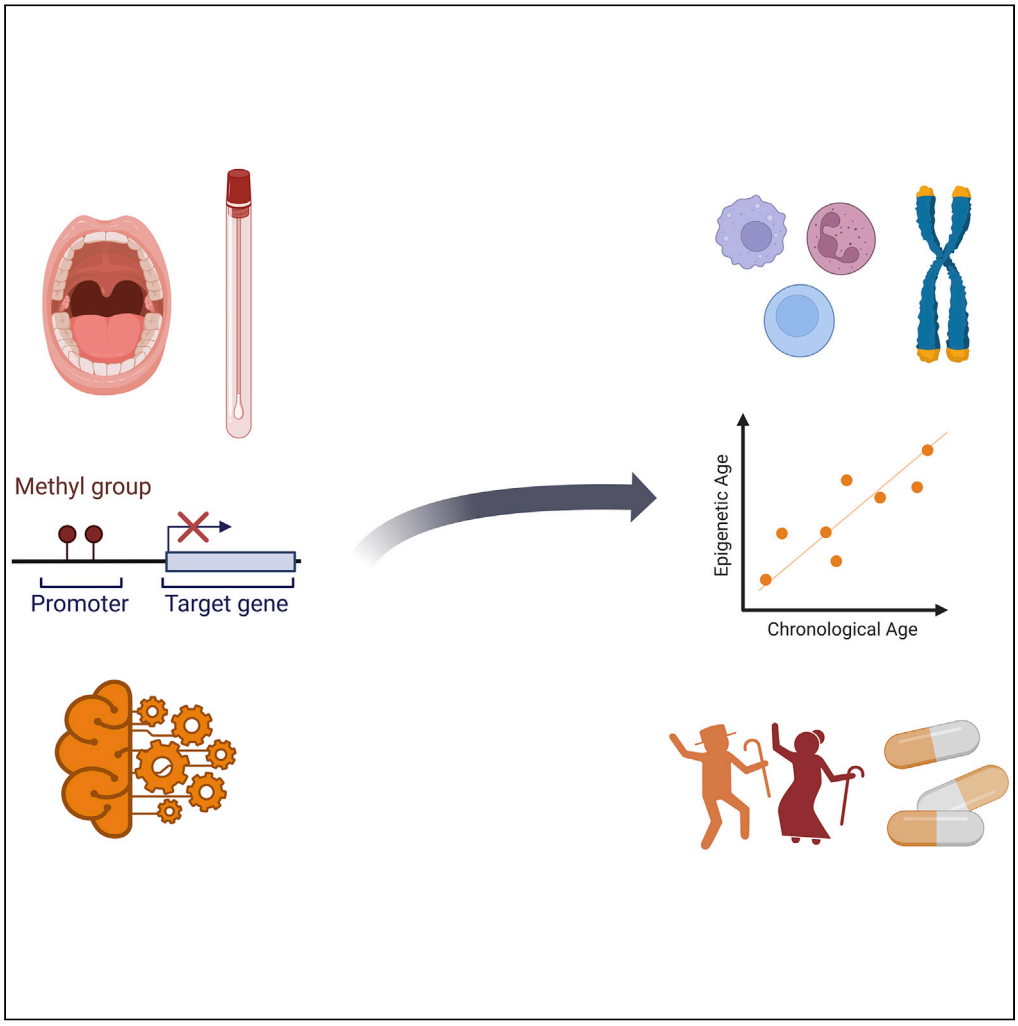
iScience

CellPress
OPEN ACCESS

Article

A set of common buccal CpGs that predict epigenetic age and associate with lifespan-regulating genes



Adiv A. Johnson, Nicole S. Torosin, Maxim N. Shokhirev, Trinna L. Cuellar

adiv@tallyhealth.com

Highlights

130 CpGs have been used by two or more aging clocks applied to human buccal tissue

Common CpG genes are linked to the adaptive immune system and telomere maintenance

Common CpGs can be used to build a novel, proof-of-concept epigenetic aging clock

Several compounds associated with common CpG genes regulate lifespan in animals

Johnson et al., iScience 25, 105304
November 18, 2022
https://doi.org/10.1016/j.isci.2022.105304

iScience

CellPress
OPEN ACCESS

Article

A set of common buccal CpGs that predict epigenetic age and associate with lifespan-regulating genes

Adiv A. Johnson,[1,2,*] Nicole S. Torosin,[1] Maxim N. Shokhirev,[1] and Trinna L. Cuellar[1]

SUMMARY

Epigenetic aging clocks are computational models that use DNA methylation sites to predict age. Since cheek swabs are non-invasive and painless, collecting DNA from buccal tissue is highly desirable. Here, we review 11 existing clocks that have been applied to buccal tissue. Two of these were exclusively trained on adults and, while moderately accurate, have not been used to capture health-relevant differences in epigenetic age. Using 130 common CpGs utilized by two or more existing buccal clocks, we generate a proof-of-concept predictor in an adult methylomic dataset. In addition to accurately estimating age (r = 0.95 and mean absolute error = 3.88 years), this clock predicted that Down syndrome subjects were significantly older relative to controls. A literature and database review of CpG-associated genes identified numerous genes (e.g., *CLOCK*, *ELOVL2*, and *VGF*) and molecules (e.g., alpha-linolenic acid, glycine, and spermidine) reported to influence lifespan and/or age-related disease in model organisms.

INTRODUCTION

A rich body of literature points to various biomarkers of aging that help distinguish between individuals of the same chronological age and provide insights into how an individual is aging relative to a larger population. For example, a recent analysis involving 405,981 participants in the UK BioBank found that leukocyte telomere length positively correlated with self-reported walking pace (Dempsey et al., 2022). A separate illustration is the protein GDF15, which dramatically increases with age in human plasma (Lehallier et al., 2019). In a 20-year prospective study, levels of circulating GDF15 were measured in 4,143 subjects without cardiovascular disease and found to be predictive of all-cause mortality (Bao et al., 2021). Classical clinical markers can also be mined to provide insights into age-related health outcomes. For instance, elevated levels of high-density lipoprotein-related and glycemic markers were respectively associated with protective and deleterious mortality effects in 12,098 individuals (Li et al., 2021). Non-molecular biomarkers also have predictive utility, such as polypharmacy (Chen et al., 2021), grip strength (Pavasini et al., 2019), the timed up and go test (Chun et al., 2021), and perceived age (Uotinen et al., 2005).

One rich resource for aging biomarkers is the methylome, which becomes substantially remodeled with age (Johnson et al., 2012). Various research groups have used machine learning models to unbiasedly identify CpG sites whose methylation status can be quantified and combined into a predicted age score. Exemplar of this, Castle et al. utilized the elastic net regression algorithm to identify 286 DNA methylation sites to predict age in breast tissue with a Pearson correlation of 0.88 and a median absolute error of 4.2 years. Using this model, the authors found that epigenetic age was ~seven years higher than chronological age in tissue samples from breast tumors (Castle et al., 2020). This deviation between epigenetic age and chronological age—termed Δage—can correlate with various age-related health outcomes, including morbidity and mortality (Bell et al., 2019). Early-generation clocks—such as the Horvath 2013 (Horvath, 2013) and Hannum 2013 (Hannum et al., 2013) predictors—were trained to estimate chronological age. By incorporating a more complex training regimen, subsequent methylomic models were generated that were especially adept at predicting mortality and health outcomes (Belsky et al., 2022; Levine et al., 2018; Lu et al., 2019). In the case of GrimAge, a composite biomarker was created based on epigenetic estimators of plasma proteins and smoking pack-years (Lu et al., 2019). A separate next-generation clock is DNAm PhenoAge, which was trained to predict age calculated by distinct clinical markers (Levine et al.,

[1]Longevity Sciences, Inc. (dba Tally Health), Greenwich, CT, USA

[2]Lead contact

*Correspondence: adiv@tallyhealth.com

https://doi.org/10.1016/j.isci.2022.105304

2018). Another contemporary model called DunedinPACE was generated using information about longitudinal variation in organ-system integrity (Belsky et al., 2022).

In addition to their diagnostic utility, aging clocks can be used as research tools for drug discovery. For example, Janssens et al. created a transcriptomic aging clock and tested the ability of 1,309 different compounds to influence predicted age in human cells. This approach identified a list of candidate drugs, a subset of which was selected for *in vivo* testing in nematodes. The authors showed that the candidate compounds valproic acid, LY-294002, rapamycin, monorden, and tanespimycin all extend lifespan in *Caenorhabditis elegans* (Janssens et al., 2019). In a separate study, an accelerometer-based model called MoveAge was applied in a large human cohort to demonstrate that the antihypertensive drug doxazosin associates with age deceleration. When given to worms, doxazosin enhanced longevity and late-life motility (McIntyre et al., 2021). A separate, *in vitro* epigenetic predictor called CellAgeClock was created to estimate age in human primary cells. In addition to confirming the efficacy of known aging interventions (e.g., rapamycin), the candidate drugs torin2 and Dactolisib lowered epigenetic age *in vitro*. When these compounds were given to flies, they extended lifespan relative to controls (Lujan et al., 2020). Various interventions, including the adoption of a Mediterranean diet, have also been reported to decrease predicted age in humans (Johnson et al., 2022).

The majority of epigenetic aging clocks were trained exclusively in blood or in multi-tissue datasets containing blood samples (Simpson and Chandra, 2021). Although comparatively less common, more recent work demonstrates that buccal tissue is also a viable sample type for age prediction in pediatric (McEwen et al., 2020) and adult populations (Jung et al., 2019). Given the appeal of cheek swabs as a painless and non-invasive collection method, we wanted to better understand the buccal epigenetic clock landscape and compare the efficacy of disparate models. In doing so, we identify a set of 130 common CpGs utilized by two or more different aging clocks (Figure 1A). Given that most of these CpGs were identified using an unbiased machine learning model, this degree of overlap was unexpected. Using both bioinformatics and machine learning, we use these CpGs to reveal age-related enrichment processes (Figure 1B) and generate a proof-of-concept age prediction model (Figure 1C). Finally, we explore existing databases and literature (Figure 1D) to suggest that a subset of clock CpGs as well as their associated genes represent molecules of interest pertinent to longevity and age-related disease.

RESULTS

A common set of epigenetic clock CpGs

To begin, we performed a literature search to categorize methylomic models that have been used to predict age in human buccal tissue. In total, we identified 11 distinct epigenetic aging clocks (Table 1). The number of CpG inputs used varied from 1 to 513 and the Pearson correlation ranged from 0.83 to 0.98. Seven of these clocks were quite minimalistic and used 20 CpGs or less. Error metrics were reported as either mean absolute error or median absolute error. The lowest mean and median absolute errors were respectively 4.14 and 0.35 years. Conversely, the highest mean absolute error was 7.8 years and the largest median absolute error was 6.9 years (Table 1). Both the Eipel 2016 (Eipel et al., 2016) and McEwen 2020 (McEwen et al., 2020) models employed cell-type correction. The best performing model was developed by McEwen et al., which was trained on 1,721 pediatric buccal samples spanning an age range of 0–20 years (McEwen et al., 2020). Of the buccal models we identified (Table 1), only two exclusively focused on cheek swabs derived from adults, namely the minimalistic Jung 2019 (Jung et al., 2019) and Schwender 2021 (Schwender et al., 2021) clocks. Involving a sample set spanning an age range of 18–74 years, the Jung 2019 clock had a Pearson correlation of 0.93 and a mean absolute error of 4.29 years. The sample age range for the Schwender 2021 clock was narrower (21–69 years) and had a Pearson correlation of 0.88 and a mean absolute error of 5.33 years.

For each model, we list the CpG inputs as well as any associated genes and accession IDs in Table S1. Taking an approach inspired by our previous proteomics work (Johnson et al., 2020), we compared all 11 CpG lists to see if any of the inputs overlapped between different clocks (Figure 2). This comparison revealed a total of 130 common CpGs (Table S2) that were used by two or more distinct age predictors. Although 105 of these were utilized by two different clocks and 21 emerged in three distinct models, only the following four CpGs showed up four or more times: cg16867657 (*ELOVL2*), cg06144905 (*PIPOX*), cg09809672 (*EDARADD*), and cg17861230 (*PDE4C*). The DNA methylation sites cg16867657 (*ELOVL2*) and cg06144905 (*PIPOX*) were shared by four clocks, cg09809672 (*EDARADD*) occurred five times, and cg17861230

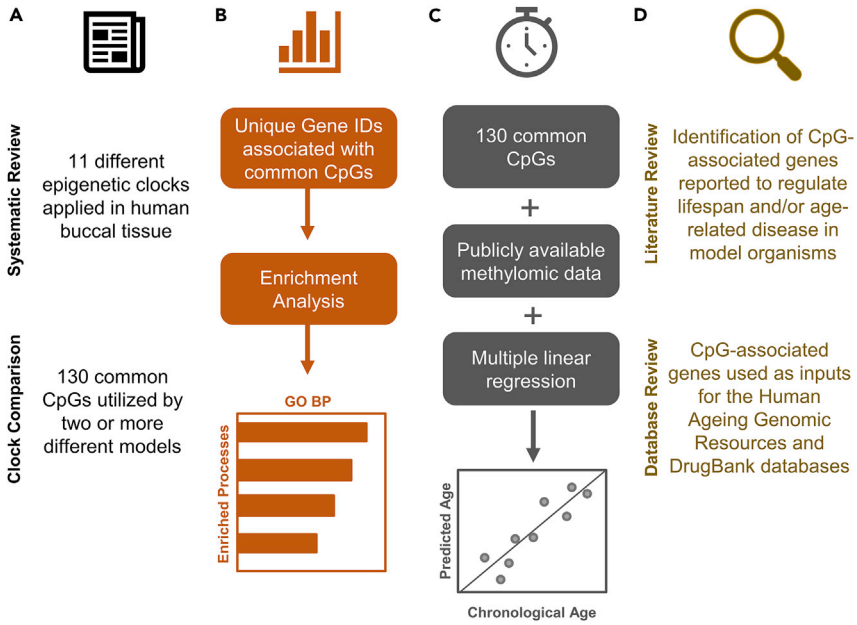


Figure 1. Analytical workflow

(A) To begin, we performed a survey of the literature to identify 11 different epigenetic aging clocks that have been used to predict age in human buccal tissue. By comparing the inputs of each model, we uncovered a set of 130 common CpGs used by two or more different clocks.

(B) We utilized the unique gene IDs associated with the common CpGs to perform a network topology-based analysis and identify significantly enriched processes in the Gene Ontology Biological Process (GO BP) database.

(C) Using publicly available methylomic data in the Gene Expression Omnibus database, we used multiple linear regression in conjunction with the common CpGs to generate a proof-of-concept clock that can accurately predict age in adult data.

(D) A literature and database analysis revealed that many of the genes connected to common CpGs have been reported to influence lifespan and/or age-related disease in model organisms. Several molecules associated with common CpG genes are also implicated in the DrugBank database, some of which have been previously reported to elongate lifespan in animals.

(*PDE4C*) was utilized by six different models. The appearance of the *ELOVL2*-associated CpG cg16867657 is not surprising given that this DNA methylation site has been reported to undergo hypermethylation with age across multiple tissue types (Slieker et al., 2018).

The majority of common CpGs were independently prioritized via elastic net regression in different methylomic datasets (Figure 2 and Table S2). One possibility for the degree of overlap is that elastic net is selecting CpGs that work well in combination for the task of epigenetic age prediction. While the Horvath 2013 clock was trained using CpGs shared between Illumina's Infinium HumanMethylation27 and Infinium HumanMethylation450 arrays, the Horvath 2018 and McEwen 2020 clocks used CpGs present on both the Infinium HumanMethylation450 and Infinium MethylationEPIC arrays. The Levine 2018 clock focused on CpGs shared by all three array platforms. Of the 130 common buccal CpGs, 128 were utilized by at least two of these unbiased predictors. The directionality (i.e., hypermethylated or hypomethylated) of these 128 CpGs was largely consistent, with 90 (70.31%) DNA methylation sites being reported to change in the same direction across different studies (Table S2). Among these 90 consistent CpGs, 39 (43.33%) had a positive coefficient (i.e., hypermethylated with age) and 51 (56.67%) had a negative coefficient (i.e., hypomethylated with age). We hypothesize that the consistency of directionality was not higher due to the fact that these models have been trained on different tissue sets. Both the Horvath 2013 and Horvath 2018 clocks were built using methylation information from multiple tissue sources, including buccal tissue. The McEwen 2020 model was exclusively trained using buccal samples and the Levine 2018 predictor was initially trained on whole blood.

As an example of a CpG with consistent directionality, cg05442902 (*P2RX6*) exhibits age-dependent hypomethylation according to the Horvath 2013, Horvath 2018, and Levine 2018 clocks (Table S2). Interestingly,

Table 1. Metrics for 11 epigenetic clocks that have been used to predict age in human buccal tissue

Sample Size	Age Range (Years)	# of CpG Inputs	Correlation	Error (Years)	Reference
221	0–68	353	R = 0.83	Median error = 0.37	(Horvath, 2013)
55	1–85	3	R = 0.96	Mean error = 7.03	(Eipel et al., 2016)
91	6–73	2	R = 0.86	Mean error = 7.1	(Alghanim et al., 2017)
485	1–60	391	R = 0.88	Median error = 2	(Horvath et al., 2018)
Not reported	Not reported	513	R = 0.88	Not reported	(Levine et al., 2018)
148	18–74	5	R = 0.93	Mean error = 4.29	(Jung et al., 2019)
95	0–80	9	R = 0.87	Median error = 6.9	(Han et al., 2020)
1721	0–20	94	R = 0.98	Median error = 0.35	(McEwen et al., 2020)
142	0–89	1	R = 0.92	Mean error = 7.8	(Koop et al., 2021)
141	21–69	3	R = 0.88	Mean error = 5.33	(Schwender et al., 2021)
370	10–65	20	Not reported	Mean error = 4.14 or 4.38	(Becker et al., 2022)

protein levels of P2RX6 significantly trend toward increased expression with age (q = 0.0019 and coefficient = 0.00044) in human plasma (Lehallier et al., 2019). Although most changes in DNA methylation do not correlate with gene expression (Murphy et al., 2013), a subset of age-related methylomic alterations strongly correspond with age-related changes in gene expression (Chatsirisupachai et al., 2021). Multi-omics analyses are warranted to determine if the methylation status of these common CpG sites influences gene and protein expression. Suggesting that such efforts are worthwhile, various CpG-related genes were previously identified to change with age according to our prior systematic review of human proteomic studies (Johnson et al., 2021), our previous aging meta-analysis of human transcriptomic studies (Shokhirev and Johnson, 2021), and the Digital Aging Atlas database (Craig et al., 2015). The overlap between each CpG-associated gene and these resources is provided in Table S2. It is also worth noting that the common CpG cg13547237 is linked with the global transcription repressor *DRAP1* (Kim et al., 1997).

Enrichment analysis implicates telomere maintenance and the adaptive immune system

We were curious to learn what biological processes were associated with these unique CpGs. To do this, we performed a network topology-based analysis (Wang et al., 2017) using WebGestalt (Liao et al., 2019) and filtered for enrichment results with a false discovery rate (FDR) < 0.05 in the Gene Ontology Biological Process database (The Gene Ontology, 2019). Unlike a typical over-representation enrichment analysis, a network topology-based enrichment analysis expands the number of inputs by identifying and including interacting partners. This was done in the BioGRID database (Oughtred et al., 2021), which stores curated information about protein, genetic, and chemical interactions.

The 14 significantly enriched results (Figure 3 and Table S3) highlight distinct themes of telomere maintenance ("telomere maintenance via telomerase", "RNA-dependent DNA biosynthetic process", and "telomere maintenance via telomere lengthening") and the adaptive immune system ("T cell costimulation" and "lymphocyte costimulation"). Several results were pertinent to stimulus-induced responses ("cellular response to stimulus", "response to abiotic stimulus", and "response to stimulus"), cell signaling ("cell communication", "signaling", and "cell-cell signaling"), and regulation ("regulation of biological quality", "negative regulation of multicellular organismal process", and "regulation of DNA-binding transcription factor activity"). Three processes were tied for having the lowest FDR value, namely "cellular response to stimulus", "cell communication", and "regulation of biological quality". The telomere maintenance and adaptive immune system themes respectively implicate the canonical aging hallmarks telomere attrition and altered intercellular communication (Lopez-Otin et al., 2013). Interestingly, telomeres were also incriminated in a recent transcriptomic comparison between immortal and mortal jellyfish (Pascual-Torner et al., 2022).

Collating publicly available data to build a proof-of-concept aging clock

Looking at the 11 clocks listed in Table 1, two were built using multi-tissue datasets containing buccal tissue (Horvath, 2013; Horvath et al., 2018) and one was initially trained on whole blood and subsequently applied to buccal samples (Levine et al., 2018). Of the remaining eight, one focused exclusively on pediatric subjects (McEwen et al., 2020), five included pediatric and adult samples (Alghanim et al., 2017; Becker

	Horvath 2013	Eipel 2016	Alghanim 2017	Horvath 2018	Levine 2018	Jung 2019	Han 2020	McEwen 2020	Koop 2021	Schwender 2021	Becker 2022
Horvath 2013	353	1	0	60	41	0	0	1	0	1	1
Eipel 2016	1	3	0	3	2	0	0	0	1	1	1
Alghanim 2017	0	0	2	0	0	0	0	0	0	0	0
Horvath 2018	60	3	0	391	58	0	2	12	1	2	4
Levine 2018	41	2	0	58	513	0	0	1	1	2	2
Jung 2019	0	0	0	0	0	5	0	0	0	0	0
Han 2020	0	0	0	2	0	0	9	1	0	0	1
McEwen 2020	1	0	0	12	1	0	1	94	0	0	1
Koop 2021	0	1	0	1	1	0	0	0	1	1	1
Schwender 2021	1	1	0	2	2	0	0	0	1	3	2
Becker 2022	1	1	0	4	2	0	1	1	1	2	20

Figure 2. A set of common clock CpGs

Across 11 different epigenetic aging clocks used to predict age in human buccal tissue, a total of 130 common buccal CpGs were identified that were utilized by two or more models. The bulk of overlap was seen between the Horvath 2013, Horvath 2018, Levine 2018, and McEwen 2020 models. Lighter colors represent less overlap and darker colors represent more overlap. Gray boxes indicate the number of CpG inputs used by each clock.

et al., 2022; Eipel et al., 2016; Han et al., 2020; Koop et al., 2021), and two focused exclusively on adult samples (Jung et al., 2019; Schwender et al., 2021). The adult-specific buccal clocks exhibited moderate accuracy and their ability to detect epigenetic age acceleration in association with specific conditions or behaviors has not been tested. As such, the field is currently lacking validated models optimized for prediction in adult buccal tissue.

Given that these common buccal CpGs were useful for multiple clocks, we hypothesized that they could be used to accurately predict epigenetic age. To test this, we first searched for adult buccal methylomic data in the Gene Expression Omnibus database (Clough and Barrett, 2016). Three HumanMethylation450 adult buccal datasets with age information were identified (Table S4) containing a total of 390 samples, 300 of which were not associated with a disease, condition, or tobacco-related behavior. Of these 300 samples, 255 were derived from females, 45 originated from males, and the age range was 25–60 years. A reported race or ethnicity was not listed in association with 260 samples. In order to expand the age range and sample number, we incorporated 131 controls from a saliva dataset (Table S4). A principal component analysis (PCA) shows that the combined methylation data exhibit distinct variance based on batch (Figure S1A), which is to be expected given our inclusion of four different datasets. These 431 samples had an age range of 25–88 years and included 312 females and 119 males. For the buccal and saliva datasets that reported race or ethnicity, 65.5% were Caucasian, 27.5% were Hispanic, and 7% were African American. We justified the merging of these datasets given that cellular content is similar between buccal tissue and saliva (Theda et al., 2018). Furthermore, Jung et al. previously showed that an epigenetic predictor can be built which predicts age in both tissue types (Jung et al., 2019). Lastly, various models have been constructed that can measure epigenetic age in both saliva and cheek swab samples (Horvath, 2013; Horvath et al., 2018; Levine et al., 2018).

Building and assessing the accuracy of a proof-of-concept aging clock

Using multiple linear regression, all 130 common CpGs, estimated cell type proportions (Figure S1B) as additional features, and 10-fold cross-validation to repeatedly split samples into training and validation sets, we find that age can be predicted with a Pearson correlation of 0.95 and a mean absolute error of 3.88 years (Figure 4A). While these accuracy metrics stand out as being relatively high (Table 1), we were curious if other clocks could be designed that outperform our common CpG clock. As such, we used elastic net regression to create a second proof-of-concept clock trained on the top 10,000 CpGs correlated with age along with the predicted cell type proportions as features. The training automatically selected 335 CpGs optimal for age prediction, three of which—cg11896923 (*PNMT*), cg01820374 (*LAG3*), and cg22809047 (*RPL31*)—overlapped with our common buccal CpGs. This minimal amount of overlap may be due to our dataset only including adult samples as well as the pre-filtering for the top 10,000 age-correlated CpGs prior to running the elastic net model. The unbiased clock exhibited comparable accuracy metrics, namely a Pearson correlation of 0.94 and a mean absolute error of 3.85 years (Figure 4B). For both models, removing the saliva dataset markedly impacted accuracy metrics (Figure S2). Given that the

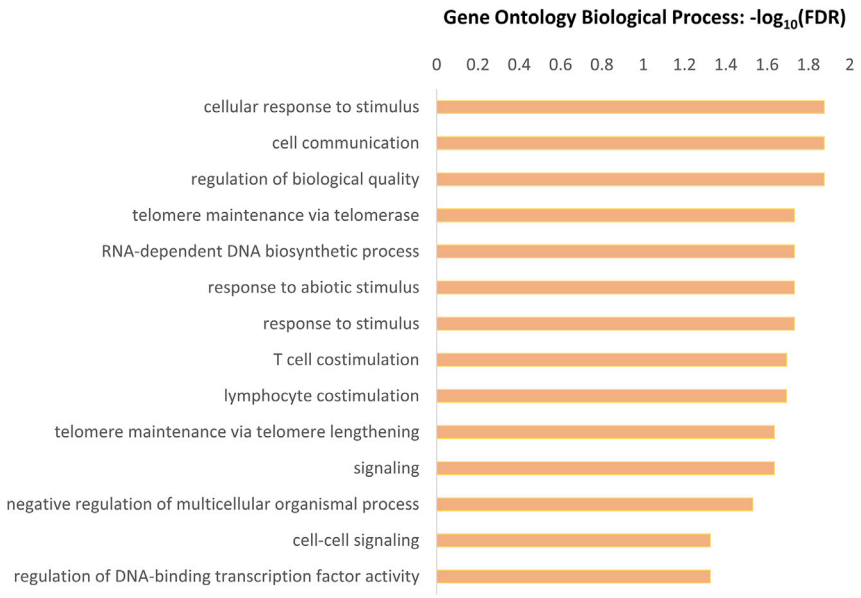


Figure 3. Common CpGs highlight telomere maintenance and the adaptive immune system

Using WebGestalt, a network topology-based analysis was performed on unique gene IDs associated with common buccal clock CpGs. All 14 significantly enriched results in the Gene Ontology Biological Process database are presented as -$\log_{10}$(FDR). The threshold for significance was a false discovery rate (FDR) less than 0.05.

gene *ELOVL2* is known to be hypermethylated with age (Slieker et al., 2018), we created a third clock that uses multiple linear regression and the two common CpGs (cg16867657 and cg21572722) associated with *ELOVL2*. While not as accurate, this minimalistic model is commendable with a Pearson correlation of 0.88 and a mean absolute error of 5.22 years (Figure S3). CpG lists as well as their associated coefficients, genes, and accession IDs for all three clocks are provided in Table S5. An enrichment analysis for the elastic net clock revealed prominent themes of programmed cell death and proteostasis (Figure S4 and Table S6). Predicted age color-coded by estimated cell type is visualized in Figure S5.

The above-mentioned models included estimated cell types as separate features. However, it is also possible to use predicted cell type information to modify coefficients associated with CpGs. To determine which was optimal, we tried the latter approach for both the common CpG (Figure S6A) and elastic net (Figure S6B) clocks. While the common CpG accuracy metrics were comparable (Pearson = 0.94 and mean absolute error = 3.84 years), the elastic net clock displayed a noticeable dip in performance with a Pearson correlation of 0.87 and a mean absolute error of 5.73 years. Regardless of the correction method used, the majority of cells were predicted to be epithelial cells or neutrophils. A small proportion was labeled as other immune cells, which include B cells, monocytes, eosinophils, CD8 T cells, and CD4 T cells.

While our proof-of-concept clocks (Figure 4) display predictive power, they do not outcompete the pediatric buccal clock developed by McEwen et al. (Table 1). We posit that the McEwen 2020 clock is exceptionally accurate for three reasons. First, it was trained on a narrow pediatric age range and therefore able to filter for CpGs that uniquely undergo changes between 0 and 20 years of age. In contrast, the adult age range is broader and expected to involve greater variability. Second, a large number of samples were utilized to train (n = 1,032) and validate (n = 689) the McEwen 2020 clock. Given the dearth of publicly available adult buccal data, our sample size was limited in this study. Third, McEwen et al. only analyzed cheek swabs and were therefore able to identify the most predictive inputs for buccal tissue. In our study, it was necessary to combine buccal data with saliva data given the limited amount of public data. Thus, efforts are warranted to develop a more powerful, adult-specific clock using a greater number of buccal samples from individuals aged 18 years or older.

Down syndrome subjects and moist snuff users exhibit epigenetic age acceleration

The collated datasets (Table S4) containing control samples additionally had samples derived from Down syndrome subjects (n = 10), moist snuff users (n = 40), cigarette smokers (n = 40), and patients with

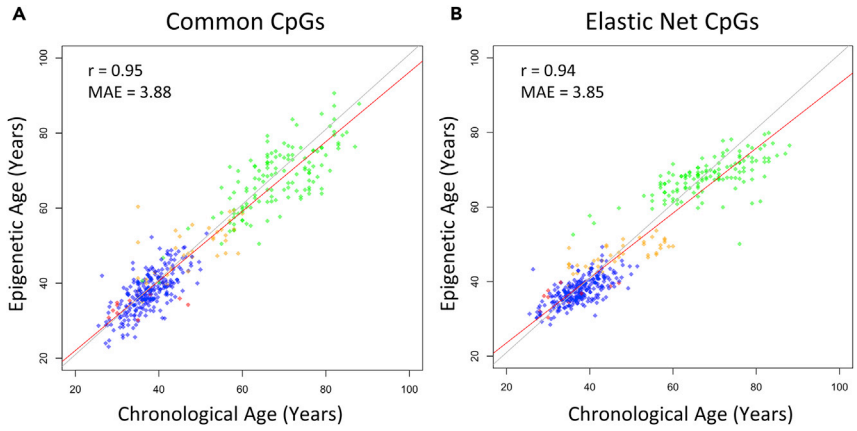


Figure 4. Proof-of-concept epigenetic aging clocks that predict age in human buccal tissue and saliva

(A) In conjunction with multiple linear regression, the 130 common CpGs were used to predict epigenetic age in human methylomic data with a Pearson correlation of 0.95 and a mean absolute error (MAE) of 3.88 years.

(B) A set of 335 CpGs were selected by elastic net regression and used to build an unbiased epigenetic clock in the same set of methylomic data. This model had a Pearson correlation of 0.94 and a MAE of 3.85 years. The four different colors (red, blue, green, and orange) represent four different publicly available datasets, three of which were derived from buccal tissue (red, blue, and orange) and one of which was generated in saliva (green). This dataset includes 431 samples (age range of 25–88 years) not associated with a disease, condition, or tobacco-associated behavior.

Parkinson's disease (n = 128). Using our proof-of-concept aging clocks, we next investigated whether or not these samples exhibit epigenetic age acceleration in comparison to controls not associated with a condition, disease, or tobacco-related behavior.

We found that individuals with Down syndrome have a significantly higher Δage compared to controls (Figure 5). The Wilcoxon p values were 0.005 and 0.004 for the common CpG (Figure 5A) and elastic net (Figure 5B) clocks, respectively. These data corroborate prior reports that epigenetic age is elevated in individuals with Down syndrome (Horvath et al., 2015; Xu et al., 2022). In line with evidence that smokeless tobacco increases the risk of mortality and chronic disease (Siddiqi et al., 2015), moist snuff users were predicted to have a higher Δage (Wilcoxon p = 0.004) according to the elastic net clock (Figure 5B). For both clocks, Δage differences between controls, Down syndrome subjects, moist snuff users, cigarette smokers, and Parkinson's disease patients are visualized in Figure S7. To assess the impact of sample size on accuracy metrics (Figure S8) and ability to detect age acceleration (Figure S9), we performed 100 random sampling runs of either 200 or 300 samples for the common CpG and elastic net clocks. These results (Figures S8 and S9) collectively demonstrate that clock performance is influenced by the number of samples available for training.

Clock inputs as molecules of interest

Previously, we collaborated with Drs. Benoit Lehallier and Tony-Wyss Coray to develop a uniquely accurate plasma proteomic aging clock. This model utilizes 491 proteins as inputs and was used to show that aerobic exercise-trained individuals have a lower proteomic age compared to sedentary subjects (Lehallier et al., 2020). A detailed bioinformatics analysis of this model showed that these proteins were significantly associated with immunological processes. Furthermore, an extensive literature review concluded that the majority of these proteins were previously reported to influence lifespan and/or age-related disease in animal models (Johnson et al., 2021). We similarly found that genes commonly prioritized for transcriptomic age prediction exhibit intimate connections to the immune system and age-related disease (Johnson and Shokhirev, 2021). It is therefore tempting to speculate that a portion of molecules important for age prediction may be more than simple biomarkers. Instead, some may actively underlie and contribute to age-related dysfunction.

In the present work, we have identified 130 common CpGs that were prioritized by two or more different epigenetic clocks used to predict age in buccal tissue. Based on our previous work, we were curious if the genes associated with these CpGs had known roles in the regulation of lifespan. To begin this exploration,

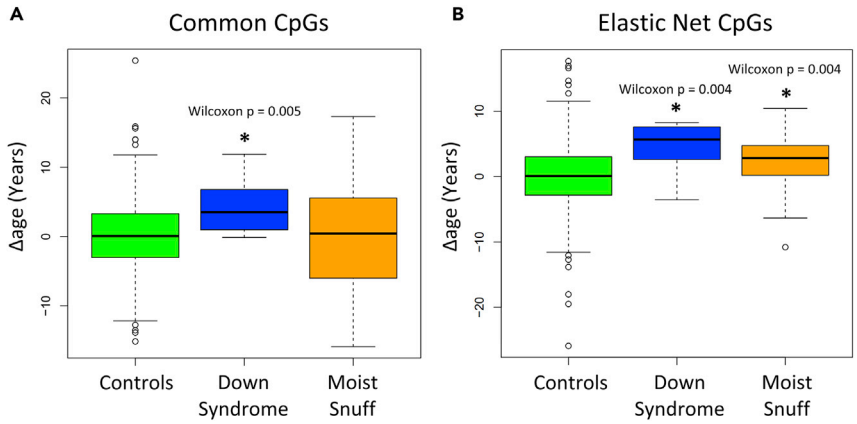


Figure 5. Epigenetic age acceleration in Down syndrome subjects and moist snuff users

Compared to samples not associated with a condition, disease, or tobacco-associated behavior, Δage was assessed in Down syndrome subjects (n = 10) and moist snuff users (n = 40).

(A) Using the common CpG clock, individuals with Down syndrome were predicted to be significantly older (Wilcoxon p = 0.005).

(B) Significantly higher Δage values were observed in both Down syndrome subjects (Wilcoxon p = 0.004) and moist snuff users (Wilcoxon p = 0.004) according to the elastic net clock. An asterisk indicates statistical significance compared to controls.

we uploaded our list of gene IDs to the Human Aging Genomic Resources database (Tacutu et al., 2018). Interestingly, nine genes had existing entries (Table S7), including existing lifespan effects for *CLOCK*, *FXN*, and *HTRA2*. In mice, creating a deficiency in *Clock* induces a premature aging phenotype characterized by a reduced lifespan and a higher likelihood of developing dermatitis and cataracts (Dubrovsky et al., 2010). Disrupting the expression of *Fxn* in mouse hepatocytes impairs mitochondrial function, lowers life expectancy, and leads to the development of hepatic tumors (Thierbach et al., 2005). Mutant mice that lack *Htra2* exhibit a phenotype similar to Parkinson's disease and die prematurely (Jones et al., 2003). Transgenically expressing human *HTRA2* in these mice elongates their lifespan and rescues their neurodegeneration phenotype. These mutant mice lacking mouse *Htra2* and expressing human *HTRA2* display a phenotype of accelerated aging and die between 12 and 17 months of age (Kang et al., 2013). For all three genes, the loss of expression is tethered to a longevity-lowering effect. It would be intriguing to assess if age-related methylation changes in *CLOCK*, *FXN*, and *HTRA2* are functionally significant.

This inspired us to do a mini-literature review (Table S7) to see if other genes associated with these common buccal CpGs influence lifespan in animal models. Using UniProt (UniProt, 2021) for up-to-date nomenclature and the Alliance of Genome Resources (Alliance of Genome Resources, 2020) for established orthologs, we searched for genes alone or in conjunction with either "lifespan" or "life span" in PubMed. This approach additionally implicated the following 16 genes in the regulation of lifespan in yeast, worms, flies, or mice: *BSN*, *CAMKK1*, *DLX5*, *ELOVL2*, *FZD9*, *GALC*, *GLO1*, *KCNC2*, *KCNC4*, *KLF2*, *MBNL1*, *RPL31*, *TBC1D23*, *TSPAN32*, *UBA7*, and *VPS18* (Table S7). Of the 19 genes identified (including those listed in the Human Aging Genomic Resources database), 10 examples are highlighted in Table 2. For instance, overexpressing *klf-3* (ortholog of *KLF2*) in worms enhances health and lifespan (Hsieh et al., 2017) while the genetic depletion of *Tsp2A* (ortholog of *TSPAN32*) in flies shortens lifespan and causes intestinal barrier dysfunction (Izumi et al., 2019). Our literature review also revealed that several of these genes have roles in the regulation of age-related dysfunction. Although by no means comprehensive, 10 literature connections are provided in Table 3 where the manipulation of an associated gene or protein influences neurodegeneration, visual decline, osteoporosis, atherosclerosis, diabetes, obesity, or cancer *in vivo*. An illustration of this is a study by Sharma et al., which showed that recombinant Scgn can protect against insulin insensitivity, obesity, and cardiovascular risk in mice fed a high-fat diet (Sharma et al., 2019). As a separate example, the overexpression of *Snx8* in Alzheimer's disease mice decreases amyloid-beta levels and guards against cognitive impairment (Xie et al., 2019).

Compounds associated with CpG-linked genes

Inspired by these literature connections, we searched the DrugBank database (Wishart et al., 2018) to see if any of the CpG-linked genes had known drug relations. Of the 130 common CpGs, a total of 39

Table 2. For the 130 common CpGs (CpGs used by two or more epigenetic clocks applied to human buccal tissue), we provide 10 examples of a CpG-associated gene influencing lifespan in an animal model. Whether or not lifespan is increased or decreased is respectively highlighted with an upward (↑) or downward (↓) arrow

Common CpG	Associated Gene	Reported Lifespan Effect
cg16933388	*BSN*	Transgenically overexpressing rat *Bsn* in neurons reduces climbing ability and lifespan in fruit flies (Schattling et al., 2019) ↓
cg05960024	*CLOCK*	*Clock*$^{-/-}$ mice are shorter-lived and display an increased incidence of dermatitis and cataracts (Dubrovsky et al., 2010) ↓
cg16867657, cg21572722	*ELOVL2*	Deleting *Elovl2* in mice causes an accelerated aging phenotype, including a reduced lifespan (Li et al., 2022) ↓
cg07158339	*FXN*	Disrupting the expression of *Fxn* lowers lifespan and leads to the development of hepatic tumors in mice (Thierbach et al., 2005) ↓
cg20692569	*FZD9*	Broadly inhibiting *fz3* (ortholog of *FZD9*) in fruit flies increases lifespan (Bouska and Bai, 2022) ↑
cg26824091	*GL O 1*	Lifespan in fruit flies is elongated via RNAi knockdown against *Glo1* (Bouska and Bai, 2022) ↑
cg02154074	*HTRA2*	Although mice lacking *Htra2* die prematurely and exhibit neurodegeneration, their short lifespan can be partially rescued by transgenically expressing human *HTRA2* in the CNS (Kang et al., 2013) ↓
cg26842024	*KLF2*	Both health and lifespan in worms are enhanced by the overexpression of *klf-3* (ortholog of *KLF2*) (Hsieh et al., 2017) ↑
cg17627559	*TSPAN32*	Genetically depleting *Tsp2A* (ortholog of *TSPAN32*) causes intestinal barrier dysfunction and shortens lifespan in flies (Izumi et al., 2019) ↓
cg19381811	*UBA7*	In fruit flies, *Uba1* (ortholog of *UBA7*) mutants are shorter-lived and exhibit motor impairment (Liu and Pfleger, 2013) ↓

were connected to a DrugBank entry via their associated gene (Table S7). Intriguingly, some of the implicated molecules have been reported to extend lifespan in model organisms. The compounds glycine, alpha-linolenic acid, S-adenosyl-L-homocysteine, and spermidine were identified in relation to *PIPOX*, *ELOVL2*, *PNMT*, and *ADRB1*, respectively. In 2019, the Interventions Testing Program demonstrated that glycine supplementation in genetically heterogeneous mice extends lifespan in males and females (Miller et al., 2019). In nematode worms, treatment with the omega-3 fatty acid alpha-linolenic acid leads to a dose-dependent increase in lifespan (Qi et al., 2017). Supplementation with S-adenosyl-*l*-homocysteine recapitulates the benefits of methionine restriction in *C. elegans*, including the activation of AMPK and enhanced longevity (Ogawa et al., 2022). In mice, the polyamine spermidine was reported to prolong lifespan and protect against cardiovascular disease (Eisenberg et al., 2016).

DISCUSSION

In this work, we identify a set of 130 common CpGs that were used by two or more different epigenetic aging clocks to predict age in human buccal tissue. Interestingly, many of the genes associated with these CpGs influence lifespan and/or age-related disease *in vivo*. For example, overexpressing *Vgf* in Alzheimer's disease mice rescues memory impairment and improves neuropathology (Beckmann et al., 2020). This is intriguing as VGF is reportedly underexpressed in biofluids from patients with Alzheimer's disease (Pedrero-Prieto et al., 2020), Parkinson's disease (Virreira Winter et al., 2021), amyotrophic lateral sclerosis (Zhao et al., 2008), frontotemporal dementia (Remnestal et al., 2020), and Lewy body dementia (van Steenoven et al., 2019). In the rat hippocampus, the expression of *Vgf* declines with age (Blalock et al., 2003; Porter et al., 2012). The concentration of *VGF* negatively correlates with age in the human brain (Meng et al., 2016). As part of our previous transcriptomic meta-analysis of human studies (Shokhirev and Johnson, 2021), we collated publicly available brain transcriptomes derived from 91 samples

Table 3. 10 examples are provided where the manipulation of a gene associated with a common CpG induces an *in vivo* effect relevant to age-related disease

Common CpG	Associated Gene	Literature Connection Pertinent to Age-related Disease
cg23124451	*CBX7*	Mice deficient in *Cbx7* develop carcinomas and adenomas in the lung and liver (Forzati et al., 2012)
cg16867657, cg21572722	*ELOVL2*	Introducing a point mutation in *Elovl2* accelerates visual decline and induces the premature appearance of age-related deposits (Chen et al., 2020)
cg21296230	*GREM1*	In atherosclerotic *Apoe*$^{-/-}$ mice, treatment with recombinant Grem1 attenuates atheroprogression and reduces the content of macrophages/monocytes in lesions (Muller et al., 2013)
cg19853760	*LGALS1*	In a mouse model of glioblastoma, deleting *LGALS1* in brain tumor stem cells increases survival and suppresses tumorigenesis (Sharanek et al., 2021)
cg06493994	*SCGN*	In mice fed a high-fat diet, treatment with recombinant Scgn protects against cardiovascular risk, obesity, and insulin insensitivity (Sharma et al., 2019)
cg22679120	*SNX8*	Overexpressing *Snx8* in Alzheimer's disease mice protects against cognitive impairment and decreases amyloid-beta levels (Xie et al., 2019)
cg09646392	*TNFSF13B*	While genetically deleting *Tnfsf13b* worsens diet-induced obesity, overexpressing *Tnfsf13b* increases energy expenditure and protects against weight gain in mice (Chan et al., 2021)
cg04084157	*VGF*	The overexpression of *Vgf* rescues neuropathology and memory impairment in Alzheimer's disease mice (Beckmann et al., 2020)
cg02071305	*VPS18*	Deleting *Vps18* in neural cells impairs autophagy and induces severe neurodegeneration (Peng et al., 2012)
cg11298786	*WNT3A*	Treating osteoporotic mice with human WNT3A protein accelerates bone repair (Liu et al., 2019a)

with an age range of 32–99 years. By analyzing the brain samples in this data, we find that *VGF* is significantly underexpressed (log-fold change = −2.27, FDR = 9.64×10^{-5}) in adults aged 70 + years compared to those aged <70 years. Furthermore, the expression of VGF along with other select proteins can be used to construct a machine learning classifier that distinguishes Alzheimer's disease samples from controls (Shokhirev and Johnson, 2022). Since three different clocks have reported that the *VGF*-associated CpG cg04084157 becomes hypermethylated with age (Horvath, 2013; Horvath et al., 2018; Levine et al., 2018), it would be interesting to explore whether or not methylation is driving this age-related decrease in gene expression.

Another gene to highlight is *ELOVL2*, which elongates long-chain, polyunsaturated fatty acids such as omega-3 (Chao and Skowronska-Krawczyk, 2020). The common buccal CpGs cg16867657 and cg21572722 are linked to *ELOVL2* and undergo hypermethylation with age (Horvath et al., 2018; McEwen et al., 2020). As we show in this study, they can also be used to create a minimalistic epigenetic aging clock. Creating a point mutation in *Elovl2* disrupts the enzymatic activity of Elovl2, causes premature visual decline, and leads to the early accumulation of age-related autofluorescent deposits in mice (Chen et al., 2020). In zebrafish, inactivating *elovl2* alters lipid composition, increases retinal thickness, and perturbs visual behavior (Dasyani et al., 2020). The deletion of *Elovl2* reduces lifespan and induces an accelerated aging phenotype in mice characterized by early hair loss as well as impairments in bone density, endurance, muscle strength, learning, and memory (Li et al., 2022). Separate work has shown that the ablation of *Elovl2* drives mitochondrial dysfunction in the mouse liver (Gomez Rodriguez et al., 2022). Given that the hypermethylation of this gene is tethered to a reduced expression level with age (Chen et al., 2020; Li et al., 2022), it is feasible that the methylation of *ELOVL2* contributes to age-related pathology. Suggestive of this possibility, reversing promoter hypermethylation via 5-Aza-2′-deoxycytidine rescues visual function in older, wild-type mice and elevates *Elovl2* expression (Chen et al., 2020). In *Elovl2*

knockout mice, supplementation with fish oil was found to partially rescue the accelerated aging phenotype (Li et al., 2022).

Notable connections also emerged when genes associated with common CpGs were analyzed in the DrugBank database. For example, the amino acid glycine (linked to the gene *PIPOX*) has been reported to extend lifespan in worms (Liu et al., 2019b), rats (Brind et al., 2011), and mice (Miller et al., 2019). In conjunction with the enzyme GNMT, glycine is essential for clearing methionine in mammals (Luka et al., 2009). Relevantly, the overexpression of *Gnmt* extends lifespan in fruit flies (Obata and Miura, 2015). *PIPOX* encodes for peroxisomal sarcosine oxidase, an enzyme that converts sarcosine to glycine (UniProt, 2021). A metabolomics screen recently identified sarcosine as a circulating metabolite that declines with age in both mice and humans. Sarcosine levels were increased by dietary restriction and elevated in long-lived Ames dwarf mice (Walters et al., 2018). Further research is justified to determine the relationship between the methylation status of this gene and concentrations of sarcosine and glycine. Efforts are also warranted to assess the ability of *PIPOX* to influence lifespan and/or healthspan in model organisms.

In this work, we also generate proof-of-concept epigenetic aging clocks that accurately predict age in publicly available methylomic data. While the common CpG clock predicted age acceleration in Down syndrome subjects, the elastic net clock predicted significantly higher Δage values in both Down syndrome subjects and moist snuff users. These novel findings indicate that clocks optimized for adult buccal and saliva data can capture key aspects of health. They also corroborate previous work by Lowe et al., which concluded that cheek swabs are an informative and useful tissue for epigenetic research (Lowe et al., 2013). This is important given that the majority of work investigating epigenetic age acceleration in buccal tissue has been done in pediatric populations. Building off of the findings presented here, we are working on creating a more powerful aging clock using in-house cheek swab samples derived exclusively from adults. By employing a more advanced training regime, our goal is to build a next-generation, adult-focused buccal clock.

Regarding age acceleration, it was interesting to observe that neither proof-of-concept clock predicted a significant Δage difference in smokers or patients with Parkinson's disease. The Eipel 2016 clock similarly reported a non-significant association between smoking and epigenetic age in cheek swab samples (Eipel et al., 2016). Simpkin et al. reported a negative association between current smoking and the Horvath 2013 clock in buccal cells (Simpkin et al., 2017). The publicly available saliva dataset used in this study was previously analyzed by Chuang et al. The authors only detected five CpGs (cg15133963, cg01820192, cg22275276, cg11748881, and cg24742912) that significantly differed between controls and patients with Parkinson's disease (Chuang et al., 2017), none of which were present in our common CpG or elastic net clocks. These findings reveal the limitations of clocks simply trained on chronological age and emphasize the importance of building a next-generation model. In order to more broadly predict age-related outcomes, it is imperative to incorporate CpGs relevant to health.

Beyond building off of this work to create a more powerful adult buccal clock, there are many intriguing research directions that could be pursued. Recently, Roudbar et al. created an especially accurate whole blood aging clock by using a non-standard approach in conjunction with over 450,000 CpG sites present on Illumina's Infinium HumanMethylation450 array. Specifically, they developed a predictor using reproducing kernel Hilbert spaces regression and a ridge regression model in a Bayesian framework (Roudbar et al., 2021). Future efforts are warranted to compare the efficacy of clocks that use the entirety of an array to those that utilize a specific subset identified via a machine learning approach like elastic net regression. Unique modeling approaches should also be explored, such as best linear unbiased prediction (Zhang et al., 2019). Work involving technical and/or biological replicates is also needed to better optimize these clocks in terms of retest reliability. While the DunedinPACE model was able to achieve consistency across replicates by filtering out noisy CpGs (Belsky et al., 2022), other clocks have been reported to benefit from an additional PCA step (Higgins-Chen et al., 2022).

In conjunction with the *in vivo* results highlighted, our findings recommend future research into CpGs important for epigenetic age prediction. Given that only a subset of methylomic changes correlate with other omics changes, how age-related alterations in methylation correspond with shifts in gene and protein expression should be assessed. Furthermore, the ability of the genes and proteins associated with clock

CpGs to influence lifespan and/or age-related disease in animals needs to be evaluated comprehensively. Finally, the feasibility of these molecules as interventional targets should be explored in pre-clinical models.

Limitations of the study

This study has two primary limitations, the first of which is that we utilized publicly available human methylomic data. Since these data were not generated in-house, key information regarding the participants is lacking. Furthermore, the public data were created using Infinium HumanMethylation450 BeadChip arrays. While a powerful research tool, they do not provide the same level of coverage of the more expansive Infinium MethylationEPIC arrays. Secondly, we mine databases and literature to suggest that the age-related methylation changes of specific CpGs sites may underlie functional alterations. While we provide evidence that some of the genes linked to these CpGs potentially represent molecules of interest, our paper remains purely *in silico* and we do not perform any *in vivo* testing. Ultimately, experiments in animal models are required to fully understand the relationship between common CpG genes, lifespan, and age-related disease.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- METHODS DETAILS
  - Resources for omics data
  - Enrichment analyses
  - Machine learning

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2022.105304.

ACKNOWLEDGMENTS

The authors are thankful to internal funding from Tally Health, Inc. (Greenwich, Connecticut, United States of America). They would additionally like to expression gratitude to Mr. Devin Baker (Tally Health, Inc.) and Dr. David Sinclair (Harvard Medical School, Boston, Massachusetts, United States of America) for helpful discussions and support. They further note that the graphical abstract was created using BioRender (https://biorender.com/).

AUTHOR CONTRIBUTIONS

A.A.J. performed the enrichment analysis as well as contributed to project design, writing, and editing. N.S.T. and M.N.S. created the novel clocks, assisted with writing, and participated in editing. T.L.C. provided oversight, helped with project design, and performed editing.

DECLARATION OF INTERESTS

All authors are full-time employees of the biotechnology company Tally Health, Inc. The authors have no other conflicts of interest to declare.

Received: July 8, 2022
Revised: August 11, 2022
Accepted: October 2, 2022
Published: November 18, 2022

REFERENCES

Alghanim, H., Antunes, J., Silva, D., Alho, C.S., Balamurugan, K., and McCord, B. (2017). Detection and evaluation of DNA methylation markers found at SCGN and KLF14 loci to estimate human age. Forensic Sci. Int. Genet. *31*, 81–88. https://doi.org/10.1016/j.fsigen.2017.07.011.

Alliance of Genome Resources, C. (2020). Alliance of Genome Resources Portal: unified model organism research platform. Nucleic Acids Res. *48*, D650–D658. https://doi.org/10.1093/nar/gkz813.

Bao, X., Borné, Y., Xu, B., Orho-Melander, M., Nilsson, J., Melander, O., and Engström, G. (2021). Growth differentiation factor-15 is a biomarker for all-cause mortality but less evident for cardiovascular outcomes: a prospective study. Am. Heart J. *234*, 81–89. https://doi.org/10.1016/j.ahj.2020.12.020.

Becker, J., Böhme, P., Reckert, A., Eickhoff, S.B., Koop, B.E., Blum, J., Gündüz, T., Takayama, M., Wagner, W., and Ritz-Timme, S. (2022). Evidence for differences in DNA methylation between Germans and Japanese. Int. J. Legal Med. *136*, 405–413. https://doi.org/10.1007/s00414-021-02736-3.

Beckmann, N.D., Lin, W.J., Wang, M., Cohain, A.T., Charney, A.W., Wang, P., Ma, W., Wang, Y.C., Jiang, C., Audrain, M., et al. (2020). Multiscale causal networks identify VGF as a key regulator of Alzheimer's disease. Nat. Commun. *11*, 3942. https://doi.org/10.1038/s41467-020-17405-z.

Bell, C.G., Lowe, R., Adams, P.D., Baccarelli, A.A., Beck, S., Bell, J.T., Christensen, B.C., Gladyshev, V.N., Heijmans, B.T., Horvath, S., et al. (2019). DNA methylation aging clocks: challenges and recommendations. Genome Biol. *20*, 249. https://doi.org/10.1186/s13059-019-1824-y.

Belsky, D.W., Caspi, A., Corcoran, D.L., Sugden, K., Poulton, R., Arseneault, L., Baccarelli, A., Chamarti, K., Gao, X., Hannon, E., et al. (2022). DunedinPACE, a DNA methylation biomarker of the pace of aging. Elife *11*, e73420. https://doi.org/10.7554/eLife.73420.

Blalock, E.M., Chen, K.C., Sharrow, K., Herman, J.P., Porter, N.M., Foster, T.C., and Landfield, P.W. (2003). Gene microarrays in hippocampal aging: statistical profiling identifies novel processes correlated with cognitive impairment. J. Neurosci. *23*, 3807–3819.

Bouska, M.J., and Bai, H. (2022). Loxl2 is a mediator of cardiac aging in Drosophila melanogaster, genetically examining the role of aging clock genes. G3 *12*, jkab381. https://doi.org/10.1093/g3journal/jkab381.

Brind, J., Malloy, V., Augie, I., Caliendo, N., Vogelman, J.H., Zimmerman, J.A., and Orentreich, N. (2011). Dietary glycine supplementation mimics lifespan extension by dietary methionine restriction in Fisher 344 rats. FASEB. J. *25*, 528.2. https://doi.org/10.1096/fasebj.25.1_supplement.528.2.

Castle, J.R., Lin, N., Liu, J., Storniolo, A.M.V., Shendre, A., Hou, L., Horvath, S., Liu, Y., Wang, C., and He, C. (2020). Estimating breast tissue-specific DNA methylation age using next-generation sequencing data. Clin. Epigenetics *12*, 45. https://doi.org/10.1186/s13148-020-00834-4.

Chan, C.C., Harley, I.T.W., Pfluger, P.T., Trompette, A., Stankiewicz, T.E., Allen, J.L., Moreno-Fernandez, M.E., Damen, M.S.M.A., Oates, J.R., Alarcon, P.C., et al. (2021). A BAFF/APRIL axis regulates obesogenic diet-driven weight gain. Nat. Commun. *12*, 2911. https://doi.org/10.1038/s41467-021-23084-1.

Chao, D.L., and Skowronska-Krawczyk, D. (2020). ELOVL2: not just a biomarker of aging. Transl. Med. Aging *4*, 78–80. https://doi.org/10.1016/j.tma.2020.06.004.

Chatsirisupachai, K., Lesluyes, T., Paraoan, L., Van Loo, P., and de Magalhães, J.P. (2021). An integrative analysis of the age-associated multi-omic landscape across cancers. Nat. Commun. *12*, 2345. https://doi.org/10.1038/s41467-021-22560-y.

Chen, D., Chao, D.L., Rocha, L., Kolar, M., Nguyen Huu, V.A., Krawczyk, M., Dasyani, M., Wang, T., Jafari, M., Jabari, M., et al. (2020). The lipid elongation enzyme ELOVL2 is a molecular regulator of aging in the retina. Aging Cell *19*, e13100. https://doi.org/10.1111/acel.13100.

Chen, L.J., Trares, K., Laetsch, D.C., Nguyen, T.N.M., Brenner, H., and Schöttker, B. (2021). Systematic review and meta-analysis on the associations of polypharmacy and potentially inappropriate medication with adverse outcomes in older cancer patients. J. Gerontol. A Biol. Sci. Med. Sci. *76*, 1044–1052. https://doi.org/10.1093/gerona/glaa128.

Chuang, Y.H., Paul, K.C., Bronstein, J.M., Bordelon, Y., Horvath, S., and Ritz, B. (2017). Parkinson's disease is associated with DNA methylation levels in human blood and saliva. Genome Med. *9*, 76. https://doi.org/10.1186/s13073-017-0466-5.

Chun, S., Shin, D.W., Han, K., Jung, J.H., Kim, B., Jung, H.W., Son, K.Y., Lee, S.P., and Lee, S.C. (2021). The Timed up and Go test and the ageing heart: findings from a national health screening of 1, 084, 875 community-dwelling older adults. Eur. J. Prev. Cardiol. *28*, 213–219. https://doi.org/10.1177/2047487319882118.

Clough, E., and Barrett, T. (2016). The gene expression Omnibus database. Methods Mol. Biol. *1418*, 93–110. https://doi.org/10.1007/978-1-4939-3578-9_5.

Craig, T., Smelick, C., Tacutu, R., Wuttke, D., Wood, S.H., Stanley, H., Janssens, G., Savitskaya, E., Moskalev, A., Arking, R., and de Magalhães, J.P. (2015). The Digital Ageing Atlas: integrating the diversity of age-related changes into a unified resource. Nucleic Acids Res. *43*, D873–D878. https://doi.org/10.1093/nar/gku843.

Dasyani, M., Gao, F., Xu, Q., Van Fossan, D., Zhang, E., Pinto, A.F.M., Saghatelian, A., Skowronska-Krawczyk, D., and Chao, D.L. (2020). Elovl2 is required for robust visual function in zebrafish. Cells *9*, 2583. https://doi.org/10.3390/cells9122583.

Dempsey, P.C., Musicha, C., Rowlands, A.V., Davies, M., Khunti, K., Razieh, C., Timmins, I., Zaccardi, F., Codd, V., Nelson, C.P., et al. (2022). Investigation of a UK biobank cohort reveals causal associations of self-reported walking pace with telomere length. Commun. Biol. *5*, 381. https://doi.org/10.1038/s42003-022-03323-x.

Dubrovsky, Y.V., Samsa, W.E., and Kondratov, R.V. (2010). Deficiency of circadian protein CLOCK reduces lifespan and increases age-related cataract development in mice. Aging *2*, 936–944. https://doi.org/10.18632/aging.100241.

Eipel, M., Mayer, F., Arent, T., Ferreira, M.R.P., Birkhofer, C., Gerstenmaier, U., Costa, I.G., Ritz-Timme, S., and Wagner, W. (2016). Epigenetic age predictions based on buccal swabs are more precise in combination with cell type-specific DNA methylation signatures. Aging *8*, 1034–1048. https://doi.org/10.18632/aging.100972.

Eisenberg, T., Abdellatif, M., Schroeder, S., Primessnig, U., Stekovic, S., Pendl, T., Harger, A., Schipke, J., Zimmermann, A., Schmidt, A., et al. (2016). Cardioprotection and lifespan extension by the natural polyamine spermidine. Nat. Med. *22*, 1428–1438. https://doi.org/10.1038/nm.4222.

Forzati, F., Federico, A., Pallante, P., Abbate, A., Esposito, F., Malapelle, U., Sepe, R., Palma, G., Troncone, G., Scarfò, M., et al. (2012). CBX7 is a tumor suppressor in mice and humans. J. Clin. Invest. *122*, 612–623. https://doi.org/10.1172/JCI58620.

Gómez Rodríguez, A., Talamonti, E., Naudi, A., Kalinovich, A.V., Pauter, A.M., Barja, G., Bengtsson, T., Jacobsson, A., Pamplona, R., and Shabalina, I.G. (2022). Elovl2-Ablation leads to mitochondrial membrane fatty acid remodeling and reduced efficiency in mouse liver mitochondria. Nutrients *14*, 559. https://doi.org/10.3390/nu14030559.

Han, Y., Franzen, J., Stiehl, T., Gobs, M., Kuo, C.C., Nikolić, M., Hapala, J., Koop, B.E., Strathmann, K., Ritz-Timme, S., and Wagner, W. (2020). New targeted approaches for epigenetic age predictions. BMC Biol. *18*, 71. https://doi.org/10.1186/s12915-020-00807-2.

Hannum, G., Guinney, J., Zhao, L., Zhang, L., Hughes, G., Sadda, S., Klotzle, B., Bibikova, M., Fan, J.B., Gao, Y., et al. (2013). Genome-wide methylation profiles reveal quantitative views of human aging rates. Mol. Cell *49*, 359–367. https://doi.org/10.1016/j.molcel.2012.10.016.

Higgins-Chen, A.T., Thrush, K.L., Wang, Y., Minteer, C.J., Kuo, P.-L., Wang, M., Niimi, P., Sturm, G., Lin, J., Moore, A.Z., et al. (2022). A computational solution for bolstering reliability of epigenetic clocks: implications for clinical trials and longitudinal tracking. Nat. Aging *2*, 644–661. https://doi.org/10.1038/s43587-022-00248-2.

Horvath, S. (2013). DNA methylation age of human tissues and cell types. Genome Biol. *14*, R115. https://doi.org/10.1186/gb-2013-14-10-r115.

Horvath, S., Garagnani, P., Bacalini, M.G., Pirazzini, C., Salvioli, S., Gentilini, D., Di Blasio, A.M., Giuliani, C., Tung, S., Vinters, H.V., and Franceschi, C. (2015). Accelerated epigenetic aging in Down syndrome. Aging Cell *14*, 491–495. https://doi.org/10.1111/acel.12325.

Horvath, S., Oshima, J., Martin, G.M., Lu, A.T., Quach, A., Cohen, H., Felton, S., Matsuyama, M., Lowe, D., Kabacik, S., et al. (2018). Epigenetic clock for skin and blood cells applied to Hutchinson Gilford Progeria Syndrome and ex vivo studies. Aging *10*, 1758–1775. https://doi.org/10.18632/aging.101508.

Hsieh, P.N., Zhou, G., Yuan, Y., Zhang, R., Prosdocimo, D.A., Sangwung, P., Borton, A.H., Boriushkin, E., Hamik, A., Fujioka, H., et al. (2017). A conserved KLF-autophagy pathway modulates nematode lifespan and mammalian age-associated vascular dysfunction. Nat. Commun. *8*, 914. https://doi.org/10.1038/s41467-017-00899-5.

Izumi, Y., Furuse, K., and Furuse, M. (2019). Septate junctions regulate gut homeostasis through regulation of stem cell proliferation and enterocyte behavior in Drosophila. J. Cell Sci. *132*, jcs232108. https://doi.org/10.1242/jcs.232108.

Janssens, G.E., Lin, X.X., Millan-Ariño, L., Kavšek, A., Sen, I., Seinstra, R.I., Stroustrup, N., Nollen, E.A.A., and Riedel, C.G. (2019). Transcriptomics-based screening identifies pharmacological inhibition of Hsp90 as a means to defer aging. Cell Rep. *27*, 467–480.e6. https://doi.org/10.1016/j.celrep.2019.03.044.

Johnson, A.A., Akman, K., Calimport, S.R.G., Wuttke, D., Stolzing, A., and de Magalhães, J.P. (2012). The role of DNA methylation in aging, rejuvenation, and age-related disease. Rejuvenation Res. *15*, 483–494. https://doi.org/10.1089/rej.2012.1324.

Johnson, A.A., English, B.W., Shokhirev, M.N., Sinclair, D.A., and Cuellar, T.L. (2022). Human age reversal: fact or fiction? Aging Cell *21*, e13664. https://doi.org/10.1111/acel.13664.

Johnson, A.A., and Shokhirev, M.N. (2021). Pan-tissue aging clock genes that have intimate connections with the immune system and age-related disease. Rejuvenation Res. *24*, 377–389. https://doi.org/10.1089/rej.2021.0012.

Johnson, A.A., Shokhirev, M.N., and Lehallier, B. (2021). The protein inputs of an ultra-predictive aging clock represent viable anti-aging drug targets. Ageing Res. Rev. *70*, 101404. https://doi.org/10.1016/j.arr.2021.101404.

Johnson, A.A., Shokhirev, M.N., Wyss-Coray, T., and Lehallier, B. (2020). Systematic review and analysis of human proteomics aging studies unveils a novel proteomic aging clock and identifies key processes that change with age. Ageing Res. Rev. *60*, 101070. https://doi.org/10.1016/j.arr.2020.101070.

Jones, J.M., Datta, P., Srinivasula, S.M., Ji, W., Gupta, S., Zhang, Z., Davies, E., Hajnóczky, G., Saunders, T.L., Van Keuren, M.L., et al. (2003). Loss of Omi mitochondrial protease activity causes the neuromuscular disorder of mnd2 mutant mice. Nature *425*, 721–727. https://doi.org/10.1038/nature02052.

Jung, S.E., Lim, S.M., Hong, S.R., Lee, E.H., Shin, K.J., and Lee, H.Y. (2019). DNA methylation of the ELOVL2, FHL2, KLF14, C1orf132/MIR29B2C, and TRIM59 genes for age prediction from blood, saliva, and buccal swab samples. Forensic Sci. Int. Genet. *38*, 1–8. https://doi.org/10.1016/j.fsigen.2018.09.010.

Kang, S., Louboutin, J.P., Datta, P., Landel, C.P., Martinez, D., Zervos, A.S., Strayer, D.S., Fernandes-Alnemri, T., and Alnemri, E.S. (2013). Loss of HtrA2/Omi activity in non-neuronal tissues of adult mice causes premature aging. Cell Death Differ. *20*, 259–269. https://doi.org/10.1038/cdd.2012.117.

Kim, S., Na, J.G., Hampsey, M., and Reinberg, D. (1997). The Dr1/DRAP1 heterodimer is a global repressor of transcription in vivo. Proc. Natl. Acad. Sci. USA *94*, 820–825. https://doi.org/10.1073/pnas.94.3.820.

Koop, B.E., Mayer, F., Gündüz, T., Blum, J., Becker, J., Schaffrath, J., Wagner, W., Han, Y., Boehme, P., and Ritz-Timme, S. (2021). Postmortem age estimation via DNA methylation analysis in buccal swabs from corpses in different stages of decomposition-a "proof of principle" study. Int. J. Legal Med. *135*, 167–173. https://doi.org/10.1007/s00414-020-02360-7.

Kuhn, M. (2008). Building predictive models in R using the caret package. J. Stat. Softw. *28*, 1–26. https://doi.org/10.18637/jss.v028.i05.

Lehallier, B., Gate, D., Schaum, N., Nanasi, T., Lee, S.E., Yousef, H., Moran Losada, P., Berdnik, D., Keller, A., Verghese, J., et al. (2019). Undulating changes in human plasma proteome profiles across the lifespan. Nat. Med. *25*, 1843–1850. https://doi.org/10.1038/s41591-019-0673-2.

Lehallier, B., Shokhirev, M.N., Wyss-Coray, T., and Johnson, A.A. (2020). Data mining of human plasma proteins generates a multitude of highly predictive aging clocks that reflect different aspects of aging. Aging Cell *19*, e13256. https://doi.org/10.1111/acel.13256.

Levine, M.E., Lu, A.T., Quach, A., Chen, B.H., Assimes, T.L., Bandinelli, S., Hou, L., Baccarelli, A.A., Stewart, J.D., Li, Y., et al. (2018). An epigenetic biomarker of aging for lifespan and healthspan. Aging *10*, 573–591. https://doi.org/10.18632/aging.101414.

Li, X., Ploner, A., Wang, Y., Zhan, Y., Pedersen, N.L., Magnusson, P.K., Jylhävä, J., and Hägg, S. (2021). Clinical biomarkers and associations with healthspan and lifespan: evidence from observational and genetic data. EBioMedicine *66*, 103318. https://doi.org/10.1016/j.ebiom.2021.103318.

Li, X., Wang, J., Wang, L., Gao, Y., Feng, G., Li, G., Zou, J., Yu, M., Li, Y.F., Liu, C., et al. (2022). Lipid metabolism dysfunction induced by age-dependent DNA methylation accelerates aging. Signal Transduct. Target. Ther. *7*, 162. https://doi.org/10.1038/s41392-022-00964-6.

Liao, Y., Wang, J., Jaehnig, E.J., Shi, Z., and Zhang, B. (2019). WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. Nucleic Acids Res. *47*, W199–W205. https://doi.org/10.1093/nar/gkz401.

Liu, H.Y., and Pfleger, C.M. (2013). Mutation in E1, the ubiquitin activating enzyme, reduces Drosophila lifespan and results in motor impairment. PLoS One *8*, e32835. https://doi.org/10.1371/journal.pone.0032835.

Liu, Y., Li, Z., Arioka, M., Wang, L., Bao, C., and Helms, J.A. (2019a). WNT3A accelerates delayed alveolar bone repair in ovariectomized mice. Osteoporos. Int. *30*, 1873–1885. https://doi.org/10.1007/s00198-019-05071-x.

Liu, Y.J., Janssens, G.E., McIntyre, R.L., Molenaars, M., Kamble, R., Gao, A.W., Jongejan, A., Weeghel, M.V., MacInnes, A.W., and Houtkooper, R.H. (2019b). Glycine promotes longevity in Caenorhabditis elegans in a methionine cycle-dependent fashion. PLoS Genet. *15*, e1007633. https://doi.org/10.1371/journal.pgen.1007633.

López-Otín, C., Blasco, M.A., Partridge, L., Serrano, M., and Kroemer, G. (2013). The hallmarks of aging. Cell *153*, 1194–1217. https://doi.org/10.1016/j.cell.2013.05.039.

Lowe, R., Gemma, C., Beyan, H., Hawa, M.I., Bazeos, A., Leslie, R.D., Montpetit, A., Rakyan, V.K., and Ramagopalan, S.V. (2013). Buccals are likely to be a more informative surrogate tissue than blood for epigenome-wide association studies. Epigenetics *8*, 445–454. https://doi.org/10.4161/epi.24362.

Lu, A.T., Quach, A., Wilson, J.G., Reiner, A.P., Aviv, A., Raj, K., Hou, L., Baccarelli, A.A., Li, Y., Stewart, J.D., et al. (2019). DNA methylation GrimAge strongly predicts lifespan and healthspan. Aging *11*, 303–327. https://doi.org/10.18632/aging.101684.

Lujan, C., Tyler, E.J., Ecker, S., Webster, A.P., Stead, E.R., Martinez Miguel, V.E., Milligan, D., Garbe, J.C., Stampfer, M.R., Beck, S., et al. (2020). A CellAgeClock for expedited discovery of anti-ageing compounds. Preprint at bioRxiv. https://doi.org/10.1101/803676.

Luka, Z., Mudd, S.H., and Wagner, C. (2009). Glycine N-methyltransferase and regulation of S-adenosylmethionine levels. J. Biol. Chem. *284*, 22507–22511. https://doi.org/10.1074/jbc.R109.019273.

McEwen, L.M., O'Donnell, K.J., McGill, M.G., Edgar, R.D., Jones, M.J., MacIsaac, J.L., Lin, D.T.S., Ramadori, K., Morin, A., Gladish, N., et al. (2020). The PedBE clock accurately estimates DNA methylation age in pediatric buccal cells. Proc. Natl. Acad. Sci. USA *117*, 23329–23335. https://doi.org/10.1073/pnas.1820843116.

McIntyre, R.L., Rahman, M., Vanapalli, S.A., Houtkooper, R.H., and Janssens, G.E. (2021). Biological age prediction from wearable device movement data identifies nutritional and pharmacological interventions for healthy aging. Front. Aging *2*, 708680. https://doi.org/10.3389/fragi.2021.708680.

Meng, G., Zhong, X., and Mei, H. (2016). A systematic investigation into aging related genes in brain and their relationship with Alzheimer's disease. PLoS One *11*, e0150624. https://doi.org/10.1371/journal.pone.0150624.

Miller, R.A., Harrison, D.E., Astle, C.M., Bogue, M.A., Brind, J., Fernandez, E., Flurkey, K., Javors, M., Ladiges, W., Leeuwenburgh, C., et al. (2019). Glycine supplementation extends lifespan of male and female mice. Aging Cell *18*, e12953. https://doi.org/10.1111/acel.12953.

Müller, I., Schönberger, T., Schneider, M., Borst, O., Ziegler, M., Seizer, P., Leder, C., Müller, K., Lang, M., Appenzeller, F., et al. (2013). Gremlin-1 is an inhibitor of macrophage migration inhibitory factor and attenuates atherosclerotic plaque

growth in ApoE-/- Mice. J. Biol. Chem. *288*, 31635–31645. https://doi.org/10.1074/jbc.M113.477745.

Murphy, S.K., Yang, H., Moylan, C.A., Pang, H., Dellinger, A., Abdelmalek, M.F., Garrett, M.E., Ashley-Koch, A., Suzuki, A., Tillmann, H.L., et al. (2013). Relationship between methylome and transcriptome in patients with nonalcoholic fatty liver disease. Gastroenterology *145*, 1076–1087. https://doi.org/10.1053/j.gastro.2013.07.047.

Obata, F., and Miura, M. (2015). Enhancing S-adenosyl-methionine catabolism extends Drosophila lifespan. Nat. Commun. *6*, 8332. https://doi.org/10.1038/ncomms9332.

Ogawa, T., Masumura, K., Kohara, Y., Kanai, M., Soga, T., Ohya, Y., Blackwell, T.K., and Mizunuma, M. (2022). S-adenosyl-L-homocysteine extends lifespan through methionine restriction effects. Aging Cell *21*, e13604. https://doi.org/10.1111/acel.13604.

Oughtred, R., Rust, J., Chang, C., Breitkreutz, B.J., Stark, C., Willems, A., Boucher, L., Leung, G., Kolas, N., Zhang, F., et al. (2021). The BioGRID database: a comprehensive biomedical resource of curated protein, genetic, and chemical interactions. Protein Sci. *30*, 187–200. https://doi.org/10.1002/pro.3978.

Pascual-Torner, M., Carrero, D., Pérez-Silva, J.G., Álvarez-Puente, D., Roiz-Valle, D., Bretones, G., Rodríguez, D., Maeso, D., Mateo-González, E., Español, Y., et al. (2022). Comparative genomics of mortal and immortal cnidarians unveils novel keys behind rejuvenation. Proc. Natl. Acad. Sci. USA *119*. e2118763119. https://doi.org/10.1073/pnas.2118763119.

Pavasini, R., Serenelli, M., Celis-Morales, C.A., Gray, S.R., Izawa, K.P., Watanabe, S., Colin-Ramirez, E., Castillo-Martínez, L., Izumiya, Y., Hanatani, S., et al. (2019). Grip strength predicts cardiac adverse events in patients with cardiac disorders: an individual patient pooled meta-analysis. Heart *105*, 834–841. https://doi.org/10.1136/heartjnl-2018-313816.

Pedrero-Prieto, C.M., García-Carpintero, S., Frontiñán-Rubio, J., Llanos-González, E., Aguilera García, C., Alcaín, F.J., Lindberg, I., Durán-Prado, M., Peinado, J.R., and Rabanal-Ruiz, Y. (2020). A comprehensive systematic review of CSF proteins and peptides that define Alzheimer's disease. Clin. Proteomics *17*, 21. https://doi.org/10.1186/s12014-020-09276-9.

Peng, C., Ye, J., Yan, S., Kong, S., Shen, Y., Li, C., Li, Q., Zheng, Y., Deng, K., Xu, T., and Tao, W. (2012). Ablation of vacuole protein sorting 18 (Vps18) gene leads to neurodegeneration and impaired neuronal migration by disrupting multiple vesicle transport pathways to lysosomes. J. Biol. Chem. *287*, 32861–32873. https://doi.org/10.1074/jbc.M112.384305.

Porter, N.M., Bohannon, J.H., Curran-Rauhut, M., Buechel, H.M., Dowling, A.L.S., Brewer, L.D., Popovic, J., Thibault, V., Kraner, S.D., Chen, K.C., and Blalock, E.M. (2012). Hippocampal CA1 transcriptional profile of sleep deprivation: relation to aging and stress. PLoS One *7*, e40128. https://doi.org/10.1371/journal.pone.0040128.

Qi, W., Gutierrez, G.E., Gao, X., Dixon, H., McDonough, J.A., Marini, A.M., and Fisher, A.L. (2017). The omega-3 fatty acid alpha-linolenic acid extends Caenorhabditis elegans lifespan via NHR-49/PPARalpha and oxidation to oxylipins. Aging Cell *16*, 1125–1135. https://doi.org/10.1111/acel.12651.

Remnestål, J., Öijerstedt, L., Ullgren, A., Olofsson, J., Bergström, S., Kultima, K., Ingelsson, M., Kilander, L., Uhlén, M., Månberg, A., et al. (2020). Altered levels of CSF proteins in patients with FTD, presymptomatic mutation carriers and non-carriers. Transl. Neurodegener. *9*, 27. https://doi.org/10.1186/s40035-020-00198-y.

Roudbar, M.A., Mousavi, S.F., Ardestani, S.S., Lopes, F.B., Momen, M., Gianola, D., and Khatib, H. (2021). Prediction of biological age and evaluation of genome-wide dynamic methylomic changes throughout human aging. G3 *11*, jkab112. https://doi.org/10.1093/g3journal/jkab112.

Schattling, B., Engler, J.B., Volkmann, C., Rothammer, N., Woo, M.S., Petersen, M., Winkler, I., Kaufmann, M., Rosenkranz, S.C., Fejtova, A., et al. (2019). Bassoon proteinopathy drives neurodegeneration in multiple sclerosis. Nat. Neurosci. *22*, 887–896. https://doi.org/10.1038/s41593-019-0385-4.

Schwender, K., Holländer, O., Klopfleisch, S., Eveslage, M., Danzer, M.F., Pfeiffer, H., and Vennemann, M. (2021). Development of two age estimation models for buccal swab samples based on 3 CpG sites analyzed with pyrosequencing and minisequencing. Forensic Sci. Int. Genet. *53*, 102521. https://doi.org/10.1016/j.fsigen.2021.102521.

Sharanek, A., Burban, A., Hernandez-Corchado, A., Madrigal, A., Fatakdawala, I., Najafabadi, H.S., Soleimani, V.D., and Jahani-Asl, A. (2021). Transcriptional control of brain tumor stem cells by a carbohydrate binding protein. Cell Rep. *36*, 109647. https://doi.org/10.1016/j.celrep.2021.109647.

Sharma, A.K., Khandelwal, R., Kumar, M.J.M., Ram, N.S., Chidananda, A.H., Raj, T.A., and Sharma, Y. (2019). Secretagogin regulates insulin signaling by direct insulin binding. iScience *21*, 736–753. https://doi.org/10.1016/j.isci.2019.10.066.

Shokhirev, M.N., and Johnson, A.A. (2021). Modeling the human aging transcriptome across tissues, health status, and sex. Aging Cell *20*, e13280. https://doi.org/10.1111/acel.13280.

Shokhirev, M.N., and Johnson, A.A. (2022). An integrative machine-learning meta-analysis of high-throughput omics data identifies age-specific hallmarks of Alzheimer's disease. Ageing Res. Rev. *81*, 101721. https://doi.org/10.1016/j.arr.2022.101721.

Siddiqi, K., Shah, S., Abbas, S.M., Vidyasagaran, A., Jawad, M., Dogar, O., and Sheikh, A. (2015). Global burden of disease due to smokeless tobacco consumption in adults: analysis of data from 113 countries. BMC Med. *13*, 194. https://doi.org/10.1186/s12916-015-0424-2.

Simpkin, A.J., Cooper, R., Howe, L.D., Relton, C.L., Davey Smith, G., Teschendorff, A., Widschwendter, M., Wong, A., Kuh, D., and Hardy, R. (2017). Are objective measures of physical capability related to accelerated epigenetic age? Findings from a British birth cohort. BMJ Open *7*, e016708. https://doi.org/10.1136/bmjopen-2017-016708.

Simpson, D.J., and Chandra, T. (2021). Epigenetic age prediction. Aging Cell *20*, e13452. https://doi.org/10.1111/acel.13452.

Slieker, R.C., Relton, C.L., Gaunt, T.R., Slagboom, P.E., and Heijmans, B.T. (2018). Age-related DNA methylation changes are tissue-specific with ELOVL2 promoter methylation as exception. Epigenet. Chromatin *11*, 25. https://doi.org/10.1186/s13072-018-0191-3.

Tacutu, R., Thornton, D., Johnson, E., Budovsky, A., Barardo, D., Craig, T., Diana, E., Lehmann, G., Toren, D., Wang, J., et al. (2018). Human ageing genomic resources: new and updated databases. Nucleic Acids Res. *46*, D1083–D1090. https://doi.org/10.1093/nar/gkx1042.

Teschendorff, A.E., Breeze, C.E., Zheng, S.C., and Beck, S. (2017). A comparison of reference-based algorithms for correcting cell-type heterogeneity in Epigenome-Wide Association Studies. BMC Bioinf. *18*, 105. https://doi.org/10.1186/s12859-017-1511-5.

The Gene Ontology Consortium (2019). The gene Ontology resource: 20 years and still GOing strong. Nucleic Acids Res. *47*, D330–D338. https://doi.org/10.1093/nar/gky1055.

Theda, C., Hwang, S.H., Czajko, A., Loke, Y.J., Leong, P., and Craig, J.M. (2018). Quantitation of the cellular content of saliva and buccal swab samples. Sci. Rep. *8*, 6944. https://doi.org/10.1038/s41598-018-25311-0.

Thierbach, R., Schulz, T.J., Isken, F., Voigt, A., Mietzner, B., Drewes, G., von Kleist-Retzow, J.C., Wiesner, R.J., Magnuson, M.A., Puccio, H., et al. (2005). Targeted disruption of hepatic frataxin expression causes impaired mitochondrial function, decreased life span and tumor growth in mice. Hum. Mol. Genet. *14*, 3857–3864. https://doi.org/10.1093/hmg/ddi410.

UniProt Consortium (2021). UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Res. *49*, D480–D489. https://doi.org/10.1093/nar/gkaa1100.

Uotinen, V., Rantanen, T., and Suutama, T. (2005). Perceived age as a predictor of old age mortality: a 13-year prospective study. Age Ageing *34*, 368–372. https://doi.org/10.1093/ageing/afi091.

van Steenoven, I., Noli, B., Cocco, C., Ferri, G.L., Oeckl, P., Otto, M., Koel-Simmelink, M.J.A., Bridel, C., van der Flier, W.M., Lemstra, A.W., and Teunissen, C.E. (2019). VGF peptides in cerebrospinal fluid of patients with dementia with Lewy bodies. Int. J. Mol. Sci. *20*, E4674. https://doi.org/10.3390/ijms20194674.

Virreira Winter, S., Karayel, O., Strauss, M.T., Padmanabhan, S., Surface, M., Merchant, K., Alcalay, R.N., and Mann, M. (2021). Urinary proteome profiling for stratifying patients with familial Parkinson's disease. EMBO Mol. Med. *13*, e13257. https://doi.org/10.15252/emmm.202013257.

Walters, R.O., Arias, E., Diaz, A., Burgos, E.S., Guan, F., Tiano, S., Mao, K., Green, C.L., Qiu, Y., Shah, H., et al. (2018). Sarcosine is uniquely

modulated by aging and dietary restriction in rodents and humans. Cell Rep. *25*, 663–676.e6. https://doi.org/10.1016/j.celrep.2018.09.065.

Wang, J., Ma, Z., Carr, S.A., Mertins, P., Zhang, H., Zhang, Z., Chan, D.W., Ellis, M.J.C., Townsend, R.R., Smith, R.D., et al. (2017). Proteome profiling outperforms transcriptome profiling for coexpression based gene function prediction. Mol. Cell. Proteomics *16*, 121–134. https://doi.org/10.1074/mcp.M116.060301.

Wishart, D.S., Feunang, Y.D., Guo, A.C., Lo, E.J., Marcu, A., Grant, J.R., Sajed, T., Johnson, D., Li, C., Sayeeda, Z., et al. (2018). DrugBank 5.0: a major update to the DrugBank database for 2018. Nucleic Acids Res. *46*, D1074–D1082. https://doi.org/10.1093/nar/gkx1037.

Xie, Y., Niu, M., Ji, C., Huang, T.Y., Zhang, C., Tian, Y., Shi, Z., Wang, C., Zhao, Y., Luo, H., et al. (2019). SNX8 enhances non-amyloidogenic APP trafficking and attenuates abeta accumulation and memory deficits in an AD mouse. Front. Cell. Neurosci. *13*, 410. https://doi.org/10.3389/fncel.2019.00410.

Xu, K., Li, S., Muskens, I.S., Elliott, N., Myint, S.S., Pandey, P., Hansen, H.M., Morimoto, L.M., Kang, A.Y., Ma, X., et al. (2022). Accelerated epigenetic aging in newborns with down syndrome. Aging Cell *21*, e13652. https://doi.org/10.1111/acel.13652.

Zhang, Q., Vallerga, C.L., Walker, R.M., Lin, T., Henders, A.K., Montgomery, G.W., He, J., Fan, D., Fowdar, J., Kennedy, M., et al. (2019). Improved precision of epigenetic clock estimates across tissues and its implication for biological ageing. Genome Med. *11*, 54. https://doi.org/10.1186/s13073-019-0667-1.

Zhao, Z., Lange, D.J., Ho, L., Bonini, S., Shao, B., Salton, S.R., Thomas, S., and Pasinetti, G.M. (2008). Vgf is a novel biomarker associated with muscle weakness in amyotrophic lateral sclerosis (ALS), with a potential role in disease pathogenesis. Int. J. Med. Sci. *5*, 92–99. https://doi.org/10.7150/ijms.5.92.

Zheng, S.C., Breeze, C.E., Beck, S., Dong, D., Zhu, T., Ma, L., Ye, W., Zhang, G., and Teschendorff, A.E. (2019). EpiDISH web server: epigenetic dissection of intra-sample-heterogeneity with online GUI. Bioinformatics *36*, btz833. https://doi.org/10.1093/bioinformatics/btz833.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Available data		
Methylomic dataset	Gene Expression Omnibus (Clough and Barrett, 2016) database	GSE50586
Methylomic dataset	Gene Expression Omnibus (Clough and Barrett, 2016) database	GSE137688
Methylomic dataset	Gene Expression Omnibus (Clough and Barrett, 2016) database	GSE94876
Methylomic dataset	Gene Expression Omnibus (Clough and Barrett, 2016) database	GSE111223
Raw count and meta-data tables from our previous transcriptomic aging meta-analysis (Shokhirev and Johnson, 2021)	Mendeley Data	https://doi.org/10.17632/92rgnswtn8.1
Clock intercept and coefficient information	Current paper	Table S5
Software and algorithms		
R studio 2022.02.3	https://www.rstudio.com/	N/A
R software v.4.1.3	https://www.r-project.org/	N/A
caret v6.0-92 R package	https://cran.r-project.org/web/packages/caret/index.html	N/A
EpiDISH v2.10.0 R package	https://www.bioconductor.org/packages/release/bioc/html/EpiDISH.html	N/A
ggplot2 v3.3.6 R package	https://cran.r-project.org/web/packages/ggplot2/index.html	N/A
GEOquery v2.62.2 R package	https://bioconductor.org/packages/release/bioc/html/GEOquery.html	N/A
glmnet v4.1-4 R package	https://cran.r-project.org/web/packages/glmnet/index.html	N/A
minfi v1.40.0 R package	https://bioconductor.org/packages/release/bioc/html/minfi.html	N/A
Code used in this paper	https://github.com/Tally-Health/BuccalComparison/	N/A

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to the Lead Contact, Adiv Johnson (adiv@tallyhealth.com).

Materials availability

No new materials were generated in this study.

Data and code availability

- The methylomic datasets used in the study are all publicly available via the Gene Expression Omnibus (Clough and Barrett, 2016) database under the following accession IDs: GSE50586, GSE137688, GSE94876, and GSE111223. The raw count and meta-data tables from our previous transcriptomic aging meta-analysis (Shokhirev and Johnson, 2021) are available on Mendeley Data under the digital object identifier https://doi.org/10.17632/92rgnswtn8.1. Clock intercept and coefficient information is provided in Table S5.

- Code, coefficients, and plots are available at https://github.com/Tally-Health/BuccalComparison/.
- Additional information is available from the lead contact upon request.

METHODS DETAILS

Resources for omics data

Multiple omics resources are used throughout this study. An online tool developed by Lehallier et al (Lehallier et al., 2019) allows for plasma protein expression with age to be interrogated in a cohort of 4,263 adults aged 18–95 years old: https://twc-stanford.shinyapps.io/aging_plasma_proteome/. Previously, we performed a systematic review of human proteomics aging studies and identified 1,128 proteins reported to significantly change with age according to multiple studies. Information for these common proteins can be found in Table S2 of that paper (Johnson et al., 2020). In a prior meta-analysis of human transcriptomic data containing age information, we collated and filtered 3,060 RNA-Seq datasets from various tissues. Table S2 of that work lists the 1,000 genes that were either the most differentially expressed or displayed the most variability in expression with age (Shokhirev and Johnson, 2021). The Digital Aging Atlas contains information regarding age-related changes across different types of omics data and can be accessed online: https://ageing-map.org/.

Enrichment analyses

We performed network topology-based enrichment analyses (Wang et al., 2017) using the PPI BioGRID database (Oughtred et al., 2021) and WebGestalt (Liao et al., 2019). In the "Advanced Parameters" section, the "Network Construction Method" was "Network Expansion", the "Set Number of Top Ranking Neighbors" was equivalent to half of the input size, the "Significance Level" was set to an FDR cut-off value of 0.05, and seeds were highlighted. The output was significantly enriched results in the Gene Ontology Biological Process database (The Gene Ontology, 2019). Inputs were unique gene IDs (Tables S3 and S6) associated with specific DNA methylation sites. WebGestalt can be accessed via the following website: http://www.webgestalt.org/.

Machine learning

Four methylomic datasets were downloaded from the Gene Expression Omnibus database (Clough and Barrett, 2016) and combined along shared CpGs, resulting in a 649 sample by 485,512 CpG matrix of beta values. Since cell-type composition variation can affect model training, we used the popular reference based EpiDISH (Zheng et al., 2019) package (v2.6.1) for estimating cell type proportions. We used the hierarchical EpiDISH procedure to estimate the fraction of epithelial, fibroblast, and various immune subsets via the centEpiFibIC.m and centBloodSub.m references. The robust partial corrections method was utilized (Teschendorff et al., 2017). After estimating cell types, the beta values were translated into M values for downstream analysis:

$$M = log_2\left(\frac{Beta}{1 - Beta}\right)$$

We trained an unbiased model on 431 samples from individuals not associated with a specific disease, condition, or tobacco-related behavior in the combined dataset and filtered the 485,512 CpG sites. First, CpGs with variance $\leq$0.1 and CpGs with average gaps between sorted M values $\geq$0.02 were removed to avoid modeling noisy data. Then, the top 10,000 CpGs most correlated with age were used as potential inputs. We added the cell type fractions for each sample as additional training features (since fractions add to 1, we omitted the eosinophil fraction). We fit an elastic net (glmnet) model to the data using the R caret (Kuhn, 2008) package (v 6.0–86) train function. The elastic net model is a generalized linear regression model that finds and groups predictors that are correlated with each other. From there, it selects a feature from the group to keep or omit.

Tuning parameters included 10-fold cross-validation repeated once and allowed for a range of alpha and lambda values that minimize root-mean-square error. 10-fold cross-validation operates by training the model 10 times, each time using a randomly selected nine-tenths of the input data to train the model and reserving one-10th as the test set. Optimal elastic net hyperparameters were α = 0.05 and λ = 0.95. Given that our dataset had an age range of 25–88 years, predictions were constrained to a range of 0–100 years during training. The output is a linear model with the following format:

$$Epigenetic\ Age = \beta_0 + \beta_1 X_1 + \ldots + \beta_n X_n$$

In this equation, B_0 is the intercept, B_1 - B_n are the coefficients, and X_1 - X_n represent CpGs sites.

In addition to the unbiased model, we used the 130 common buccal CpG sites to train a multiple linear regression (LM) model using the caret package. This model uses all of the input features, 130 common buccal CpG sites (in this case), and assigns a coefficient based on the feature's correlation with age and weight. One M value was missing from the dataset and imputed using the median. As with the unbiased model, cell type fractions for each sample were added as additional training features. 10-fold cross-validation was repeated once to optimize coefficients and allow for resampling of training and test data. The final linear model is in the same format as described for the unbiased model. We additionally used the *ELOVL2*-associated CpG sites cg16867657 and cg21572722 to train an LM model using the caret package.

In comparison to controls not associated with a disease, condition, or tobacco-related behavior, both models were used to estimate epigenetic age in cigarette smokers, moist snuff users, Parkinson's disease patients, and Down syndrome subjects. Since models were trained on the controls, we used the values predicted when each control sample was held out during 10-fold cross validation to avoid predicting on trained data. The Δage for each individual was calculated (epigenetic age – chronological age) and the values for each of the listed groups were compared to those in the control subset using a Wilcoxon rank-sum test to measure significance. Significance was defined as a Wilcoxon p value less than 0.05.

As described above, we used the reference based EpiDISH (Zheng et al., 2019) package (v2.6.1) for estimating cell type proportions. Instead of including cell type proportions as separate features for model training, we explored the effect of adjusting the beta values *a priori*. For this analysis, we adjusted the methylation beta values of the matrix based on the predicted cell type proportions of epithelial cells, neutrophils, and combined immune subsets. To accomplish this, the original beta values were first modeled as a linear model of cell type compositions and then the betas were replaced with the mean plus the residuals. Adjusted beta values were then translated into M values for downstream analysis, as described above. We continued to train an unbiased model on the 431 individuals not associated with a specific disease, condition, or tobacco-related behavior in the combined dataset using the top 10,000 CpGs most correlated with age and did not include cell type fractions for each sample as training input. We trained a separate LM model on the 431 control samples that did not include cell type fractions for each sample as training inputs.