ClockBase: a comprehensive platform for biological age profiling in human and mouse

Kejun Ying [1,2], Alexander Tyshkovskiy [1], Alexandre Trapp [1], Hanna Liu [1,3], Mahdi Moqri [1,4,5], Csaba Kerepesi [1,6], Vadim N. Gladyshev [1, ✉]

[1] Division of Genetics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA.

[2] T. H. Chan School of Public Health, Harvard University, Boston, MA.

[3] Massachusetts College of Pharmacy and Health Sciences, Boston, MA, USA

[4] Department of Obstetrics & Gynecology, Stanford School of Medicine, Stanford University, Stanford, CA, USA

[5] Institute for Stem Cell Biology and Regenerative Medicine, Stanford School of Medicine, Stanford, CA, USA

[6] Institute for Computer Science and Control (SZTAKI), Budapest, Hungary

✉ email: vgladyshev@rics.bwh.harvard.edu

ABSTRACT

Aging represents the greatest risk factor for chronic diseases and mortality, but to understand it we need the ability to measure biological age. In recent years, many machine learning algorithms based on omics data, termed aging clocks, have been developed that can accurately predict the age of biological samples. However, there is currently no resource for systematic profiling of biological age. Here, we describe *ClockBase*, a platform that features biological age estimates based on multiple aging clock models applied to more than 2,000 DNA methylation datasets and nearly 200,000 samples. We further provide an online interface for statistical analyses and visualization of the data. To show how this resource could facilitate the discovery of biological age-

modifying factors, we describe a novel anti-aging drug candidate, zebularine, which reduces the biological age estimates based on all aging clock models tested. We also show that pulmonary fibrosis accelerates epigenetic age. Together, *ClockBase* provides a resource for the scientific community to quantify and explore biological ages of samples, thus facilitating discovery of new longevity interventions and age-accelerating conditions.

INTRODUCTION

Aging is an extremely complex biological process that represents the greatest risk factor for chronic diseases [1,2]. This makes the aging process a desirable target for preventing age-related diseases and reducing their global burden [3–5]. However, to associate aging with diseases and interventions that target aging, it is important to be able to measure the rate of aging [6,7]. In recent years, various machine-learning models based on omics data have emerged (also known as aging clocks), which can accurately predict the age of samples derived from different tissues, cell types, and even single cells [6,8,9]. Various molecular markers have been shown to have the potential to profile the rate of aging, including DNA methylation, transcriptome, proteome, metabolome, microbiome, and other types of omics data [6]. In addition to assessing chronological age, many aging clock models were trained to reveal associations with various aging-related phenotypes and mortality [10].

Aging clocks based on the methylation levels of CpGs are the earliest and some of the most accurate age predictors [11,12]. Such clocks are represented by the blood biomarker developed by Hannum and colleagues [12] and the human pan-tissue epigenetic clock developed by Horvath, the latter trained on 51 different tissue types [11]. These epigenetic clocks could accurately predict the chronological age of samples, but since they are trained based only on age, only a fraction of the biological variation of the sample could be captured by them. Subsequently, the “second-generation” epigenetic clocks emerged: instead of training solely on chronological age, these clocks incorporated health-related phenotypic information and therefore could reveal a stronger association with aging-related phenotypes. For example, PhenoAge was trained based on the phenotypic age score, which was derived from chronological age and certain mortality-related blood test parameters [13]. Additionally, GrimAge emerged as a robust predictor that is based on multiple phenotypes and the remaining time to death [14]. More recently, DunedinPOAm and

DunedinPACE biomarkers were reported that are trained based on the pace of biological aging that was derived from multiple clinical biomarkers measured in the Dunedin longitudinal cohort [15,16]. We also developed DamAge and AdaptAge, which are causality-informed clock models that could separately measure age-related damage and adaptation [17]. Similar to humans, multiple mouse epigenetic clocks were developed and shown to be able to robustly predict the chronological age of mice [18,19].

One of major applications of aging clocks is the identification of conditions or treatments that modify the aging rate of individuals or reduce their biological age, which could potentially lead to the development of anti-aging therapies [20,21]. For example, parabiosis and iPSC reprogramming were shown to be associated with the decrease in epigenetic age [22–24], and unhealthy lifestyles such as smoking and stress could accelerate epigenetic aging [10,25]. Although this type of research has been described in many publications, there has been no systematic effort to identify the impact of available interventions on biological age. One reason is that various clock models require different transformations and pre-processing of omics data, making it difficult to use them and compare them across studies. Some tools exist that try to tackle this problem, but they typically only utilize a small subset of human methylation clocks [26]. Moreover, although a large amount of omics data has been acquired by the scientific community that is publicly available through the databases such as Gene Expression Omnibus (GEO) [27], there is currently no public resource that could uniformly process them for biological age profiling.

To address this problem, we created *ClockBase*, a comprehensive platform for biological age profiling in humans and mice. We curated 11 best-performing aging clock models, including epigenetic clocks for humans and mice, and used them to profile the biological age of samples (Figure 1). We re-processed over 2,000 publicly available DNA methylation datasets from GEO. In total, *ClockBase* contains the biological age information for around 200,000 samples in both mice and humans under various experimental settings. Besides preprocessed data, users can upload their data to *ClockBase* for biological age calculation. *ClockBase* provides an interactive analysis tool to allow users to perform statistical analyses and visualization of biological age online. We believe that *ClockBase* may provide a valuable resource for the scientific community to explore the biological age of samples, and thus facilitate the discovery of new longevity interventions and age-accelerating conditions.

Results

Overview of *ClockBase*

To develop *ClockBase*, we processed over 2,000 publicly available DNA methylation datasets from GEO and calculated biological age based on multiple aging clocks (Figure 1). In total, *ClockBase* contains biological age information for ~200,000 human and mouse samples (Figure 1). We standardized metadata for each experiment, which allows users to search for diseases and treatments of interest and examine biological age under a variety of experimental conditions. All data are available for download. Besides preprocessed data, users can upload their data to *ClockBase* and calculate predicted biological age.

ClockBase provides an interactive analysis tool that allows users to perform statistical analyses and visualization of biological age online. We also embedded each sample into a low-dimensional space which allows users to explore data interactively. Our toolkit includes group comparison, which allows users to compare biological age across different experimental groups; correlation analysis, which allows users to explore the relationship between biological age and other numeric variables, or the correlation across different clock models; and accuracy analysis, which allows users to explore the accuracy of clock models. All plots and statistical results are available for download. We also created a companion R package called ClockBasis, that allows users to calculate biological age of their samples. *ClockBase* is available at https://clockbase.org

ClockBase offers insights into the relationship among clock models

To understand the biological meaning and relationship among aging clocks, it is important to compare different clocks and have information on their correlation. Although several studies reported on this topic, all were performed with established human cohorts and biobanks [35,36], which contain only a limited number of interventions and biological variables. As ClockBase consists of a large number of samples with highly diverse biological statuses, it provides a unique opportunity for exploring the relationship among different clock models in a much more diverse sample population. We first explored the distribution of biological age measurement across 192,635 highly diverse human samples (Figure 2a). Among them, 80,346 samples also had age information. We, therefore, calculated biological age acceleration for the samples based on each clock (delta age, which is calculated as predicted age minus real age). Note that DunedinPoAm and DunedinPACE are predictors of the pace of age that is independent of the age of samples and is

centered at 1. We then examined the distribution of biological age acceleration across the samples (Figure 2b). chi-square test was performed to determine whether there are significantly more samples with accelerated biological age or decelerated biological age. Interestingly, while DunedinPACE, HannumAge, HorvathAge, and ZhangAge clocks showed that there are significantly more age-accelerated samples, DunedinPoAm and PhenoAge revealed the opposite effect (i.e. there are significantly more age-decelerated samples), whereas PedBE clock showed no significant difference. This suggests that different clocks may measure different aspects of aging and therefore have a disagreement on the biological age of samples.

We further analyzed correlation across biological age prediction based on seven aging clock models (Figure 2c, d). Prior to adjusting for age, PedBE, Horvath Clock, Zhang clock, Hannum Clock, and PhenoAge showed strong correlation with one another, with Pearson's correlation coefficients ranging from 0.59 (PhenoAge and PedBE) to 0.85 (Zhang clock and Hannum Clock). Correlations between DunedinPoAm/DunedinPACE and other clocks were low. This is expected as both DunedinPoAm and DunedinPACE measure the rate of aging, which shows only a weak correlation with chronological age [16]. Yet surprisingly, Pearson's correlation coefficient between DunedinPoAm and DunedinPACE was -0.05.

After adjusting for age, the five epigenetic age clocks (Horvath Clock, Zhang clock, Hannum Clock, PedBE, and PhenoAge) still showed a significant, yet weaker, positive correlation (Figure 2d). Pearson's correlation coefficients ranged from 0.31 (HorvathAge and PhenoAge) to 0.89 (ZhangAge and PedPE). DunedinPoAm and DunedinPACE still showed a weak correlation with all other clocks. Notably, DunedinPACE has a significant negative correlation with all other clocks except HorvathAge. These findings reveal the internal discrepancy among different aging clocks when applied to diverse biological samples.

To better visualize inconsistency among different aging clocks, we embedded each sample into two-dimensional space by performing UMAP on biological age predictions from each clock model (Figure 3a, b). Locations of the samples on UMAP embedding indicated the relationship among biological age prediction for different aging clocks.

As a demonstration, we show that although DunedinPACE has a very weak correlation with other aging clock models, it predicts iPSCs and ESCs to have extremely slow rates of aging, which

is related to other clock models that revealed consistently low ages of these cells following long-term maintenance in culture. Therefore, iPSCs/ESCs form a unique cluster in the UMAP space. Similarly, cells overexpressing DNA methyltransferases (DNMTs) are predicted to be relatively young based on Horvath Clock and PhenoAge, and also have a very slow rate of aging based on DunedinPACE [37]. In contrast, during induced differentiation *in vitro*, hepatocytes appear to be more than 200 years old based on Horvath Clock and PhenoAge and also exhibit an extremely fast rate of aging [38].

In general, samples form a trajectory in the UMAP space, where the upper left corner represents biologically older samples and the lower right and lower left corners younger samples (Figure 3a). The branching of the trajectory indicates disagreement among different aging clocks. For example, the lower left branch has low biological age prediction based on HannumAge, HorvathAge, PhenoAge, and ZhangAge clocks. Yet PedBE shows a moderate biological age prediction, and DunedinPoAm shows that this region contains samples with an accelerated pace of aging. The discrepancy becomes even more obvious when we used biological age acceleration (delta age) as the attribute for t-SNE embedding (Figure 3c). The interactive three-dimensional UMAP and t-SNE embedding are available in the ClockBase online analysis tool.

ClockBase facilitates the discovery of longevity interventions and age-accelerating conditions

To demonstrate the utility of *ClockBase* for identifying novel longevity interventions and age-accelerating conditions, we show two datasets that to our knowledge have not been studied in the context of biological aging. In the first dataset (GSE60446), two different cholangiocarcinoma cell types, TFK-1 and HuCCT1, were treated with a DNA methyltransferase inhibitor zebularine (1-(β-D-ribofuranosyl)-1,2-dihydropyridine-2-one) [39]. Through only a few clicks on the *ClockBase* online statistical analysis tool, we found that the zebularine treatment significantly reduces the epigenetic age based on almost all clock models and in both cell lines (Figure 4a). In addition, both DunedinPoAm and DunedinPACE showed that the zebularine-treated cells exhibited a slower pace of aging. Zebularine has never been studied for its role in rejuvenation, and our results suggest that this compound is a potential longevity intervention, which may be further studied in future studies.

The second example is from GSE63704, which includes 204 plasma DNA methylation samples representing healthy controls and lung cancer, pulmonary fibrosis, and chronic obstructive pulmonary disease (COPD) patients [40]. We observed that pulmonary fibrosis patients exhibit a significantly higher epigenetic age compared to control patients, based on Horvath Clock, PedBE, Zhang clock, and Hannum Clock models (Figure 4b). Additionally, DunedinPACE showed that pulmonary fibrosis patients had a significantly faster pace of aging. Notably, both of these examples were semi-randomly selected for demonstration purposes, suggesting that there are many other potential associations that remain to be explored by future *ClockBase* users.

Discussion

The emergence of aging clocks provided researchers with promising tools to estimate the age of biological samples and shed light on the associated biology. However, there are currently multiple dozens of aging clocks that have been created, making it increasingly important to understand the relationship between different aging clocks [8,41–43]. There have been some efforts to compare clocks based on established human cohorts and biobanks [35,36], but these studies are limited in both clocks examined and the dataset used. *ClockBase* currently contains DNA methylation for both mice and humans, with much more diverse sample coverage compared to human biobanks. We believe that this resource can be used to help researchers to understand the relationship between clocks in different experimental settings.

Another challenge is that it is currently hard for non-computational experts in the field to use aging clocks, as they usually require different transformations and data preprocessing. Even for computational biologists, downloading individual datasets from GEO and preprocessing each of them is a time-consuming task. *ClockBase* is designed to provide a simple and easy-to-use interface for biologists to perform statistical analyses and visualization of biological age. Only a GSE accession identifier and a few clicks are required for analyzing a dataset from GEO. This could remove the barrier for researchers and domain experts to use and understand the aging clocks.

We illustrated the utility of ClockBase by discovering zebularine, a potent DNMT inhibitor, which affects the methylation status of the samples by directly targeting the DNA methylation machinery [44]. Our data suggest that zebularine is a candidate longevity intervention, as it significantly reduced the epigenetic age of cultured cells based on almost all clock models. However,

as zebularine affects the DNA methylation machinery, DNA methylation clocks should be used with caution. Further investigation and *in vivo* studies are required to understand the role of zebularine in the aging process.

We believe many other potential anti-aging interventions are hidden in a large number of available experimental conditions, that could be explored and explained by domain experts.

MATERIALS AND METHODS

Data collection

The data used in this study were downloaded before July 30th, 2022, from Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo). Raw data were downloaded using the R package GEOquery (https://bioconductor.org/packages/release/bioc/html/GEOquery.html), and metadata were extracted using the R package GEOmetadb (https://www.bioconductor.org/packages/release/bioc/html/GEOmetadb.html). For mouse DNA methylation analyses, all GEO entries associated with "Methylation profiling by high throughput sequencing" were collected; for human DNA methylation analyses, all GEO entries associated with "Methylation profiling by (genome tiling, SNP, or other) array" were collected. Methylation level data were then downloaded from supplementary files of each GEO entry. Only the datasets with at least 6 samples were used for downstream analyses.

Methylation data preprocessing

Existing mouse DNA methylation data are not uniformly structured. A custom R script was used to identify CpG sites and methylation levels of each sample and then standardize data format. Metadata are standardized based on the custom pipeline aspired by refine.bio [28]. Datasets with missing information or in unrecognized format were excluded. Then, for both mouse and human DNA methylation data, the range of methylation levels was standardized to the 0-1 scale. The data with out-of-range values were replaced with missing values. We impute missing methylation level data using mean methylation for the reference dataset. For humans, we used 2,664 blood samples measured using the 450k Human Methylation Beadchip as a reference [29]. For mice, since sequencing-based methods were used, DNA methylation data were more sparse compared to array-based data. Therefore, we first imputed missing values based on mean methylation levels within 100 base-pair regions, as it was reported in a previous study that the nearby

sites tend to exhibit a high correlation with regard to methylation levels [30]. For the sites still having missing values, we imputed missing values based on the mean methylation levels of the reference dataset from Petkovich et al. [18]. We report the ratio of missingness for each clock model. In general, samples with more than 20% missing values were considered unreliable for biological age prediction; however, we included them in the database with a warning message as they may still provide information.

The code for all the preprocessing steps is included in the ClockBasis R package (https://github.com/albert-ying/ClockBasis).

Aging clock models implementation

Aging clock models were implemented on the web server and precalculated for all datasets, including 4 mouse epigenetic clocks and 7 human epigenetic clocks. The following mouse epigenetic clocks were included: Petkovich blood clock (90 sites) [18], Meer multi-tissue clock (435 sites) [19], Thompson multi-tissue clock (582 sites) [31], and Wang liver clock (148 sites) [32]. The following human epigenetic clocks were included: Horvath multi-tissue clock (353 sites) [11], Hannum clock (71 sites) [12], PhenoAge (513 sites) [13], PedBE pediatric buccal clock (94 sites) [33], Zhang blood clock (514 sites) [34], DunedinPOAm (46 sites) [15], and DunedinPACE (173 sites) [16].

All clock models are publicly available and could be downloaded from the original source. All epigenetic clocks are also available as functions in the ClockBasis R package.

Online statistical analysis

Three types of statistical analysis were implemented in the *ClockBase* online interface.

(1) Group comparison: the group comparison function allows users to compare the biological age or another numeric variable across different experimental groups in the dataset. The pairwise T-test is performed across each group and p-value is adjusted by the number of comparisons using the Benjamini-Hochberg procedure. p-value for ANOVA across all groups is also reported. The result is an output in the form of a boxplot followed by the result table.

(2) Correlation: the correlation function allows users to calculate Pearson's correlation across two numeric variables in the dataset. The result is an output in the form of a scatter plot with regression lines. Pearson's correlation coefficient and p-value are also reported. Users can further calculate correlations within each subgroup of the dataset and report statistics separately. Notably, this function is also useful for quality control by visualizing correlation between biological age prediction and percentage missingness of the data. This

could avoid reporting false positive results due to imbalanced missingness across experimental groups.

(3) Accuracy: the accuracy function allows users to calculate accuracy of biological age prediction when the true age is given in the dataset. Pearson's R, RMSE, MAE, and p-value are reported.

Data availability

All data are available on the *ClockBase* online resource (https://clockbase.org) and GEO (https://www.ncbi.nlm.nih.gov/geo/).

Code availability

All codes are available in the ClockBasis R package (https://github.com/albert-ying/ClockBasis)

Acknowledgments

We thank all the Gladyshev lab members for their helpful discussions and suggestions. We especially thank Miss. Ying Fang for her help in designing the *ClockBase* logo and webpage. Supported by NIA grants to VNG.

Author contributions

K.Y. and A.T. initiated the study; K.Y. collected the data; V.N.G supervised the study. K.Y., A.T., and H.L. performed data analyses; All authors contributed to paper preparation and data interpretation.

Competing interest statement

The authors declare no competing financial interests.

Figure legends

Figure 1. *ClockBase* overview. Schematic diagram shows data sources and main functionalities of the *ClockBase* online resource.

Figure 2. Distribution and correlation of biological age in human samples. a. Density plot shows the distribution of biological age measurements across 192,635 samples based on 7 human DNAm aging clocks. **b.** Density plot shows the distribution of biological age acceleration across 80,346 samples with chronological age annotation. Dashed line shows the boundary between samples with accelerated and decelerated ages. chi-square test was performed, and the ratio and p-value are shown above the plots **c,d.** Correlation plot shows Pearson's correlation across different biological age measurements (**c**) and biological age acceleration (**d**). Upper triangle: Pearson's correlation plot; lower triangle: Pearson's correlation coefficient. Areas of the squares represent the absolute value of corresponding Pearson's correlations. P values are corrected using Bonferroni correction for 21 tests with $P_{adjusted} < 0.05$.

Figure 3. *ClockBase* reveals discrepancy across different biological age measurements. a, b. The UMAP plot of 192,635 human DNA methylation samples. Colors of the dots represent biological age prediction based on HorvathAge (**a**) and other human DNAm clocks (**b**). **c.** t-SNE plot of 80,346 human DNA methylation samples with chronological age annotation. Colors of the dots represent biological acceleration based on each clock.

Figure 4. Identifying novel biological age-modifying conditions using *ClockBase*. a. Epigenetic age comparison between zebularine-treated and untreated cells. **b.** Epigenetic age comparison across healthy control, lung cancer, pulmonary fibrosis (fibrosis), and COPD patients. Boxes indicate 25−75% interquartile ranges, and whiskers indicate minimum to maximum. * $P_{adjusted} < 0.05$, ** $P_{adjusted} < 0.01$, *** $P_{adjusted} < 0.001$, **** $P_{adjusted} < 0.0001$. COPD: chronic obstructive pulmonary disease.

References

1. Chang, A. Y., Skirbekk, V. F., Tyrovolas, S., Kassebaum, N. J. & Dieleman, J. L. Measuring population ageing: an analysis of the Global Burden of Disease Study 2017. *Lancet Public Health* **4**, e159–e167 (2019).
2. Chauhan, A. *et al.* Systems biology approaches in aging research. *Interdiscip Top Gerontol* **40**, 155–76 (2015).
3. Bakula, D. *et al.* Aging and drug discovery. *Aging* **10**, 3079–3088 (2018).
4. Gladyshev, T. V. & Gladyshev, V. N. A Disease or Not a Disease? Aging As a Pathology. *Trends Mol Med* **22**, 995–996 (2016).
5. Barzilai, N., Cuervo, A. M. & Austad, S. Aging as a Biological Target for Prevention and Therapy. *JAMA* **320**, 1321–1322 (2018).
6. Rutledge, J., Oh, H. & Wyss-Coray, T. Measuring biological age using omics data. *Nat. Rev. Genet.* (2022) doi:10.1038/s41576-022-00511-7.
7. Ludwig, F. C. & Smoke, M. E. The measurement of biological age. *Exp. Aging Res.* **6**, 497–522 (1980).
8. Macdonald-Dunlop, E. *et al.* A catalogue of omics biological ageing clocks reveals substantial commonality and associations with disease risk. *Aging* **14**, 623–659 (2022).
9. Trapp, A., Kerepesi, C. & Gladyshev, V. N. Profiling epigenetic age in single cells. *Nat. Aging* **1**, 1189–1201 (2021).
10. Quach, A. *et al.* Epigenetic clock analysis of diet, exercise, education, and lifestyle factors. *Aging* **9**, 419–446 (2017).
11. Horvath, S. DNA methylation age of human tissues and cell types. *Genome Biol.* **14**, R115 (2013).
12. Hannum, G. *et al.* Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol. Cell* **49**, 359–367 (2013).
13. Levine, M. E. *et al.* An epigenetic biomarker of aging for lifespan and healthspan. *Aging* **10**, 573–591 (2018).
14. Lu, A. T. *et al.* DNA methylation GrimAge strongly predicts lifespan and healthspan. *Aging* **11**, 303–327 (2019).
15. Belsky, D. W. *et al.* Quantification of the pace of biological aging in humans through a blood test, the DunedinPoAm DNA methylation algorithm. *eLife* **9**, e54870 (2020).
16. Belsky, D. W. *et al.* DunedinPACE, a DNA methylation biomarker of the pace of aging. *eLife* **11**, e73420 (2022).
17. Ying, K. *et al.* Causal Epigenetic Age Uncouples Damage and Adaptation. 2022.10.07.511382 Preprint at https://doi.org/10.1101/2022.10.07.511382 (2022).
18. Petkovich, D. A. *et al.* Using DNA methylation profiling to evaluate biological age and longevity interventions. *Cell Metab.* **25**, 954-960.e6 (2017).
19. Meer, M. V., Podolskiy, D. I., Tyshkovskiy, A. & Gladyshev, V. N. A whole lifespan mouse multi-tissue DNA methylation clock. *eLife* **7**, e40675 (2018).
20. Galkin, F., Zhang, B., Dmitriev, S. E. & Gladyshev, V. N. Reversibility of irreversible aging. *Ageing Res Rev* **49**, 104–114 (2019).
21. Fahy, G. M. *et al.* Reversal of epigenetic aging and immunosenescent trends in humans. *Aging Cell* **18**, e13028 (2019).
22. Lu, Y. *et al.* Reprogramming to recover youthful epigenetic information and restore vision. *Nature* **588**, 124–129 (2020).
23. Alle, Q. *et al.* A single short reprogramming early in life improves fitness and increases lifespan in old age. *bioRxiv* 2021.05.13.443979 (2021) doi:10.1101/2021.05.13.443979.

24. Maherali, N. *et al.* Directly Reprogrammed Fibroblasts Show Global Epigenetic Remodeling and Widespread Tissue Contribution. *Cell Stem Cell* **1**, 55–70 (2007).

25. Poganik, J. R. *et al.* Biological age is increased by stress and restored upon recovery. 2022.05.04.490686 Preprint at https://doi.org/10.1101/2022.05.04.490686 (2022).

26. Di Lena, P., Sala, C. & Nardini, C. Estimage: a webserver hub for the computation of methylation age. *Nucleic Acids Res.* **49**, W199–W206 (2021).

27. Edgar, R., Domrachev, M. & Lash, A. E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* **30**, 207–210 (2002).

28. Casey S. Greene, Hu, D., Jones, R. W. W. & Stephanie Liu, David S. Mejia, Rob Patro, Stephen R. Piccolo, Ariel Rodriguez Romero, Hirak Sarkar, Candace L. Savonen, Jaclyn N. Taroni, William E. Vauclain, Deepashree Venkatesh Prasad, Kurt G. refine.bio: a resource of uniformly processed publicly available gene expression datasets.

29. 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).

30. Zhang, W., Spector, T. D., Deloukas, P., Bell, J. T. & Engelhardt, B. E. Predicting genome-wide DNA methylation using methylation marks, genomic position, and DNA regulatory elements. *Genome Biol.* **16**, 14 (2015).

31. Thompson, M. J. *et al.* A multi-tissue full lifespan epigenetic clock for mice. *Aging* **10**, 2832–2854 (2018).

32. Wang, T. *et al.* Epigenetic aging signatures in mice livers are slowed by dwarfism, calorie restriction and rapamycin treatment. *Genome Biol.* **18**, 57 (2017).

33. McEwen, L. M. *et al.* The PedBE clock accurately estimates DNA methylation age in pediatric buccal cells. *Proc. Natl. Acad. Sci.* **117**, 23329–23335 (2020).

34. Zhang, Y. *et al.* Neuronal TORC1 modulates longevity via AMPK and cell nonautonomous regulation of mitochondrial dynamics in C. elegans. *Elife* **8**, (2019).

35. Li, X. *et al.* Longitudinal trajectories, correlations and mortality associations of nine biological ages across 20-years follow-up. *Elife* **9**, (2020).

36. Crimmins, E. M., Thyagarajan, B., Levine, M. E., Weir, D. R. & Faul, J. Associations of Age, Sex, Race/Ethnicity, and Education With 13 Epigenetic Clocks in a Nationally Representative U.S. Sample: The Health and Retirement Study. *J. Gerontol. Ser. A* **76**, 1117–1123 (2021).

37. Kobayashi, Y. *et al.* DNA methylation profiling reveals novel biomarkers and important roles for DNA methyltransferases in prostate cancer. *Genome Res.* **21**, 1017–1027 (2011).

38. Ancey, P.-B. *et al.* TET-Catalyzed 5-Hydroxymethylation Precedes HNF4A Promoter Choice during Differentiation of Bipotent Liver Progenitors. *Stem Cell Rep.* **9**, 264–278 (2017).

39. Nakamura, K. *et al.* DNA methyltransferase inhibitor zebularine induces human cholangiocarcinoma cell death through alteration of DNA methylation status. *PloS One* **10**, e0120545 (2015).

40. Wielscher, M. *et al.* Diagnostic Performance of Plasma DNA Methylation Profiles in Lung Cancer, Pulmonary Fibrosis and COPD. *EBioMedicine* **2**, 929–936 (2015).

41. Levine, M. E., Higgins-Chen, A., Thrush, K., Minteer, C. & Niimi, P. Clock Work: Deconstructing the Epigenetic Clock Signals in Aging, Disease, and Reprogramming. 2022.02.13.480245 Preprint at https://doi.org/10.1101/2022.02.13.480245 (2022).

42. Bell, C. G. *et al.* DNA methylation aging clocks: challenges and recommendations. *Genome Biol* **20**, 249 (2019).

43. Horvath, S. & Raj, K. DNA methylation-based biomarkers and the epigenetic clock theory of ageing. *Nat Rev Genet* **19**, 371–384 (2018).

44. Yoo, C. B., Cheng, J. C. & Jones, P. A. Zebularine: a new drug for epigenetic therapy. *Biochem. Soc. Trans.* **32**, 910–912 (2004).

Figures

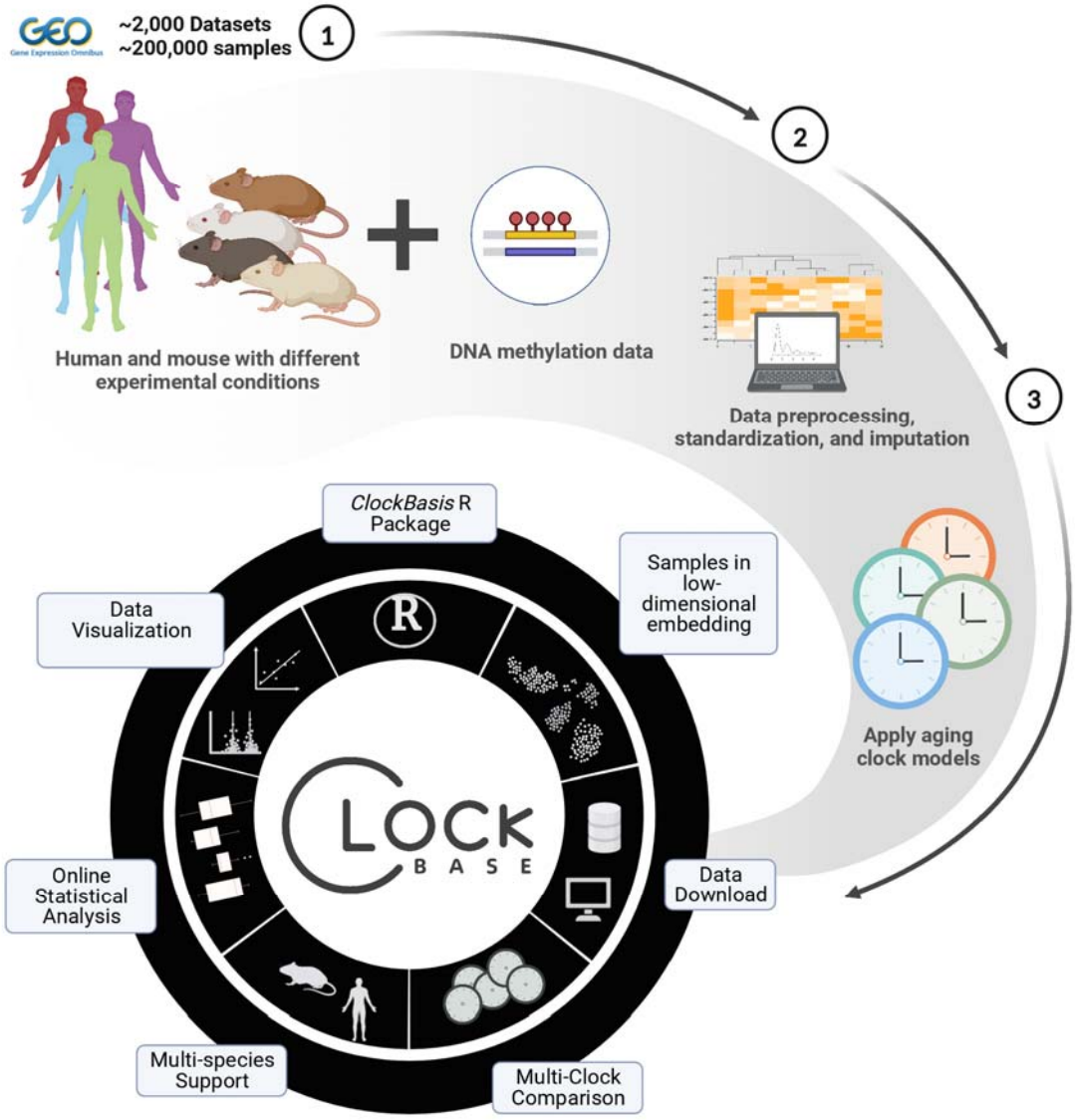

Figure 1

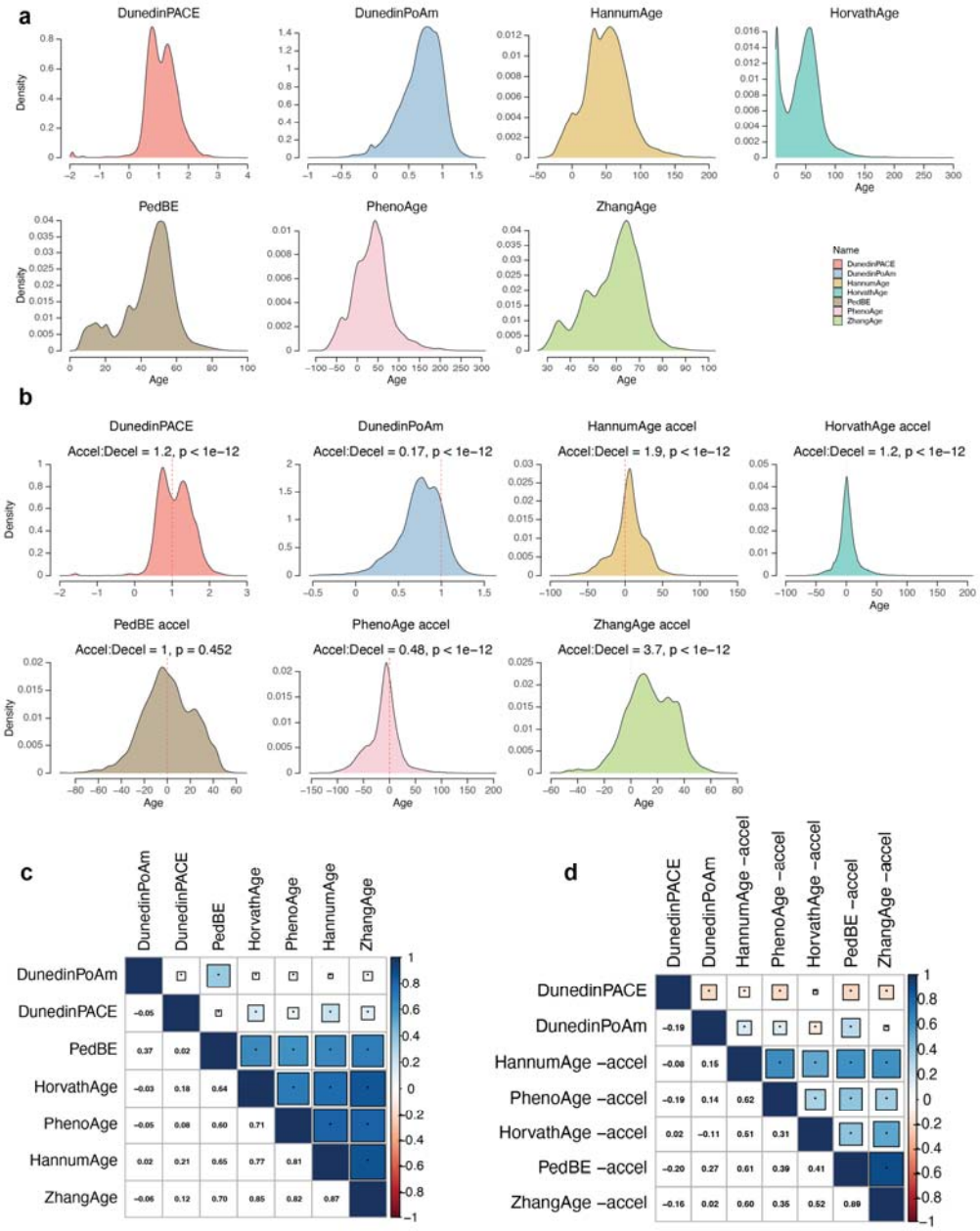

Figure 2

383

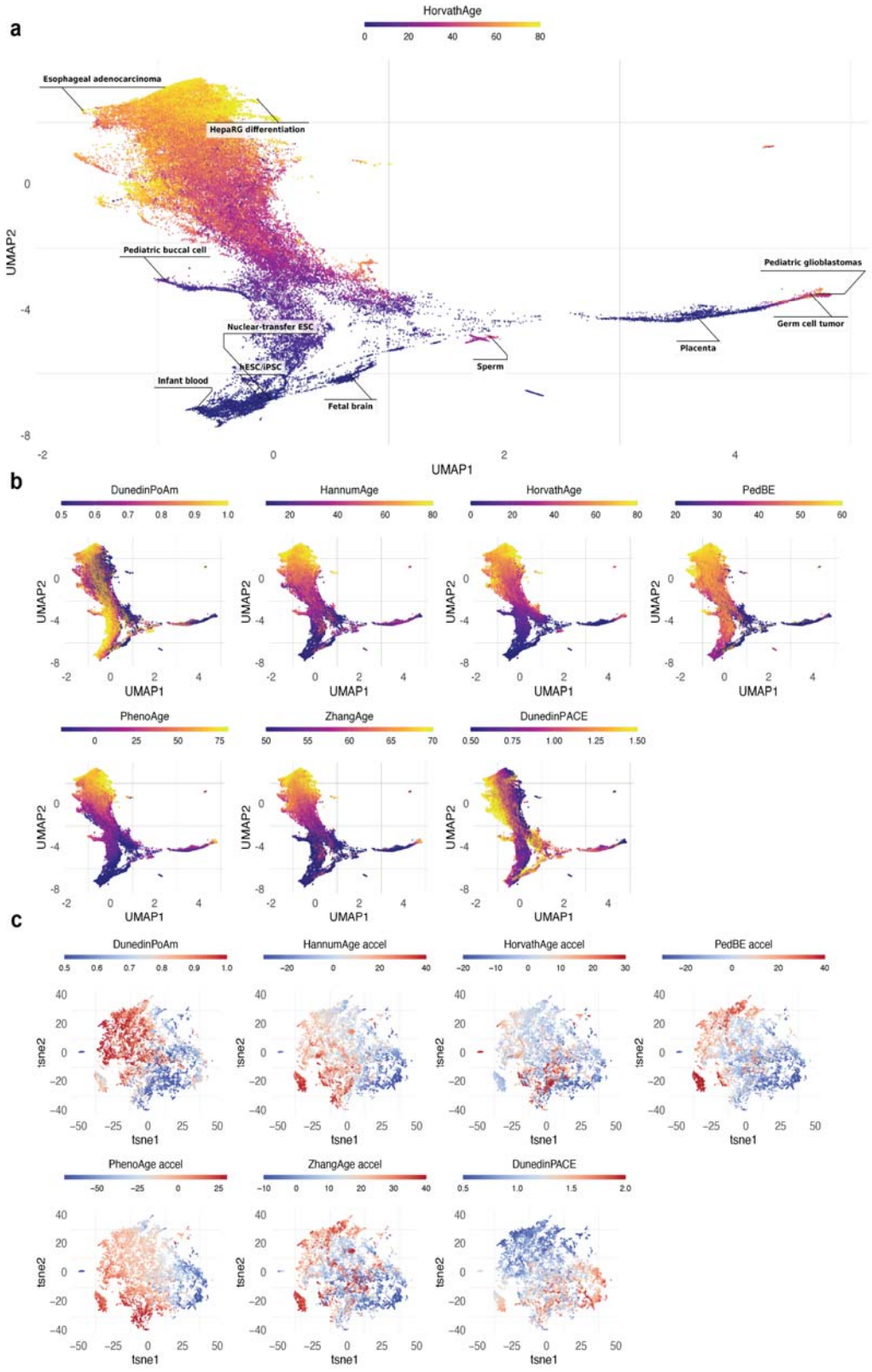

Figure 3

384

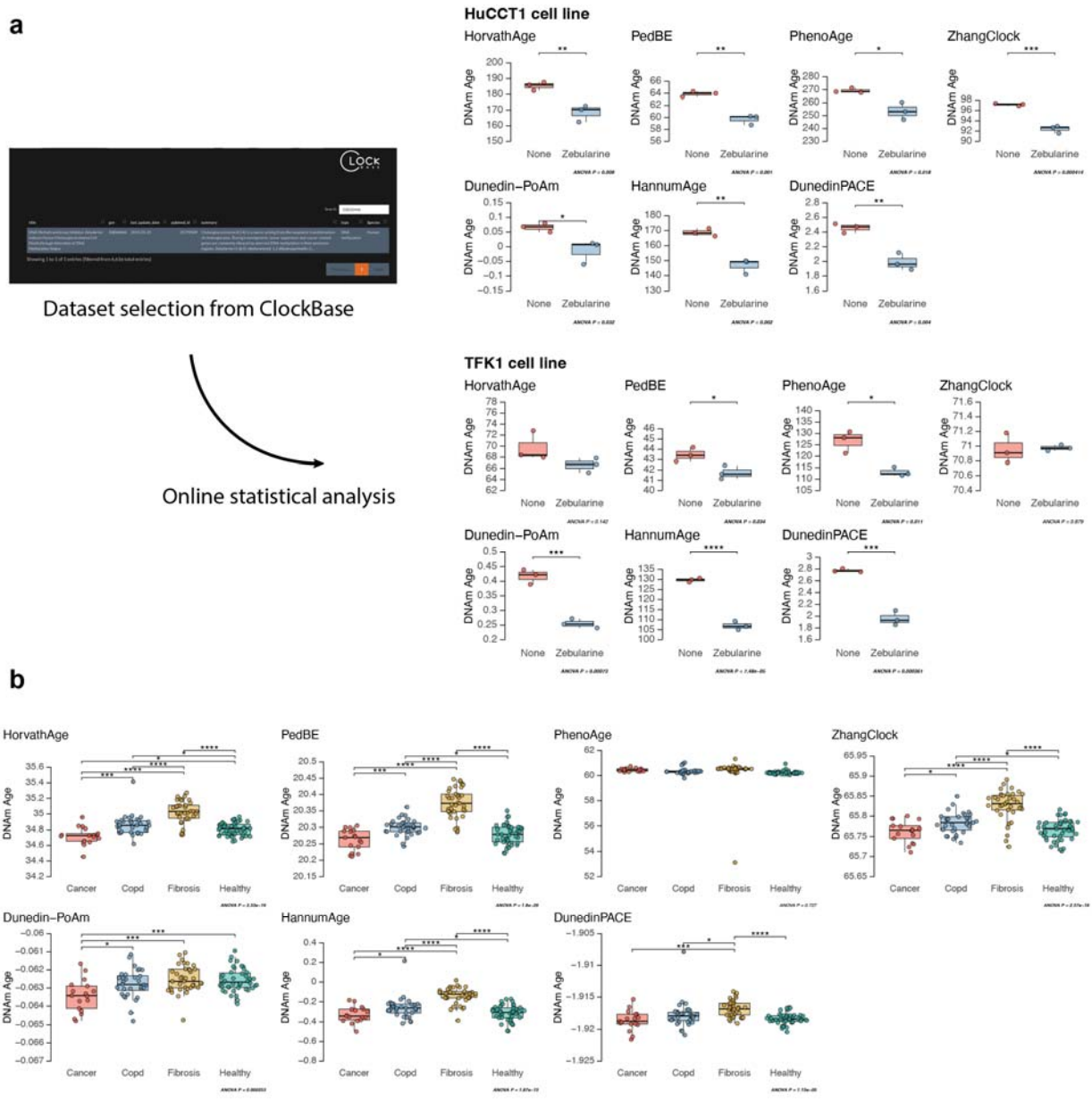

Figure 4

385