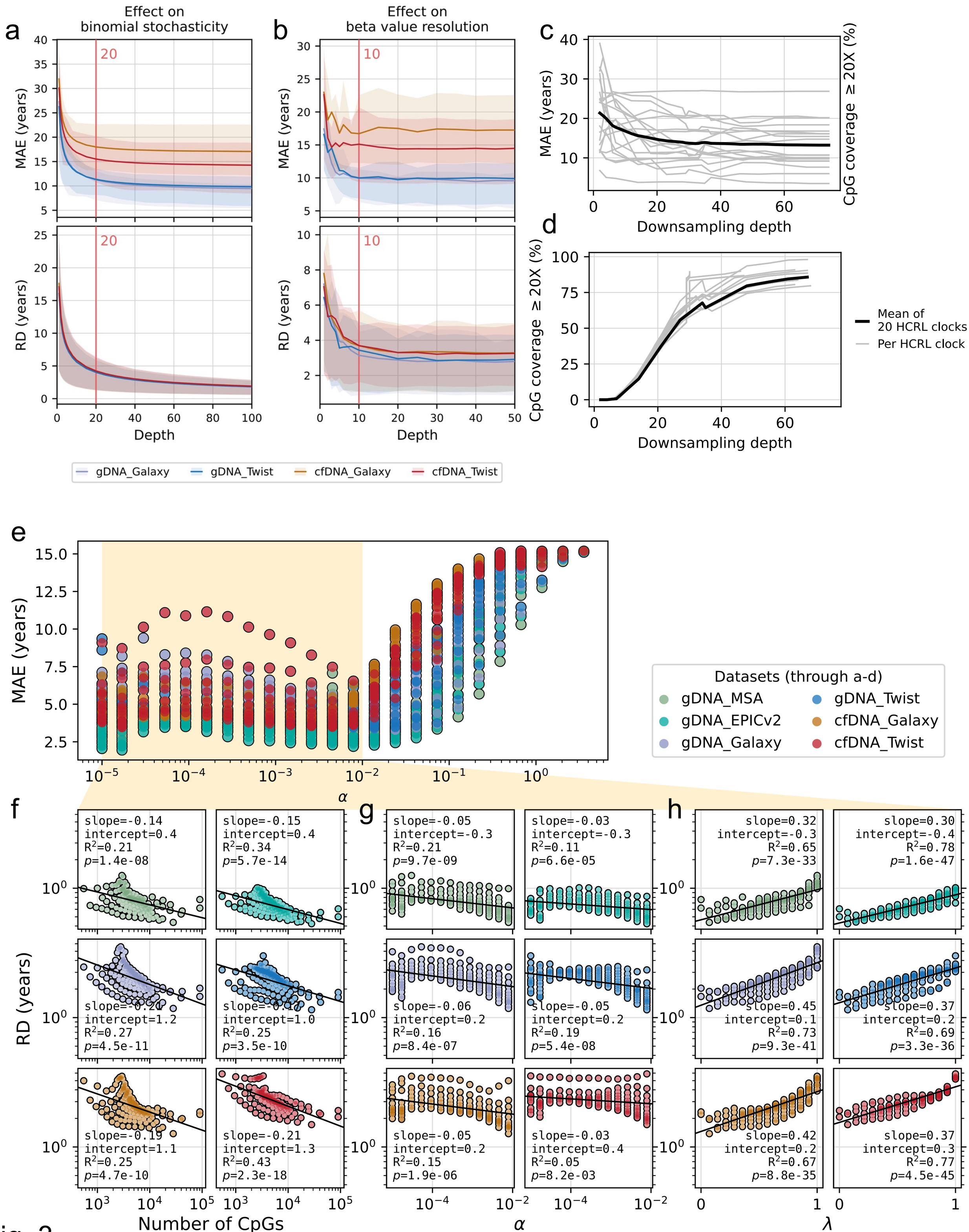
a
Effect on binomial stochasticity
MAE (years)
RD (years)
Depth
b
Effect on beta value resolution
c
MAE (years)
CpG coverage ≥ 20X (%)
Downsampling depth
d
Mean of 20 HCRL clocks
Per HCRL clock
gDNA_Galaxy
gDNA_Twist
cfDNA_Galaxy
cfDNA_Twist
e
MAE (years)
α
Datasets (through a-d)
gDNA_MSA
gDNA_EPICv2
gDNA_Galaxy
gDNA_Twist
cfDNA_Galaxy
cfDNA_Twist
f
g
h
RD (years)
Number of CpGs
α
λ

Fig. 3

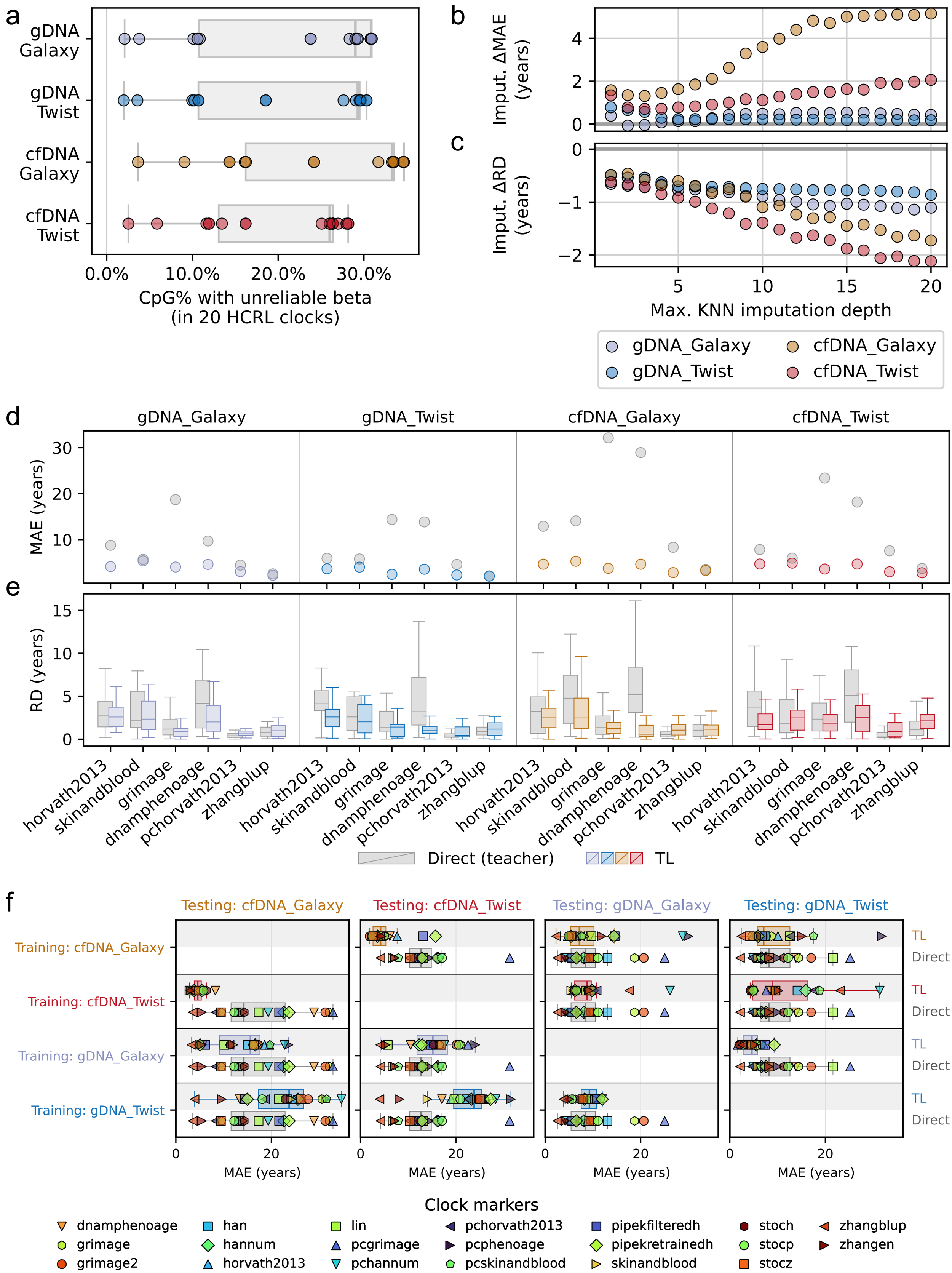
a
gDNA Galaxy
gDNA Twist
cfDNA Galaxy
cfDNA Twist
0.0%
10.0%
20.0%
30.0%
CpG% with unreliable beta (in 20 HCRL clocks)
b
Imput. ΔMAE (years)
c
Imput. ΔRD (years)
Max. KNN imputation depth
gDNA_Galaxy
gDNA_Twist
cfDNA_Galaxy
cfDNA_Twist
d
MAE (years)
e
RD (years)
horvath2013
skinandblood
grimage
dnamphenoage
pchorvath2013
zhangblup
Direct (teacher)
TL
f
Testing: cfDNA_Galaxy
Testing: cfDNA_Twist
Testing: gDNA_Galaxy
Testing: gDNA_Twist
Training: cfDNA_Galaxy
Training: cfDNA_Twist
Training: gDNA_Galaxy
Training: gDNA_Twist
TL
Direct
MAE (years)
Clock markers
dnamphenoage
grimage
grimage2
han
hannum
horvath2013
lin
pcgrimage
pchannum
pchorvath2013
pcphenoage
pcskinandblood
pipekfilteredh
pipekretrainedh
skinandblood
stoch
stocp
stocz
zhangblup
zhangen

Fig. 4

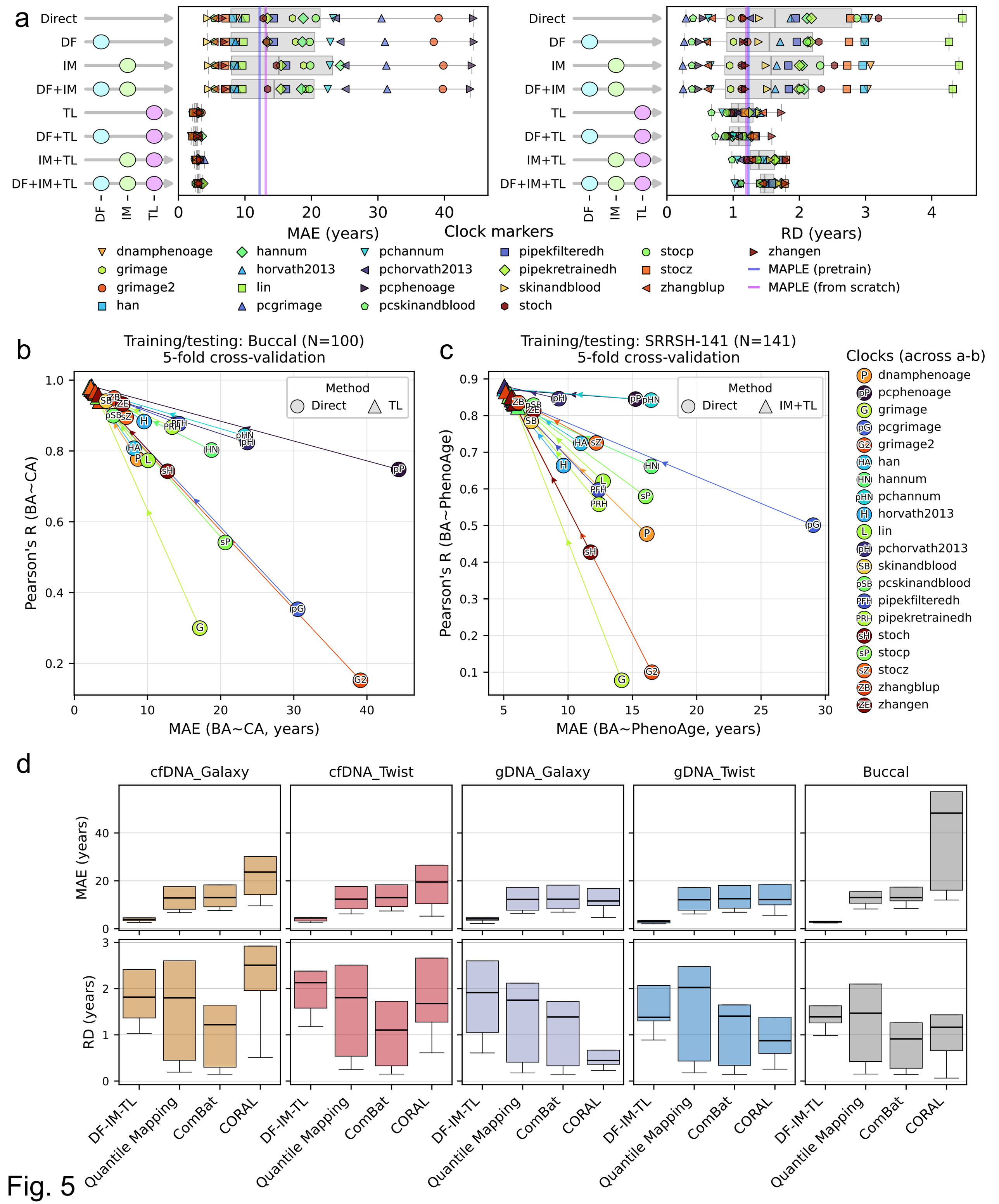
a
Direct
DF
IM
DF+IM
TL
DF+TL
IM+TL
DF+IM+TL
MAE (years)
RD (years)
Clock markers
dnamphenoage
grimage
grimage2
han
hannum
horvath2013
lin
pcgrimage
pchannum
pchorvath2013
pcphenoage
pcskinandblood
pipekfilteredh
pipekretrainedh
skinandblood
stoch
stocp
stocz
zhangblup
zhangen
MAPLE (pretrain)
MAPLE (from scratch)
b
Training/testing: Buccal (N=100)
5-fold cross-validation
Method
Direct
TL
Pearson's R (BA~CA)
MAE (BA~CA, years)
c
Training/testing: SRRSH-141 (N=141)
5-fold cross-validation
Method
Direct
IM+TL
Pearson's R (BA~PhenoAge)
MAE (BA~PhenoAge, years)
Clocks (across a-b)
dnamphenoage
pcphenoage
grimage
pcgrimage
grimage2
han
hannum
pchannum
horvath2013
lin
pchorvath2013
skinandblood
pcskinandblood
pipekfilteredh
pipekretrainedh
stoch
stocp
stocz
zhangblup
zhangen
d
cfDNA_Galaxy
cfDNA_Twist
gDNA_Galaxy
gDNA_Twist
Buccal
MAE (years)
RD (years)
DF-IM-TL
Quantile Mapping
ComBat
CORAL

Fig. 5

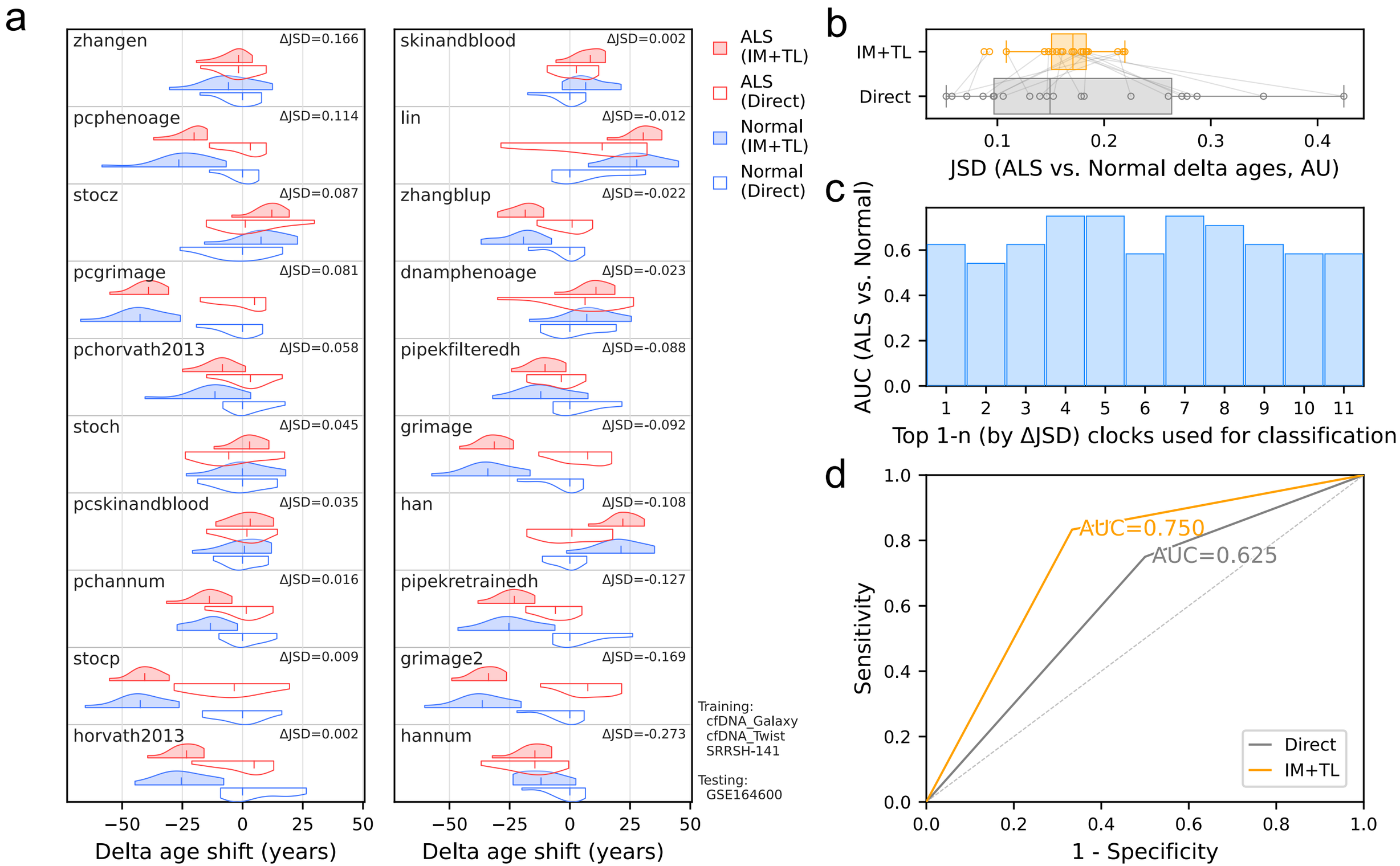
a
zhangen ΔJSD=0.166
pcphenoage ΔJSD=0.114
stocz ΔJSD=0.087
pcgrimage ΔJSD=0.081
pchorvath2013 ΔJSD=0.058
stoch ΔJSD=0.045
pcskinandblood ΔJSD=0.035
pchannum ΔJSD=0.016
stocp ΔJSD=0.009
horvath2013 ΔJSD=0.002
skinandblood ΔJSD=0.002
lin ΔJSD=-0.012
zhangblup ΔJSD=-0.022
dnamphenoage ΔJSD=-0.023
pipekfilteredh ΔJSD=-0.088
grimage ΔJSD=-0.092
han ΔJSD=-0.108
pipekretrainedh ΔJSD=-0.127
grimage2 ΔJSD=-0.169
hannum ΔJSD=-0.273
−50 −25 0 25 50
Delta age shift (years)
ALS (IM+TL)
ALS (Direct)
Normal (IM+TL)
Normal (Direct)
Training:
cfDNA_Galaxy
cfDNA_Twist
SRRSH-141
Testing:
GSE164600
b
IM+TL
Direct
0.1 0.2 0.3 0.4
JSD (ALS vs. Normal delta ages, AU)
c
AUC (ALS vs. Normal)
0.0 0.2 0.4 0.6
1 2 3 4 5 6 7 8 9 10 11
Top 1-n (by ΔJSD) clocks used for classification
d
Sensitivity
1 - Specificity
0.0 0.2 0.4 0.6 0.8 1.0
AUC=0.750
AUC=0.625
Direct
IM+TL

Fig. 6