

Development of a novel transcriptomic measure of aging: Transcriptomic Mortality-risk Age (TraMA)

Eric T. Klopach¹, Gokul Seshadri², Thalida Em Arpawong¹, Steve Cole³, Bharat Thyagarajan², Eileen M. Crimmins¹

¹Leonard Davis School of Gerontology, University of Southern California, Los Angeles, CA 90089, USA

²Department of Laboratory Medicine and Pathology, University of Minnesota, Minneapolis, MN 55455, USA

³David Geffen School of Medicine, University of California, Los Angeles, CA 90089, USA

Correspondence to: Eric T. Klopach; email: klopach@usc.edu

Keywords: biological aging, transcriptomics, mortality, accelerated aging, machine learning

Received: February 6, 2025

Accepted: June 2, 2025

Published: June 13, 2025

Copyright: © 2025 Klopach et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/) (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

ABSTRACT

Increasingly, research suggests that aging is a coordinated multi-system decline in functioning that occurs at multiple biological levels. We developed and validated a transcriptomic (RNA-based) aging measure we call Transcriptomic Mortality-risk Age (TraMA) using RNA-seq data from the 2016 Health and Retirement Study using elastic net Cox regression analyses to predict 4-year mortality hazard. In a holdout test sample, TraMA was associated with earlier mortality, more chronic conditions, poorer cognitive functioning, and more limitations in activities of daily living. TraMA was also externally validated in the Long Life Family Study and several publicly available datasets. Results suggest that TraMA is a robust, portable RNAseq-based aging measure that is comparable, but independent from past biological aging measures (e.g., GrimAge). TraMA is likely to be of particular value to researchers interested in understanding the biological processes underlying health and aging, and for social, psychological, epidemiological, and demographic studies of health and aging.

INTRODUCTION

A growing body of research suggests that aging is a coordinated multi-system decline in functioning that occurs at multiple biological levels (e.g., DNA damage accumulation, cellular aging and senescence, chronic disease morbidity, physical disability) [1, 2]. A major goal of geroscience research is to develop biomarkers of this aging process using minimally invasive methods in humans, as these markers are highly useful in evaluating interventions, understanding social inequalities in health and aging, and researching causes and consequences of accelerated aging in humans [3, 4]. Biomarkers of aging have been developed using combinations of clinical biomarkers [5, 6], DNA methylation (DNAm) [3, 7–10], inflammatory markers [11], telomere length [12], metabolomics [13, 14], and proteomics [15, 16]. These tools have been extremely

useful for understanding how social and environmental exposures affect health and aging [17–23], the long-term impact of early life adversity [24–27], how timing of exposure matters for health [28–30], among other important advances.

RNA gene expression may be a particularly valuable tool, as RNA expression is more directly related to genes and gene functioning, compared to DNAm, and may therefore be more easily interpretable [31]. DNAm largely describes what gene may or may not be transcribed; whereas, transcriptomics more directly measures active gene expression [32]. Additionally, research suggests that RNA changes may occur more rapidly than DNAm changes and may capture short-term and long-term responses not captured in DNAm [33]. Thus, RNA- and DNAm-based aging measures may be complementary in studying aging processes.

Previous transcriptomic (RNA-based) aging measures [34] were generally developed using array data (rather than RNA sequencing, which predominate in newer studies), have utilized small, specialty samples, or were estimated in tissue other than blood [34–36]. Indeed, a recent review noted the limitations of existing transcriptomic aging measures and the large number of unknowns about their reproducibility and ability to capture health and mortality risk [31]. At the same time, there has been a proliferation of research utilizing next-generation high throughput RNA sequencing (RNAseq), and several large population-based surveys (e.g., the Health and Retirement Study (HRS), the National Longitudinal Study of Adolescent to Adult Health (Add Health), Midlife in the United States (MIDUS), the Northern Ireland Cohort for the Longitudinal Study of Ageing (NICOLA)) are collecting large RNAseq samples that will be able to address questions about the causes and consequences of transcriptomic aging at the population level.

For these analyses to yield useful generalizable findings, a reliable and portable summary measure of accelerated transcriptomic aging is needed. We developed such a measure here using the 2016 HRS Venous Blood Study (VBS), a nationally representative sample of nearly 4000 US adults aged 50 and older. We utilized elastic net penalized regression to estimate a transcriptomic prediction measure of 4-year mortality risk—Transcriptomic Mortality-risk Age (TraMA)—using more than 10,000 gene transcripts in a training sub-sample. We evaluated this measure in a hold-out testing sub-sample of the HRS, in an external dataset (the Long Life Family Study; LLFS), and in several other publicly available datasets. Our plan of analysis for this study is shown in Figure 1A.

RESULTS

Training TraMA

Because we are interested in developing a measure that is accurate and portable to other human datasets, we restricted the set of genes used for training to coding genes with relatively high expression in human venous blood. To accomplish this, of the 50,611 transcripts that were measured and were successfully mapped in HRS, we restricted ourselves to the 19,291 protein coding genes. We further restricted ourselves to genes with a mean count per million greater than 3 in the total HRS sample, leaving 10,964 genes. We also included chronological age in years and sex/gender as features to reduce age and sex/gender bias in the selection of genes (i.e., they are covariates along with the 10,964 genes used to train TraMA).

We then randomly split the HRS sample into training ($N=1801$) and testing ($N=1794$) subsamples. Descriptive statistics for each subsample are shown in Table 1A and 1B. 222 participants in the training data died during the 4-year follow-up. 197 participants in the testing data died during follow-up. We ran elastic net models predicting 4-year mortality in the training sample. Hyperparameters, including the alpha and lambda penalty terms, were selected using a grid search procedure with 5x cross validation. Log₂ adjusted counts per million (log₂cpm) were used in all analyses, including the Health and Retirement Study (HRS), the Long Life Family Study (LLFS), and publicly available data described below. This procedure selected an alpha of 1 (equivalent to LASSO regression) and a lambda of 0.0198. This model selected (i.e., did not reduce regression coefficients to 0) 35 genes and age. Gene names, Ensembl IDs, and coefficients are shown in Table 2. TraMA scores were transformed to have a mean and variance equivalent to the HRS training set chronological age. In the training data, Harrel's C-index predicting survival from the TraMA score was 0.835, suggesting very good fit.

Gene ontologies, associated traits, pathway analysis, and functional enrichment analysis

Gene ontologies

A major value of biological aging measures is that they describe biological aging pathways, marking processes that underly health and aging that may not be phenotypic yet. That is, these measures help assess pre-diagnostic states before morbidities and mortalities manifest. To assess how well TraMA indexes these pathways, we assessed ontologies provided by Ensembl [37] (shown in Supplementary Table 1). According to these ontologies, there is evidence several of these genes are involved in neurological development and functioning (*CNTNAP2*, *KCNA2*, *KIFBP*, *NELL2*, *NOG*), amyloid formation and regulation (*ADAM17*, *APH1B*), immune responses (*ADAM17*, *CLEC4C*, *GPR15*, *MCOLN2*), cell cycle regulation (*CDKN2B*, *TRIM39*), and methylation and gene expression regulation (*METTL9*, *ZNF417*, *ZNF44*). These ontologies, thus, include pathways essential to aging and health.

Associated traits

We assessed associated traits from past transcriptome-wide association studies (TWAS) for the 35 genes using TWAS logged in the TWAS Atlas [38] (shown in Supplementary Table 2). A large number of these genes have been linked in past TWAS to body height (*ABTB3*, *ADAM17*, *ANGPT1*, *DSP*, *HDGFL3*, *KIFBP*, *LASP1B*, *NKD1*, *PLVAP*, *TMEM38A*, *ZNF417*, *ZNF44*), weight and BMI (*ABTB3*, *ADAM17*, *ANGPT1*, *C12orf76*, *HDGFL3*, *KIFBP*, *NOG*, *PLVAP*, *SLC4A10*),

blood pressure and hypertension (*C12orf76*, *CTTNBP2NL*, *SLC16A1*, *SLC4A10*, *TRIM39*), lung functioning (*DSP*, *LASP1NB*) and to chronological age

(*ADAM17*, *CDKN2B*, *METTL9*). Thus, these genes have been associated with a number of age and development-related traits in past research.

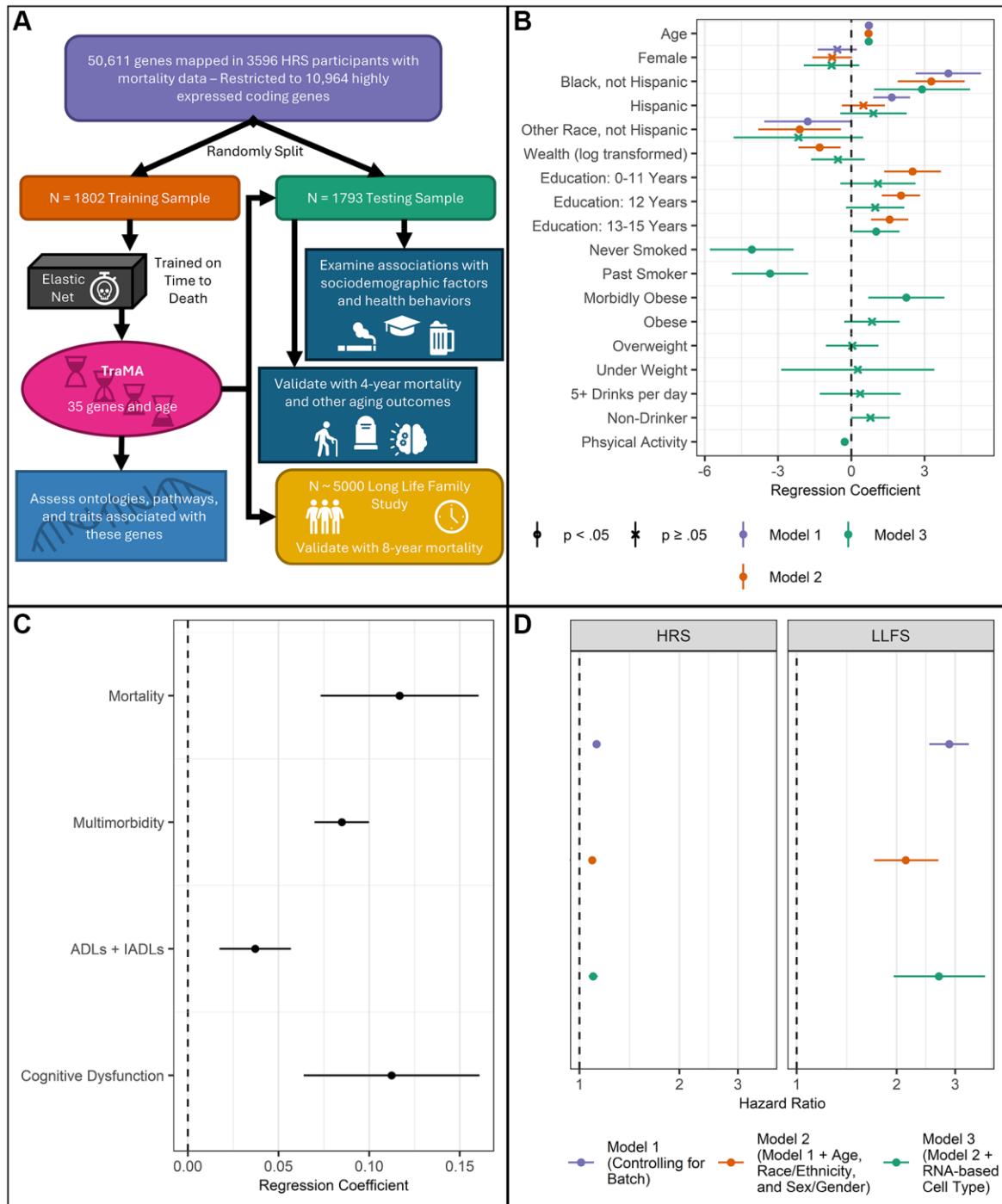


Figure 1. (A) Plan of analysis for the current study. (B) Nested regression results from the HRS testing data including associations between TraMA and sociodemographic factors and health behaviors; points represent regression coefficients and bars represent 95% confidence intervals; all models include cell type and batch as covariates. Model 1 includes demographic factors; Model 2 includes variables in Model 1, as well as socioeconomic factors; Model 3 includes variables in Model 2, as well as health behaviors. (C) Regression results from the HRS testing data of health/aging outcomes on TraMA; points represent regression coefficients and bars represent 95% confidence intervals; all models include age, race/ethnicity, sex/gender, cell type, and batch as covariates. (D) Validation results from nested regression of time to death on TraMA in HRS and LLFS. Model 1 includes batch as a covariate; Model 2 includes batch, age, race/ethnicity, and sex/gender as covariates; Model 3 includes variables from Model 2, as well as RNA-based cell type as covariates.

Table 1A. Weighted descriptive statistics for the HRS training and testing data.

	Training data		Testing data	
	Mean/Proportion	SD	Mean/Proportion	SD
Age	68.62	9.21	68.66	9.09
Race: Non-Hispanic White	0.78		0.78	
Race: Non-Hispanic Black	0.10		0.11	
Race: Hispanic	0.08		0.09	
Race: Non-Hispanic Other Race	0.04		0.02	
Gender: Female	0.55		0.54	
Mortality	0.11		0.10	
Multimorbidity	2.12	1.39	2.04	1.37
Cognitive Dysfunction	11.56	4.31	11.31	4.35
ADLs + IADLs	0.72	1.74	0.64	1.60

Table 1B. Descriptive statistics for the LLFS dataset.

	Mean/Proportion	SD
Age	70.04	15.6
Subset: Proband	0.32	
Subset: Offspring	0.50	
Subset: Spouse	0.18	
Gender: Female	0.54	
Mortality	0.22	

Table 2. Genes (and age) and their coefficients in the TraMA score.

Gene	Ensembl ID	Coefficient (from 4-year mortality hazard)
<i>ZNF44</i>	ENSG00000197857	-0.2630
<i>CRYBG3</i>	ENSG00000080200	-0.1964
<i>NOG</i>	ENSG00000183691	-0.1856
<i>ABTB3</i>	ENSG00000151136	-0.1307
<i>NELL2</i>	ENSG00000184613	-0.1149
<i>ZNF417</i>	ENSG00000173480	-0.0844
<i>CLEC4C</i>	ENSG00000198178	-0.0821
<i>PMEPA1</i>	ENSG00000124225	-0.0768
<i>TRIM39</i>	ENSG00000204599	-0.0591
<i>SLC4A10</i>	ENSG00000144290	-0.0579
<i>CNTNAP2</i>	ENSG00000174469	-0.0484
<i>NKD1</i>	ENSG00000140807	-0.0272
<i>DSP</i>	ENSG00000096696	-0.0122
<i>KIFBP</i>	ENSG00000198954	-0.0120
<i>ANGPT1</i>	ENSG00000154188	-0.0080
<i>ADGRA3</i>	ENSG00000152990	-0.0069
<i>PLVAP</i>	ENSG00000130300	-0.0067
<i>MCOLN2</i>	ENSG00000153898	0.0017
<i>CTTNBP2NL</i>	ENSG00000143079	0.0191
<i>LASP1NB</i>	ENSG00000263874	0.0222
<i>SLC16A1</i>	ENSG00000155380	0.0251
<i>NBPF3</i>	ENSG00000142794	0.0303

<i>KCNA2</i>	ENSG00000177301	0.0402
<i>EFCAB2</i>	ENSG00000203666	0.0531
Chronological Age		0.0542
<i>CDKN2B</i>	ENSG00000147883	0.0560
<i>TMEM38A</i>	ENSG00000072954	0.0576
<i>CI2orf76</i>	ENSG00000174456	0.0697
<i>HDGFL3</i>	ENSG00000166503	0.0894
<i>RRAGB</i>	ENSG00000083750	0.0980
<i>GPR15</i>	ENSG00000154165	0.1157
<i>LONRF3</i>	ENSG00000175556	0.1495
<i>MARCHF6</i>	ENSG00000145495	0.1502
<i>ADAM17</i>	ENSG00000151694	0.1751
<i>APH1B</i>	ENSG00000138613	0.2858
<i>METTL9</i>	ENSG00000197006	0.5401

Pathway and functional enrichment analysis

We assessed gene pathways using the GeneMANIA prediction server [39]. This program takes a list of genes and identifies other genes involved in genetic interactions, pathways, and co-expression. These networks are shown in Supplementary Figure 1 with purple lines indicating co-expression, red lines physical interactions, and green lines genetic interactions.

We also performed functional enrichment analysis using GeneMANIA. This program identifies gene ontology terms enriched among the list of genes identified in pathways and provides FDR corrected Q values and coverage ratios [39]. The top seven functions (based on lowest FDR corrected Q values) are shown in Supplementary Figure 1 and all functions identified are shown in Supplementary Table 3. A large number of these functions involve the renal system, including nephron development, glomerulus development, kidney vasculature development, kidney development, renal system vasculature development, and renal system development. A number are also related to basic cell functioning (cell adhesion mediator activity, regulation of transmembrane receptor protein serine/threonine kinase signaling pathway, transmembrane receptor protein serine/threonine kinase signaling pathway) and cell cycle regulation (regulation of pathway-restricted SMAD protein phosphorylation, pathway-restricted SMAD protein phosphorylation). Other functions include nervous system (main axon, neuron recognition) and other biological system functioning. These functions make sense given the essential functions of the kidney and nervous system in aging and mortality and given the importance of cell cycle regulation for cellular aging.

We also entered these genes into the GTEx portal (<https://www.gtexportal.org/home/>) to examine how much these genes are expressed in different tissues (see

Supplementary Figure 2). All of the genes are expressed at least somewhat in whole blood, though some levels are relatively low. A subset of genes is very highly expressed in brain tissues (viz., *BTBD11*, *LASP1NB* (*LINC00672* in the GTEx figure), *NELL2*, *NOG*, *CNTNAP2*, *KCNA2*, and *SLC4A10*) and therefore may be particularly important for brain aging. *METTL9* and *MARCHF6* were highly expressed in all tissues. *GPR15* and *MCOLN2* were highly expressed in lymphocytes.

Inflammation and renal functioning are both implicated by this enrichment analysis. We found that age accelerated TraMA score (i.e., age is regressed out from the score) is modestly associated with several inflammatory markers (viz., IL-6, GDF15, IL-1RA, IL-10, CRP, and albumen; $r = 0.10, 0.27, 0.20, 0.22, 0.35, -0.28$, respectively). Age accelerated TraMA was modestly correlated three blood-based markers of renal function (viz., BUN, creatinine, and cystatin-C; $r = 0.10, 0.23$, and 0.13 , respectively).

Validation in the health and retirement study testing sample

Validation with time to death

We tested this surrogate score in the $N = 1794$ HRS hold-out testing sample. TraMA was significantly associated with mortality hazard in the testing subsample with age, sex/gender, race/ethnicity, and batch as covariates (HR = 1.09, 95% CI = (1.06, 1.12), $p < 0.0001$). Thus, having a TraMA score 10 years older (a little more than a standard deviation), was associated with about a 90% increase in mortality hazard. Harrel's C index (an extension of area under the curve for survival data) was 0.81 indicating excellent fit. TraMA appears to be a robust predictor of mortality, as even a relatively young person with a high TraMA age is predicted to have a low probability of surviving 4 years (see Supplementary Figure 1 and Supplementary Figure 3C).

Associations with sociodemographic factors and health behaviors

Researchers using measures of biological aging (e.g., epigenetic clocks, telomeres) are often interested in understanding how psychosocial, demographic, and behavioral risk factors contribute to differences in mortality and other health and aging outcomes. To indicate the utility of TraMA to these researchers we conducted a series of nested regressions, first regressing TraMA on basic demographic factors (viz., age, race/ethnicity, and sex/gender), then on socioeconomic factors thought to contribute to demographic differences in health (viz., wealth and education), and finally on health behaviors thought to mediate these socio-demographic associations with health (viz., smoking status, BMI (as a proxy for diet and activity), alcohol use, and a physical activity index. All regressions included RNA-based cell type distribution in whole blood (using log₂cpm for *CD3D*, *CD19*, *CD4*, *CD8A*, *FCGR3A*, *NCAMI*, and *CD14*) and batch as covariates. Results are shown in Figure 1B.

In all models, chronological age is significantly associated with TraMA, which is expected as age is used in the calculation of TraMA. Compared to non-Hispanic White respondents, non-Hispanic Black respondents had a significantly higher TraMA in all models; Hispanic participants had an older TraMA without controlling for socioeconomic status or health behaviors; and non-Hispanic participants from other racial groups had a lower TraMA in models without health behaviors as covariates. Before controlling for health behaviors, greater wealth and higher educational attainment were associated with younger TraMA. After including health behaviors as covariates, having 13–15 years of education (vs. 16 or more) was associated with older TraMA. Compared to current smokers, never smokers and past smokers had younger TraMA. Compared to normal weight individuals, morbidly obese individuals (BMI ≥ 35) had older TraMA. Greater physical activity was associated with younger TraMA.

Associations with mortality and other health/aging outcomes

To assess TraMA as a general measure of aging, we also regressed multimorbidity (count of diagnoses with high blood pressure, diabetes, cancer, lung disease, heart disease, stroke, and arthritis), 4-year mortality, cognitive dysfunction (using errors on the Telephone Interview for Cognitive Status (TICS)), and count of at least some difficulties with activities of daily living (ADLs) and instrumental activities of daily living (IADLs), controlling for chronological age, race/ethnicity, sex/gender, RNA-based cell type, and batch effects. Regression coefficients are shown in Figure 1C.

TraMA was significantly associated with each of these outcomes in the HRS testing data.

As a sensitivity analysis, we also estimated the equations shown in Figure 1C using cell percentages of neutrophils, eosinophils, basophils, monocytes, CD4+ T cells, and B cell measured by flow cytometry. The results were extremely similar with an identical pattern of significant results. We retain RNA-based cell type control in the main analysis to ease comparison in samples without flow cytometry data. We argue that intrinsic aging (removing associations with cell type) is highly useful, as it is generally consistent across tissues and theoretically captures the underlying aging process independent of tissue composition. However, extrinsic aging (not adjusted for cell type) is also of great value. It is often a better predictor mortality, morbidity, and frailty. It is also an overall measure of aging, including tissue composition changes. This general aging measure is often of greater interest to social and behavioral researchers who are less interested in the specifics of cellular aging and/or geroscience-related processes.

Comparison with other biological aging measures

As noted above, a large number of biological aging measures have been produced using omics data, telomeres, and indices of blood-based biomarkers. To be a useful and innovative measure of aging, TraMA should (1) be associated with health outcomes to a similar or greater magnitude compared to these measures and (2) should be associated with health outcomes above and beyond these measures.

To assess this first point we regressed health and aging outcomes on TraMA, GrimAge (an epigenetic aging measure), PhenoAge (an epigenetic aging measure), and ExpandedAge (an index of biomarkers associated with aging) controlling for age, race/ethnicity, sex/gender, RNA-based cell type, and batch, shown in Supplementary Figure 3A. Each point in this panel represents results from a separate regression. Associations between TraMA and health outcomes are slightly stronger than those of PhenoAge or ExpandedAge. Association between TraMA and mortality and between TraMA and cognitive dysfunction are slightly weaker than similar associations with GrimAge. TraMA has a very similar, but slightly higher, association with multimorbidity compared to GrimAge. TraMA, PhenoAge, and ExpandedAge, but not GrimAge, were significantly associated with ADLs and IADLs. Thus, the pattern of significant results was highly consistent across these aging measures.

Given these similarities, one may wonder whether TraMA explains unique variance in health outcomes or

if it simply duplicates other existing aging measures. To address this issue, we regressed each health outcome on TraMA and GrimAge in the same model, along with age, race/ethnicity, sex/gender, RNA-based cell type, and batch, shown in Supplementary Figure 3B (correlations among these aging measures are shown in Supplementary Table 4). Both TraMA and GrimAge were significantly associated with mortality at very similar magnitudes. When in the same model together, TraMA, but not GrimAge, was associated with multimorbidity and ADLs and IADLs. When in the same model together, GrimAge, but not TraMA, was associated with cognitive dysfunction, though the association with TraMA approached significance ($p = 0.066$). Thus, TraMA appears to mostly describe unique variance in mortality, multimorbidity, and ADLs and IADLs compared to GrimAge.

Validation in the Long Life Family Study

To ensure portability of this measure, and as a further check for robustness, we also validated this measure in an external cohort that includes a large number of older adult humans, the LLFS. Using mixed effect Cox proportional hazards models, we regressed time to death in each sample on TraMA, controlling for batch (Model 1); adding age, race/ethnicity (in HRS; because nearly all LLFS participants were White, we did not control for Race/Ethnicity in LLFS), and sex/gender (Model 2); and adding RNA-based cell type (Model 3). All LLFS models were adjusted for family relatedness as a fixed effect. Results are shown in Figure 1D. TraMA was significantly associated with time to death in both LLFS and HRS with all controls included.

The hazard ratio for LLFS is higher than for HRS. This potential because LLFS participants were selected for their longevity. LLFS also had twice the number of mortality events as HRS. Because the LLFS includes a sample of older adults, their children, and spousal controls, we additionally ran analyses only in the sample of older adults (referred to in the LLFS as the proband generation). These results largely replicated those shown in Model 3 above, with a hazard ratio of 2.48 and p -value less than 0.001 for the proband generation, 4.76 and p -value less than 0.01 for the offspring generation, and 1.63 and p -value less than 0.01 for the spousal controls.

Finally, LLFS has also recently processed RNA data from visit 2 ($N = 1,263$). We additionally regressed mortality hazard in 2020 on TraMA scores from time 1 and 2, with covariates from Model 3 above. When TraMA scores from both time points were entered in the model, visit 2 TraMA was associated with mortality hazard ($HR = 1.66$, $p < 0.05$) and time 1 TraMA was

not ($HR = 1.13$, $p > 0.05$), suggesting that residual change in TraMA score is associated with mortality hazard, and that longitudinal change in TraMA may track mortality risk.

Validation in small and clinical samples

To be a maximally valuable measure of the aging process, TraMA should be useful not only in large representative samples, but also in small specialty and clinical samples. We therefore additionally validated four publicly available datasets from the Gene Expression Omnibus (GEO) with RNA-seq data from whole blood and information about chronological age. First, to validate expected associations with health behaviors, we estimated TraMA in data from 454 current and 767 former smokers from the COPD Gene Study (GEO series GSE171730), including non-Hispanic White and African American people between the ages of 45 and 80 in the US (shown in Figure 2A) [40, 41]. In these data, TraMA was positively associated with smoker status and number of smoking pack years, and negatively associated with lung functioning assessed with forced expiratory volume in one second (FEV1) predicted percentage. Current smokers had an estimated TraMA 2.83 years older than former smokers controlling for age, race, sex/gender, RNA-based cell type, and batch.

TraMA was also associated with diabetes status in a sample of 43 healthy participants and 39 participants with type 1 diabetes (GEO series GSE123658; results shown in Figure 2B) [42]. In these data, controlling for age, sex/gender, RNA-based cell type, and batch, participants with diabetes had a predicted TraMA 2.12 years older than healthy controls.

We also validated in data from the Mount Sinai Crohn's and Colitis Registry (MSCCR; GEO series GSE186507) [43] including 821 participants with inflammatory bowel disease (IBD; 432 with Crohn's disease and 389 with ulcerative colitis) and 209 healthy controls (results shown in Figure 2C). IBD status was not significantly associated with TraMA after statistically controlling for age, sex/gender, and RNA-based cell type; however, this association was significant without RNA-based cell type (log odds = 0.07, $p < 0.001$), and Crohn's disease participants had elevated TraMA compared to healthy controls with all covariates ($b = 1.00$, $p < 0.05$), this may be because Crohn's disease patients were more likely to have active IBD (Harvey-Bradshaw index (HBI) ≥ 5 ; $p < 0.01$). Among all IBD patients, participants with clinically active IBD (according to a physician's evaluation) had higher TraMA, and TraMA was positively associated with IBD severity using the Simple Endoscopic Score for Crohn's Disease (SESCD).

TraMA was associated with sepsis, compared to healthy controls, in a sample of 348 sepsis patients and 44 healthy controls (GEO series GSE171730; results shown in Figure 2D) [44] in a regression including age,

sex/gender, and RNA-based cell type as covariates. Compared to healthy controls, sepsis patients had an estimated TraMA 10.31 years older. TraMA was also associated with Sequential Organ Failure Assessment

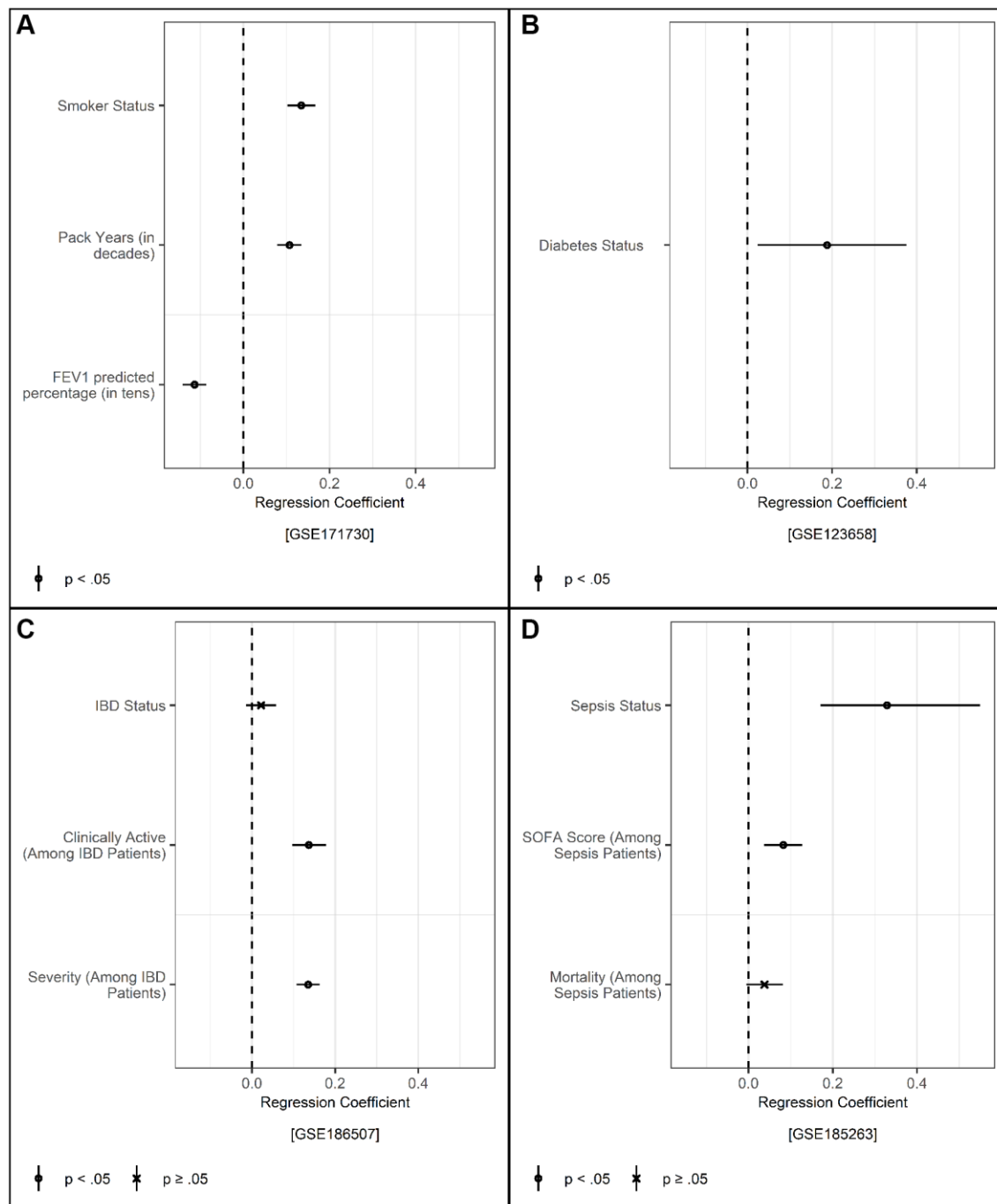


Figure 2. (A) Regressions of smoker status, cigarette pack years (divided by 10), and forced expiratory volume over one second (FEV1) predicted percentage (divided by ten) on TraMA controlling for age, race, sex/gender, cell type, and batch; points represent regression coefficients and bars represent 95% confidence intervals. (B) Regression of diabetes status on TraMA controlling for age, sex/gender, and cell type; points represent regression coefficients and bars represent 95% confidence intervals. (C) Regressions of inflammatory bowel disease (IBD) status, clinically active IBD, and IBD severity from endoscopy on TraMA controlling for age, sex/gender, and cell type; points represent regression coefficients and bars represent 95% confidence intervals. (D) Regressions of sepsis status, sequential organ failure assessment (SOFA) score, and mortality among sepsis patients on TraMA controlling for age, sex/gender, and cell type; points represent regression coefficients and bars represent 95% confidence intervals.

(SOFA) scores among sepsis patients. The association between TraMA and mortality among sepsis patients approached significance ($p = 0.09$), though this association was significant in a model without cell type (log odds = 0.05, $p < 0.001$).

Validation with other platforms and model species

A truly portable measure of transcriptomic aging would also be valid in human data using other RNA measurement platforms (e.g., array data) and in model species. To that end we also estimated TraMA in GEO data using an Affymetrix array, an Illumina array, and in a mouse (*Mus musculus*) sample. Because expression values from arrays and from RNAseq have different distributions, we do not expect the means and variances of TraMA calculated from arrays to be meaningful. We therefore use standardized scores in all of these analyses.

We begin by analyzing blood samples from 1,013 human cancer patients and 1,832 control samples with RNA profiled on an Affymetrix array (GEO series GSE203024; results shown in Figure 3A). Participants with cancer had higher TraMA scores compared to those without cancer controlling for age, sex, and RNA-based cell type. TraMA was also associated with lupus diagnosis in a sample of 134 juvenile patients and 36 healthy controls, statistically controlling for age, sex, race, RNA-based cell type, and batch (GEO series GSE65391 [45]; results shown in Figure 3B). Thus, TraMA appears to still perform robustly in a juvenile sample.

Finally, we estimated TraMA in a sample of mice exposed to 1.5, 3, 6 or 10 Gy of gamma-rays or sham irradiated controls and sacrificed at either 1, 2, 3, 5 or 7 days after exposure (GEO series GSE124612) [46]. Greater levels of radiation were associated with more TraMA aging controlling for number of days after exposure, but this pattern largely disappeared after controlling for RNA-based cell type (mice exposed to 1.5 Gy were slightly lower on TraMA compared to controls and no other differences were significant, see Figure 3C). However, there appears to be a joint effect of time and radiation (Figure 3D). Mice exposed to 1.5 Gy were mostly not significantly different on TraMA compared to controls sacrificed on all days, except on day 5 they were slightly lower. Mice exposed to 3 Gy experienced a spike in TraMA age on days 2 and 3 that appears to dissipate by day 5, and they are slightly lower than controls on day 7. Mice exposed to 6 Gy showed a similar pattern of a spike on days 2 and 3 that dissipates. Mice exposed to 10 Gy, showed higher TraMA age on all days after day 1. We also generated volcano plots from regressions of each gene in all of the human GEO datasets on age (see Supplementary Figure 4).

DISCUSSION

Large, population-based studies of aging are collecting omic-level biological data, creating a unique and exciting opportunity to understand both population-level and potentially individual-level biological aging processes. Measures of aging have been developed using DNAm and sets of clinical and blood-based biomarkers, and these measures have rapidly advanced research on aging and health. However, a similar measure does not exist for RNAseq data. This is a major gap in past research, as RNA represents a critical step in gene expression and, ultimately, nearly all biological processes. To address this gap, we developed Transcriptomic Mortality-risk Age (TraMA) using RNAseq data in the HRS and validated this measure in the LLFS and in several publicly available datasets.

We used an elastic net approach to identify genes associated with 4-year all-cause mortality. This method is in line with so-called second-generation DNAm-based epigenetic clocks (e.g., PhenoAge [3], GrimAge [9]) that focus on phenotypic indicators of aging as criterion variables. These second-generation clocks have been found to be much more strongly associated with both health outcomes and health risk exposures compared with first-generation clocks that used chronological age alone as a criterion variable [21, 25].

Thus, TraMA is in line with similar DNAm-based measures that have been consistently associated with health outcomes, health risk exposures, and socio-demographic factors [47, 48]. Indeed, analyses here indicate that TraMA is similarly associated with health outcomes, including mortality, multimorbidity, ADLs and IADLs, and cognitive functioning. However, these analyses also find TraMA captures unique variance in age-related health outcomes, compared to GrimAge.

Genes selected by this elastic net procedure represented a number of developmental and health processes we would expect to be associated with aging and mortality (e.g., immune response, cell cycle regulation, gene expression regulation, body weight, blood pressure, and chronological age). Thus, this measure captures biologically plausible cellular and multi-system processes involved in aging. Many of these genes are associated with height, weight, and BMI, suggesting that TraMA indexes general physical development. However, TraMA's associations with aging related health outcomes are generally unchanged from the values shown in Figure 3, with identical p -value thresholds, after additionally controlling for BMI (results not shown).

This study is not without limitations. The HRS is representative of the older US population, but the aging

process likely differs across national and cultural context. Because we were interested in assessing mortality risk as an indicator of aging in older adults, we utilized data from the HRS, where participants were all aged 50 or older. However, we validated this

measure in samples that included relatively young participants (as young as 19).

Blood-biomarker-based and DNAm-based aging measures have rapidly accelerated aging research. We believe our

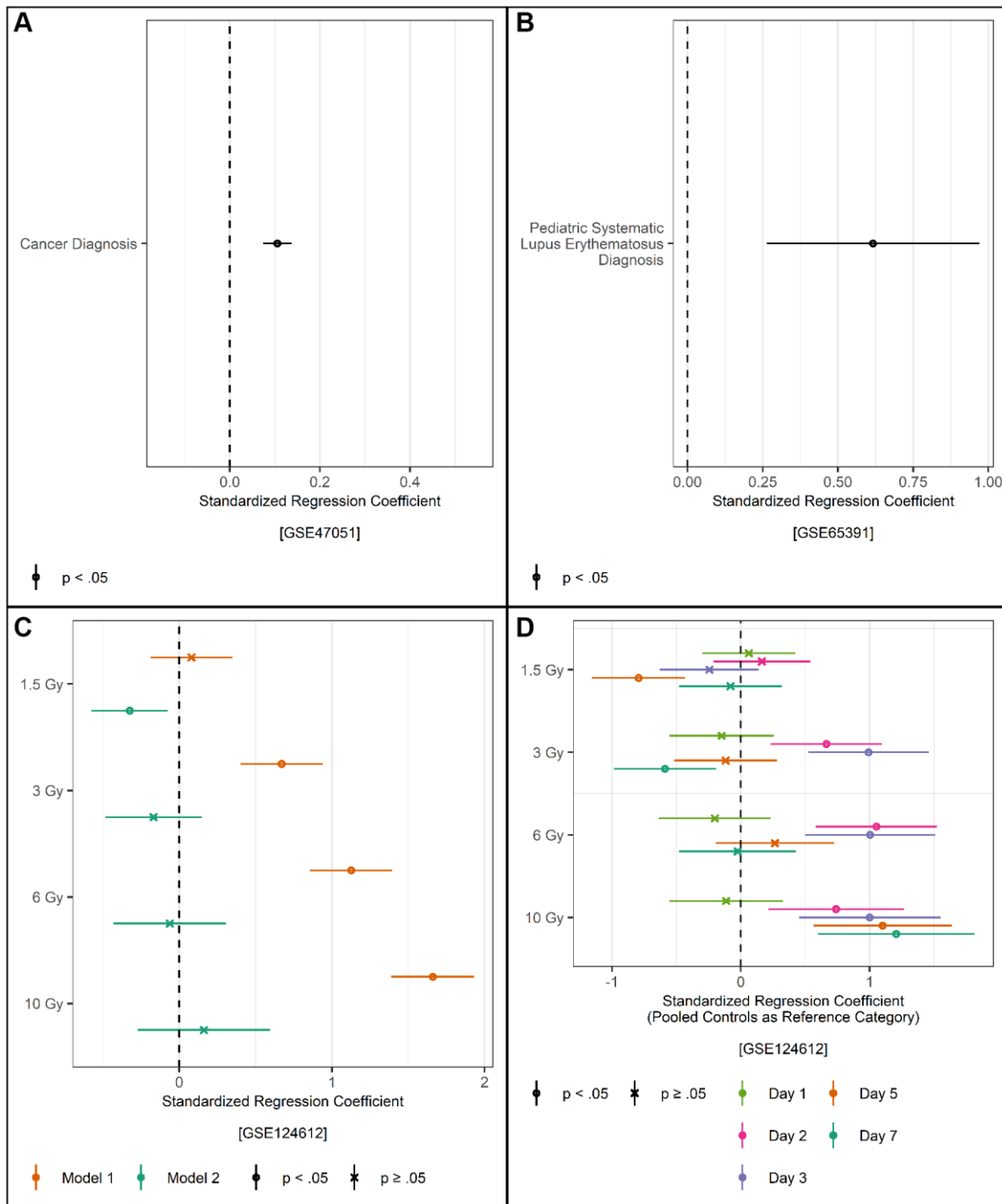


Figure 3. (A) Regression of standardized TraMA on cancer diagnosis controlling for age, sex/gender, and cell type; points represent regression coefficients and bars represent 95% confidence intervals. (B) Regression of standardized TraMA on systematic lupus erythematosus diagnosis controlling for age, sex/gender, race/ethnicity, batch, and cell type in a pediatric sample; points represent regression coefficients and bars represent 95% confidence intervals. (C) Regression of standardized TraMA on irradiation level in a sample of *Mus musculus*; points represent regression coefficients and bars represent 95% confidence intervals; Model 1 controls for number of days since exposure, Model 2 controls for number days since exposure and cell type. (D) Regression of standardized TraMA on irradiation level and number of days since exposure controlling for cell type in a sample of *Mus musculus*.

RNA-based measure has the capacity to contribute to this highly active and quickly evolving literature. Associations between TraMA and health outcomes were robust and consistent in the HRS testing sample, the LLFS, and other validation samples. Thus, this measure appears to be a useful, portable indicator of the aging process. It appears to explain a large, unique portion of aging-related health outcomes and is associated with health risks in expected directions. Our results show its utility in both large, population-based samples, and smaller clinical, specialty, and community-based samples. We, thus, believe this measure can be a useful tool for researchers interested in understanding the aging process in humans.

METHODS

Cohorts

The Health and Retirement Study (HRS) is an ongoing panel study of older adults since 1992 that is designed to be representative of older US adults when weighted. As part of 2016 data collection, venous blood was collected from a subsample of the HRS. 2.5 ml of blood was collected in PAXgene tubes from about 4000 participants. Total RNA extraction was performed on the QIAcube semi-automated method using the PAXgene Blood miRNA Kit. Assays used 200-500 ng of RNA for each sample. All RNA species were extracted and stored for future use. RNA was extracted from only half a PAXgene tube to ensure RNA storage in a variety of formats. Ribosomal RNA and globin reduction was performed using the TruSeq stranded Total Library Prep Gold kit - Ribozero Gold kit. RNAseq was performed on a NovaSeq (Illumina Inc.) using 50 bp paired end reads. All samples were sequenced to a minimum depth of 20 M reads. RNA-Seq was successfully performed on 3685 participants. The HRS pipeline closely mirrors the TOPMed/GTEX RNA-Seq analysis pipeline with minor modifications. More information about RNAseq pipelines is available elsewhere [49] including the HRS website (<https://hrs.isr.umich.edu/about>).

The Long Life Family Study (LLFS) is a longitudinal sample of nearly 5000 participants from 539 families that were selected because of their exceptional longevity. There have been three waves of data collected 6-8 years apart. The first and second waves of data included blood collection. We use data from the first wave to align with HRS. More information is available at the LLFS website <https://longlifefamilystudy.com/>.

RNA sequencing for Visit 1 was performed using RNA extracted from PAXgene™ Blood RNA tubes, processed with the Qiagen PreAnalytiX PAXgene Blood miRNA

Kit. Library preparation, quality control, and sequencing were carried out by the Division of Computation and Data Sciences at Washington University, using the nf-core/rnaseq 3.14.0 pipeline for read alignment, duplicate marking, and transcript quantification. Genes with low expression (fewer than 4 counts per million in at least 98.5% of samples) and those with significant intergenic overlap were filtered out. This resulted in a final dataset of 1,810 samples and 16,418 genes. For this study, we utilized RNAseq data from the LLFS dataset, with the filtered raw counts converted to a Log₂CPM (counts per million) scale for further analysis.

The COPD Gene study (GSE171730) that is publicly available includes 454 current and 767 former smokers, including non-Hispanic White and African American men and women between the ages of 47 and 86 in the US. RNAseq was performed on whole blood using the Illumina HiSeq 2000 platform. More information is available on the COPD Gene website (<https://copdgene.org/>). Information about current smoker status, pack years, and forced expiratory volume in one second (FEV1) predicted percentage, race, sex/gender, and batch are available.

GSE123658 is a sample of 43 healthy donors and 39 type 1 diabetes patients between ages 19 and 73. RNAseq was assessed in whole blood using Illumina NextSeq 500 or HiSeq 4000 platforms. Information about diabetes status, age, and sex/gender are available.

The Mount Sinai Crohn's and Colitis Registry (GSE186507) includes 821 irritable bowel disease (IBD) patients and 209 healthy controls aged 19 to 82 recruited during an endoscopy appointment from December 2013 to September 2016. RNAseq was assessed in whole blood using the Illumina HiSeq 2500 platform. Information about IBD status, active IBD status (Harvey-Bradshaw index (HBI) ≥ 5), disease severity Simple Endoscopic Score for Crohn's Disease (SES-CD), age, and sex/gender are available.

GSE185263 is a sample of 348 sepsis patients and 44 healthy controls aged 18 to 96 from countries, including Australia, Colombia, the Netherlands, and Canada (sites in Toronto and Vancouver). RNAseq was assessed in whole blood using the Illumina HiSeq 2500 platform. Information about sepsis status, severity using Sequential Organ Failure Assessment (SOFA) scores, mortality, age, and sex/gender are available.

GSE203024 includes blood samples from 1,013 human cancer patients with 11 different types of cancer or colorectal polyps and 1,832 control samples without a cancer diagnosis with RNA profiled on Affymetrix U133 Plus 2.0 GeneChips. Expression values were log₂

transformed and missing values were set to the median. Affymetrix IDs were matched to ensembl IDs using biomaRt [37]. If an ensembl ID matched more than one probe, the mean of the values was taken. 34 of the 35 TraMA genes were available.

GSE65391 is a sample of pediatric lupus patients. Expression values were log₂ transformed and missing values were set to the median. Illumina IDs were matched to ensembl IDs using biomaRt [37]. If an ensembl ID matched more than one probe, the mean of the values was taken. 32 of the 35 TraMA genes were available.

GSE124612 is a sample of male C57BL/6 mice exposed to 1.5, 3, 6 or 10 Gy of gamma-rays or sham irradiated controls and sacrificed at either 1, 2, 3, 5 or 7 days after exposure. 10 mice were in each experimental group except for 10 Gy on day 7, which only had 8 mice. RNA was profiled using the Agilent-026655 Whole Mouse Genome Microarray. Expression values were log₂ transformed and missing values were set to the median. *Mus musculus* genes were matched to homologous *Homo sapiens* genes using biomaRt [37]. If an ensembl ID matched more than one probe, the mean of the values was taken. 15 of the 35 TraMA genes were available. Because the ages of all of the mice were equal, age was arbitrarily set 8 to calculate TraMA. Two of the genes used to indicate cell type were not available, so we only use *CD3D*, *CD19*, *CD4*, *CD8A*, and *CD14* for these analyses.

Measures

Time to death in the HRS was assessed using information about date of interview and date of death from the HRS tracker file. We use 4-year mortality with participants known to be deceased to the HRS. Time to death was calculated as the difference between the 2016 interview month and the known month of death. For participants who survived (i.e., were not known to have died), time at risk was calculated as the time between the 2016 interview month and the most recent interview month available. This measure was used in elastic net models.

Mortality in the HRS for logistic regression. We created a binary indicator of 4-year mortality used in logistic regression models.

Time to death in the LLFS. Because relatively few of the LLFS participants died in 4 years compared to HRS, we used 8-year mortality for the replication analysis using participants known to be deceased by the LLFS. Time to death was calculated as the difference between 2006 blood sample collection date and the known date of death censored on 31 December 2014 (about 8 years after Wave 1 data collection).

Multimorbidity was calculated in HRS as the sum of diseases that a participant had ever been told by a doctor that they had, including high blood pressure, diabetes, cancer, lung disease, heart disease, stroke, and arthritis.

Cognitive dysfunction was assessed in HRS using the Telephone Interview for Cognitive Status (TICS). To make this a measure of dysfunction, we used errors (27 minus the sum of these scores) in immediate recall (10 words), delayed recall (the same 10 words after about 5 minutes of other survey questions), serial 7s (participants were asked to subtract 7 from 100 and continue subtracting for 5 trials), and backwards counting from 20 (participants were asked to count backward from 20 to 10 and given 2 points for a correct first try and 1 for a correct second try).

ADLs and IADLs in HRS are the sum of self-reported difficulty with walking across a room, dressing, bathing, eating, getting in and out of bed, using the toilet, using a map, using the phone, taking medications, managing money, shopping for groceries, and preparing a meal.

Other biological aging measures used in the current study include two epigenetic aging measures produced by HRS [50], PhenoAge [3] and GrimAge [9], and Expanded Biological Age (ExpandedAge) [6]. PhenoAge and GrimAge are both so-called second-generation epigenetic clocks. They are indices of cytosine-phosphate-guanine (CpG) sites where differential methylation is associated with age-related biomarkers and mortality. These have been widely used in past research and have been of extraordinary value in advancing geroscience and in clarifying the biological processes underlying social, psychological, and demographic differences in health and aging [48, 51–53]. We use so-called principal component versions of these clocks which have been shown to be more reliable [54]. ExpandedAge is an index of 22 biomarker of phenotypic aging that has been linked to mortality and other health outcomes [6].

Demographic factors used in regression analyses include chronological age in years, sex/gender (female as the reference group), and race/ethnicity (non-Hispanic Black, Hispanic, non-Hispanic other race, and non-Hispanic White as the reference group).

Socioeconomic factors include years of education as reported to HRS (0–11 years, 12 years (the typical number of years for a high school degree in the US), 13–15 years, and 16 or more years as the reference group), as well as total wealth as calculated by RAND for HRS [55].

Health behaviors include smoker status as reported by respondents (never smoked, past smoker, and current smoker as the reference group), BMI split into five categories (underweight (less than 18.4), overweight (25 to 29.9), obese (30 to 34.9), morbidity obese (greater than or equal to 35), and normal weight (18.5 to 24.9) as the reference group), alcohol use based on number of drinks per day drinking (non-drinker, five or more drinks, and one to four drink as the reference group), and an index of physical activity (the sum of respondent reported light, moderate, and vigorous activity, each ranging from 1 (*never*) to 5 (*every day*)).

Covariates

We control for batch and plate effects using batch (with the first batch as the reference group). Because percentages of blood cells change with age and blood cell composition can affect transcription levels, we control for blood cell composition using 7 genes indicative of cell type, *CD3D*, *CD19*, *CD4*, *CD8A*, *FCGR3A*, *NCAM1*, and *CD14*.

Analytic plan

Machine learning analyses were conducted in R 4.4.0 “Puppy Cup” [56] using the tidyverse [57], glmnet [58, 59], and lubridate [60] packages. We restricted the set of genes used for training to coding genes with relatively high expression in human venous blood. Of the 50,611 genes that were measured and were successfully mapped in HRS, we restricted ourselves to the 19,291 protein coding genes and to genes with a mean count per million greater than 3 in the total HRS sample, leaving 10,964 genes.

As sensitivity analysis, we also estimated models using the full 50,611 genes in HRS and the 19,291 protein coding genes in HRS and evaluated them in the HRS testing set. These elastic net models had similar C-index in testing data (0.805 and 0.804, respectively), suggesting that the subset of sites captures similar variance. We use the reduced gene set for greatest portability.

We ran elastic net models using Cox regression to predict 4-year mortality hazard using these 10,964 genes, chronological age, and sex in a $N = 1802$ training set randomly selected from HRS [58, 59]. Mortality hazard was assessed using month of death of participants known by HRS to have died. Hyperparameters, including the alpha and lambda penalty term, were selected using 5x cross validation with a grid search procedure. 11 alpha values were tested with more values close to 0 (viz., 0.000, 0.001, 0.008, 0.027, 0.064, 0.125, 0.216, 0.343, 0.512, 0.729, 1.000). The alpha and lambda values that produced the lowest mean square error were selected. We

adjusted the mean and variance to be the same as the mean and variance of HRS age in 2016 to make this an age-like variable.

$$\text{TraMA} = -18.07571 + \text{pred.prob} \times 10.02709 \quad (1)$$

Where pred.prob is the log hazard of mortality from the Cox elastic net. We then tested this surrogate score in the $N = 1794$ testing set. If a gene was missing in a validation cohort, the mean value of that gene from the HRS sample was imputed for all samples (n.b., LLFS had all necessary genes). The R code used to train the TraMA measure, as well as code needed to reproduce this measure in other data is available on Github at <https://github.com/etklopack/TraMA>.

Validation in the HRS was conducted in R 4.4.0 “Puppy Cup” [56] using the tidyverse [57] and survey [61] packages. Validation regressions and descriptive statistics in Table 1A, 1B and Figures 1 and 2 used survey weights provided by HRS for use in the RNAseq subsample (vbsi16wgtra). The total testing sample was 1794 participants. Because some participants were missing data on individual outcome variables and covariates, for mortality analyses $N = 1791$, for multimorbidity analyses $N = 1791$, for cognitive dysfunction $N = 1791$, and for ADLs and IADLs analyses $N = 1588$.

Validation in the LLFS was conducted in R 4.4.0 “Puppy Cup” [56] using coxme and dplyr packages. The total sample size was 1920 participants belonging to visit 1 in LLFS. TraMA was calculated using the algorithm developed in HRS. Mortality was analyzed using mixed-effect Cox proportional hazards regression models with adjustment for family effects.

Validation in public datasets was conducted in R 4.4.0 “Puppy Cup” [56] using the tidyverse [57] and MASS [62] packages. For GSE171730, smoker status was analyzed using logistic regression, and pack years and FEV1 predicted percentage were assessed with linear regression. For GSE123658, diabetes status was assessed using logistic regression. For GSE186507, IBD status and active IBD status were assessed using logistic regression, and severity level was assessed using ordinal logistic regression. For GSE185263, sepsis status and mortality were assessed using logistic regression, and severity was assessed using linear regression.

Abbreviations

TraMA: Transcriptomic Measure of Aging; RNA: Ribonucleic acid; HRS: Health and Retirement Study; LLFS: Long Life Families Study; TWAS:

Transcriptome-wide Association Study; VBS: Venous Blood Study; HR: Hazard Ratio; FDR: Benjamini-Hochberg False Discovery Rate; BMI: Body Mass Index; ADLs: Activities of Daily Living; IADLs: Instrumental Activities of Daily Living; TICS: Telephone Interview for Cognitive Status; GEO: Gene Expression Omnibus; FEV1: forced expiratory volume in one second; COPD: Chronic Obstructive Pulmonary Disease; MSCCR: Mount Sinai Crohn's and Colitis Registry; IBD: inflammatory bowel disease; HBI: Harvey-Bradshaw index; SESCD: Simple Endoscopic Score for Crohn's Disease; SOFA: Sequential Organ Failure Assessment; CpG: Cytosine-phosphate-guanine; DNAm: DNA methylation.

AUTHOR CONTRIBUTIONS

Eric T. Klopak (Conceptualization; Methodology; Formal analysis; Writing - Original Draft); Gokul Seshadri (Formal analysis; Validation; Writing - Review and Editing); Thalida Em Arpawong (Resources; Data Curation; Writing - Review and Editing); Steve Cole (Resources; Writing - Review and Editing); Bharat Thyagarajan (Resources; Data Curation; Writing - Review and Editing); Eileen M. Crimmins (Resources; Data Curation; Writing - Review and Editing; Supervision; Project administration; Funding acquisition).

ACKNOWLEDGMENTS

We would like to thank the research staff and LLFS participants for their substantial contribution to the study.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest related to this study.

ETHICAL STATEMENT AND CONSENT

This research was approved by the University of Southern California Institutional Review Board, No. UP-24-00661. This study is the secondary analysis of existing data. Consent for human subjects was conducted by the original researchers for each study.

FUNDING

Research reported in this study was supported by the National Institute on Aging of the National Institutes of Health under Award Numbers U01AG058499-06, T32AG000037, R01AG060110 and P30AG017265. The content is solely the responsibility of the authors

and does not necessarily represent the official views of the National Institutes of Health.

The HRS (Health and Retirement Study) is sponsored by the National Institute on Aging (grant number NIA U01AG009740) and is conducted by the University of Michigan.

The Long Life Family Study is supported by National Institute on Aging – National Institutes of Health grants (U01-AG023746, U01-AG023712, U01-AG023749, U01-AG023755, U01-AG023744, and U19-AG063893). We would like to thank the research staff and LLFS participants for their substantial contribution to the study.

REFERENCES

1. López-Otín C, Blasco MA, Partridge L, Serrano M, Kroemer G. The hallmarks of aging. *Cell*. 2013; 153:1194–217.
<https://doi.org/10.1016/j.cell.2013.05.039>
PMID:[23746838](https://pubmed.ncbi.nlm.nih.gov/23746838/)
2. Kennedy BK, Berger SL, Brunet A, Campisi J, Cuervo AM, Epel ES, Franceschi C, Lithgow GJ, Morimoto RI, Pessin JE, Rando TA, Richardson A, Schadt EE, et al. Geroscience: linking aging to chronic disease. *Cell*. 2014; 159:709–13.
<https://doi.org/10.1016/j.cell.2014.10.039>
PMID:[25417146](https://pubmed.ncbi.nlm.nih.gov/25417146/)
3. Levine ME, Lu AT, Quach A, Chen BH, Assimes TL, Bandinelli S, Hou L, Baccarelli AA, Stewart JD, Li Y, Whitset EA, Wilson JG, Reiner AP, et al. An epigenetic biomarker of aging for lifespan and healthspan. *Aging (Albany NY)*. 2018; 10:573–91.
<https://doi.org/10.18632/aging.101414>
PMID:[29676998](https://pubmed.ncbi.nlm.nih.gov/29676998/)
4. Xia X, Wang Y, Yu Z, Chen J, Han JJ. Assessing the rate of aging to monitor aging itself. *Ageing Res Rev*. 2021; 69:101350.
<https://doi.org/10.1016/j.arr.2021.101350>
PMID:[33940202](https://pubmed.ncbi.nlm.nih.gov/33940202/)
5. Levine ME. Modeling the rate of senescence: can estimated biological age predict mortality more accurately than chronological age? *J Gerontol A Biol Sci Med Sci*. 2013; 68:667–74.
<https://doi.org/10.1093/gerona/gls233>
PMID:[23213031](https://pubmed.ncbi.nlm.nih.gov/23213031/)
6. Crimmins EM, Thyagarajan B, Kim JK, Weir D, Faul J. Quest for a summary measure of biological age: the health and retirement study. *Geroscience*. 2021; 43:395–408.
<https://doi.org/10.1007/s11357-021-00325-1>

PMID:[33544281](#)

7. Horvath S. DNA methylation age of human tissues and cell types. *Genome Biol.* 2013; 14:R115.
<https://doi.org/10.1186/gb-2013-14-10-r115>
PMID:[24138928](#)
8. Hannum G, Guinney J, Zhao L, Zhang L, Hughes G, Sada S, Klotzle B, Bibikova M, Fan JB, Gao Y, Deconde R, Chen M, Rajapakse I, et al. Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol Cell.* 2013; 49:359–67.
<https://doi.org/10.1016/j.molcel.2012.10.016>
PMID:[23177740](#)
9. Lu AT, Quach A, Wilson JG, Reiner AP, Aviv A, Raj K, Hou L, Baccarelli AA, Li Y, Stewart JD, Whitsel EA, Assimes TL, Ferrucci L, Horvath S. DNA methylation GrimAge strongly predicts lifespan and healthspan. *Aging (Albany NY).* 2019; 11:303–27.
<https://doi.org/10.18632/aging.101684>
PMID:[30669119](#)
10. Belsky DW, Caspi A, Corcoran DL, Sugden K, Poulton R, Arseneault L, Baccarelli A, Chamarti K, Gao X, Hannon E, Harrington HL, Houts R, Kothari M, et al. DunedinPACE, a DNA methylation biomarker of the pace of aging. *Elife.* 2022; 11:e73420.
<https://doi.org/10.7554/eLife.73420>
PMID:[35029144](#)
11. Sayed N, Huang Y, Nguyen K, Krejcirova-Rajaniemi Z, Grawe AP, Gao T, Tibshirani R, Hastie T, Alpert A, Cui L, Kuznetsova T, Rosenberg-Hasson Y, Ostan R, et al. An inflammatory aging clock (iAge) based on deep learning tracks multimorbidity, immunosenescence, frailty and cardiovascular aging. *Nat Aging.* 2021; 1:598–615.
<https://doi.org/10.1038/s43587-021-00082-y>
PMID:[34888528](#)
12. Shalev I, Hastings WJ. Psychosocial stress and telomere regulation. In: Miu AC, Homberg JR, Lesch KP, eds. *Genes, Brain, and Emotions: Interdisciplinary and Translational Perspectives.* Oxford University Press; 2019; 247–61.
<https://doi.org/10.1093/oso/9780198793014.003.0017>
13. Menni C, Kastenmüller G, Petersen AK, Bell JT, Psatha M, Tsai PC, Gieger C, Schulz H, Erte I, John S, Brosevan MJ, Wilson SG, Tsaprouni L, et al. Metabolomic markers reveal novel pathways of ageing and early development in human populations. *Int J Epidemiol.* 2013; 42:1111–9.
<https://doi.org/10.1093/ije/dyt094>
PMID:[23838602](#)
14. van den Akker EB, Trompet S, Barkeby Wolf JJH, Beekman M, Suchiman HED, Deelen J, Asselbergs FW, Boersma E, Cats D, Elders PM, Geleijnse JM, Ikram MA, Kloppenburg M, et al. Metabolic Age Based on the BBMRI-NL ¹H-NMR Metabolomics Repository as Biomarker of Age-related Disease. *Circ Genom Precis Med.* 2020; 13:541–7.
<https://doi.org/10.1161/CIRCGEN.119.002610>
PMID:[33079603](#)
15. Tanaka T, Biancotto A, Moaddel R, Moore AZ, Gonzalez-Freire M, Aon MA, Candia J, Zhang P, Cheung F, Fantoni G, Semba RD, Ferrucci L, and CHI consortium. Plasma proteomic signature of age in healthy humans. *Aging Cell.* 2018; 17:e12799.
<https://doi.org/10.1111/ace1.12799>
PMID:[29992704](#)
16. Lehallier B, Gate D, Schaum N, Nanasi T, Lee SE, Yousef H, Moran Losada P, Berdnik D, Keller A, Verghese J, Sathyan S, Franceschi C, Milman S, et al. Undulating changes in human plasma proteome profiles across the lifespan. *Nat Med.* 2019; 25:1843–50.
<https://doi.org/10.1038/s41591-019-0673-2>
PMID:[31806903](#)
17. Avila-Rieger J, Turney IC, Vonk JMJ, Esie P, Seblova D, Weir VR, Belsky DW, Manly JJ. Socioeconomic Status, Biological Aging, and Memory in a Diverse National Sample of Older US Men and Women. *Neurology.* 2022; 99:e2114–24.
<https://doi.org/10.1212/WNL.0000000000201032>
PMID:[36038275](#)
18. Lei MK, Beach SRH. Neighborhood disadvantage is associated with biological aging: Intervention-induced enhancement of couple functioning confers resilience. *Fam Process.* 2023; 62:818–34.
<https://doi.org/10.1111/famp.12808>
PMID:[36008918](#)
19. Simons RL, Lei MK, Klopach E, Berg M, Zhang Y, Beach SSR. Re(Setting) Epigenetic Clocks: An Important Avenue Whereby Social Conditions Become Biologically Embedded across the Life Course. *J Health Soc Behav.* 2021; 62:436–53.
<https://doi.org/10.1177/00221465211009309>
PMID:[34528488](#)
20. Farina MP, Kim JK, Crimmins EM. Racial/Ethnic Differences in Biological Aging and Their Life Course Socioeconomic Determinants: The 2016 Health and Retirement Study. *J Aging Health.* 2023; 35:209–20.
<https://doi.org/10.1177/08982643221120743>
PMID:[35984401](#)
21. McCrory C, Fiorito G, Hernandez B, Polidoro S, O'Halloran AM, Hever A, Ni Cheallaigh C, Lu AT, Horvath S, Vineis P, Kenny RA. GrimAge Outperforms Other Epigenetic Clocks in the Prediction of Age-Related Clinical Phenotypes and All-Cause Mortality. *J Gerontol A Biol Sci Med Sci.* 2021; 76:741–9.
<https://doi.org/10.1093/gerona/glaa286>

PMID:[33211845](#)

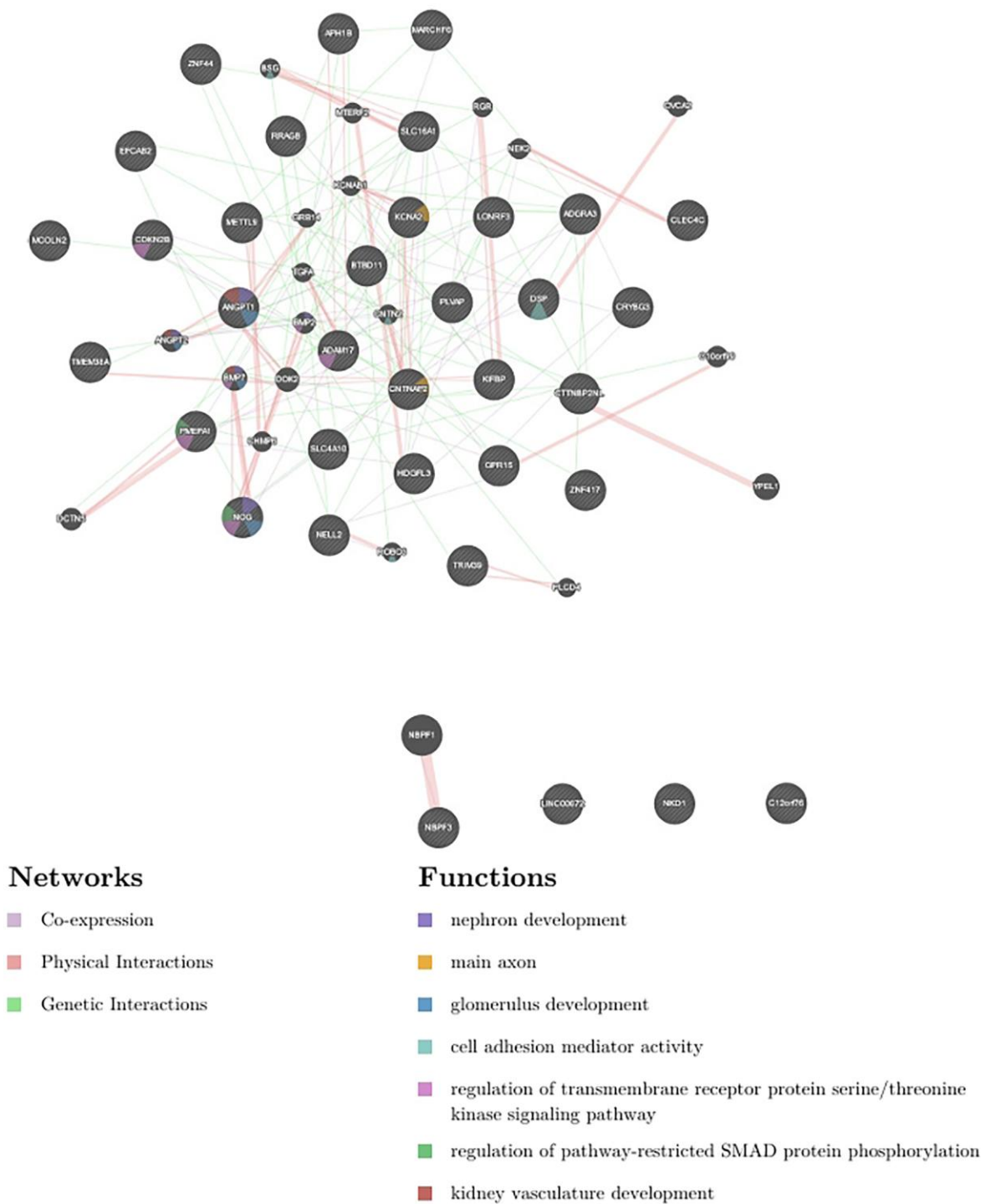
22. Courtney MG, Roberts J, Godde K. How social/environmental determinants and inflammation affect salivary telomere length among middle-older adults in the health and retirement study. *Sci Rep*. 2022; 12:8882.
<https://doi.org/10.1038/s41598-022-12742-z>
PMID:[35614304](#)
23. Farina MP, Hayward MD, Kim JK, Crimmins EM. Racial and Educational Disparities in Dementia and Dementia-Free Life Expectancy. *J Gerontol B Psychol Sci Soc Sci*. 2020; 75:e105–12.
<https://doi.org/10.1093/geronb/gbz046>
PMID:[31111926](#)
24. McCrory C, Fiorito G, O'Halloran AM, Polidoro S, Vineis P, Kenny RA. Early life adversity and age acceleration at mid-life and older ages indexed using the next-generation GrimAge and Pace of Aging epigenetic clocks. *Psychoneuroendocrinology*. 2022; 137:105643.
<https://doi.org/10.1016/j.psyneuen.2021.105643>
PMID:[34999481](#)
25. Klopach ET, Crimmins EM, Cole SW, Seeman TE, Carroll JE. Accelerated epigenetic aging mediates link between adverse childhood experiences and depressive symptoms in older adults: Results from the Health and Retirement Study. *SSM Popul Health*. 2022; 17:101071.
<https://doi.org/10.1016/j.ssmph.2022.101071>
PMID:[35313610](#)
26. Belsky DW, Caspi A, Arseneault L, Baccarelli A, Corcoran DL, Gao X, Hannon E, Harrington HL, Rasmussen LJ, Houts R, Huffman K, Kraus WE, Kwon D, et al. Quantification of the pace of biological aging in humans through a blood test, the DunedinPoAm DNA methylation algorithm. *Elife*. 2020; 9:e54870.
<https://doi.org/10.7554/eLife.54870>
PMID:[32367804](#)
27. Hamlat EJ, Prather AA, Horvath S, Belsky J, Epel ES. Early life adversity, pubertal timing, and epigenetic age acceleration in adulthood. *Dev Psychobiol*. 2021; 63:890–902.
<https://doi.org/10.1002/dev.22085>
PMID:[33423276](#)
28. Klopach ET, Carroll JE, Cole SW, Seeman TE, Crimmins EM. Lifetime exposure to smoking, epigenetic aging, and morbidity and mortality in older adults. *Clin Epigenetics*. 2022; 14:72.
<https://doi.org/10.1186/s13148-022-01286-8>
PMID:[35643537](#)
29. Bao Y, Gorrie-Stone T, Hannon E, Hughes A, Andrayas A, Neilson G, Burrage J, Mill J, Schalkwyk L, Kumari M. Social mobility across the lifecourse and DNA methylation age acceleration in adults in the UK. *Sci Rep*. 2022; 12:22284.
<https://doi.org/10.1038/s41598-022-26433-2>
PMID:[36566336](#)
30. George A, Hardy R, Castillo Fernandez J, Kelly Y, Maddock J. Life course socioeconomic position and DNA methylation age acceleration in mid-life. *J Epidemiol Community Health*. 2021; 75:1084–90.
<https://doi.org/10.1136/jech-2020-215608>
PMID:[33906906](#)
31. Rutledge J, Oh H, Wyss-Coray T. Measuring biological age using omics data. *Nat Rev Genet*. 2022; 23:715–27.
<https://doi.org/10.1038/s41576-022-00511-7>
PMID:[35715611](#)
32. Singh KP, Miaskowski C, Dhruva AA, Flowers E, Kober KM. Mechanisms and Measurement of Changes in Gene Expression. *Biol Res Nurs*. 2018; 20:369–82.
<https://doi.org/10.1177/1099800418772161>
PMID:[29706088](#)
33. Migliaccio G, Morikka J, Del Giudice G, Vaani M, Möbus L, Serra A, Federico A, Greco D. Methylation and transcriptomic profiling reveals short term and long term regulatory responses in polarized macrophages. *Comput Struct Biotechnol J*. 2024; 25:143–52.
<https://doi.org/10.1016/j.csbj.2024.08.018>
PMID:[39257962](#)
34. Peters MJ, Joehanes R, Pilling LC, Schurmann C, Conneely KN, Powell J, Reinmaa E, Sutphin GL, Zhernakova A, Schramm K, Wilson YA, Kobes S, Tukiainen T, et al, and NABEC/UKBEC Consortium. The transcriptional landscape of age in human peripheral blood. *Nat Commun*. 2015; 6:8570.
<https://doi.org/10.1038/ncomms9570>
PMID:[26490707](#)
35. Mamoshina P, Volosnikova M, Ozerov IV, Putin E, Skibina E, Cortese F, Zhavoronkov A. Machine Learning on Human Muscle Transcriptomic Data for Biomarker Discovery and Tissue-Specific Drug Target Identification. *Front Genet*. 2018; 9:242.
<https://doi.org/10.3389/fgene.2018.00242>
PMID:[30050560](#)
36. Holzschek N, Falckenhayn C, Söhle J, Kristof B, Siegner R, Werner A, Schössow J, Jürgens C, Völzke H, Wenck H, Winnefeld M, Grönniger E, Kaderali L. Modeling transcriptomic age using knowledge-primed artificial neural networks. *NPJ Aging Mech Dis*. 2021; 7:15.
<https://doi.org/10.1038/s41514-021-00068-5>
PMID:[34075044](#)
37. Durinck S, Spellman PT, Birney E, Huber W. Mapping identifiers for the integration of genomic datasets

- with the R/Bioconductor package biomaRt. *Nat Protoc.* 2009; 4:1184–91.
<https://doi.org/10.1038/nprot.2009.97>
PMID:[19617889](https://pubmed.ncbi.nlm.nih.gov/19617889/)
38. Lu M, Zhang Y, Yang F, Mai J, Gao Q, Xu X, Kang H, Hou L, Shang Y, Qain Q, Liu J, Jiang M, Zhang H, et al. TWAS Atlas: a curated knowledgebase of transcriptome-wide association studies. *Nucleic Acids Res.* 2023; 51:D1179–87.
<https://doi.org/10.1093/nar/gkac821>
PMID:[36243959](https://pubmed.ncbi.nlm.nih.gov/36243959/)
 39. Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, Franz M, Grouios C, Kazi F, Lopes CT, Maitland A, Mostafavi S, Montojo J, et al. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res.* 2010; 38:W214–20.
<https://doi.org/10.1093/nar/gkq537>
PMID:[20576703](https://pubmed.ncbi.nlm.nih.gov/20576703/)
 40. Regan EA, Hokanson JE, Murphy JR, Make B, Lynch DA, Beaty TH, Curran-Everett D, Silverman EK, Crapo JD. Genetic epidemiology of COPD (COPDGene) study design. *COPD.* 2010; 7:32–43.
<https://doi.org/10.3109/15412550903499522>
PMID:[20214461](https://pubmed.ncbi.nlm.nih.gov/20214461/)
 41. Xu Z, Platig J, Lee S, Boueiz A, Chase R, Jain D, Gregory A, Suryadevara R, Berman S, Bowler R, Hersh CP, Laederach A, Castaldi PJ, and COPDGene Investigators. Cigarette smoking-associated isoform switching and 3' UTR lengthening via alternative polyadenylation. *Genomics.* 2021; 113:4184–95.
<https://doi.org/10.1016/j.ygeno.2021.11.004>
PMID:[34763026](https://pubmed.ncbi.nlm.nih.gov/34763026/)
 42. Valentim L, Mariotti-Ferrandiz E, Klatzmann D, Six A, Konza O. Transimmunom whole blood RNA-seq data from type 1 diabetic patients and healthy volunteers. 2020.
<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE123658>
 43. Argmann C, Hou R, Ungaro RC, Irizar H, Al-Taie Z, Huang R, Kosoy R, Venkat S, Song WM, Di'Narzo AF, Losic B, Hao K, Peters L, et al. Biopsy and blood-based molecular biomarker of inflammation in IBD. *Gut.* 2023; 72:1271–87.
<https://doi.org/10.1136/gutjnl-2021-326451>
PMID:[36109152](https://pubmed.ncbi.nlm.nih.gov/36109152/)
 44. Baghela A, Pena OM, Lee AH, Baquir B, Falsafi R, An A, Farmer SW, Hurlburt A, Mondragon-Cardona A, Rivera JD, Baker A, Trahtemberg U, Shojaei M, et al. Predicting sepsis severity at first clinical presentation: The role of endotypes and mechanistic signatures. *EBioMedicine.* 2022; 75:103776.
<https://doi.org/10.1016/j.ebiom.2021.103776>
PMID:[35027333](https://pubmed.ncbi.nlm.nih.gov/35027333/)
 45. Banchereau R, Hong S, Cantarel B, Baldwin N, Baisch J, Edens M, Cepika AM, Acs P, Turner J, Anguiano E, Vinod P, Kahn S, Obermoser G, et al. Personalized Immunomonitoring Uncovers Molecular Networks that Stratify Lupus Patients. *Cell.* 2016; 165:551–65.
<https://doi.org/10.1016/j.cell.2016.03.008>
PMID:[27040498](https://pubmed.ncbi.nlm.nih.gov/27040498/)
 46. Paul S, Kleiman NJ, Amundson SA. Transcriptomic responses in mouse blood during the first week after in vivo gamma irradiation. *Sci Rep.* 2019; 9:18364.
<https://doi.org/10.1038/s41598-019-54780-0>
PMID:[31797975](https://pubmed.ncbi.nlm.nih.gov/31797975/)
 47. Faul JD, Kim JK, Levine ME, Thyagarajan B, Weir DR, Crimmins EM. Epigenetic-based age acceleration in a representative sample of older Americans: Associations with aging-related morbidity and mortality. *Proc Natl Acad Sci U S A.* 2023; 120:e2215840120.
<https://doi.org/10.1073/pnas.2215840120>
PMID:[36802439](https://pubmed.ncbi.nlm.nih.gov/36802439/)
 48. Crimmins EM, Thyagarajan B, Levine ME, Weir DR, 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 A Biol Sci Med Sci.* 2021; 76:1117–23.
<https://doi.org/10.1093/gerona/glab016>
PMID:[33453106](https://pubmed.ncbi.nlm.nih.gov/33453106/)
 49. Crimmins EM, Faul JD, Thyagarajan B, Weir DR. Venous Blood Collection and Assay Protocol in the 2016 Health and Retirement Study 2016 Venous Blood Study (VBS). 2017.
<https://hrsdata.isr.umich.edu/sites/default/files/documentation/data-descriptions/HRS2016VBSDD.pdf>
 50. Crimmins EM, Kim JK, Fisher J, Faul JD. HRS Epigenetic Clocks. University of Michigan Survey Research Center. 2020.
https://hrsdata.isr.umich.edu/sites/default/files/documentation/data-descriptions/EPICLOCKS_DD.pdf
 51. Farina MP, Klopach ET, Umberson D, Crimmins EM. The embodiment of parental death in early life through accelerated epigenetic aging: Implications for understanding how parental death before 18 shapes age-related health risk among older adults. *SSM Popul Health.* 2024; 26:101648.
<https://doi.org/10.1016/j.ssmph.2024.101648>
PMID:[38596364](https://pubmed.ncbi.nlm.nih.gov/38596364/)
 52. Schmitz LL, Zhao W, Ratliff SM, Goodwin J, Miao J, Lu Q, Guo X, Taylor KD, Ding J, Liu Y, Levine M, Smith JA. The Socioeconomic Gradient in Epigenetic Ageing Clocks: Evidence from the Multi-Ethnic Study of Atherosclerosis and the Health and Retirement Study.

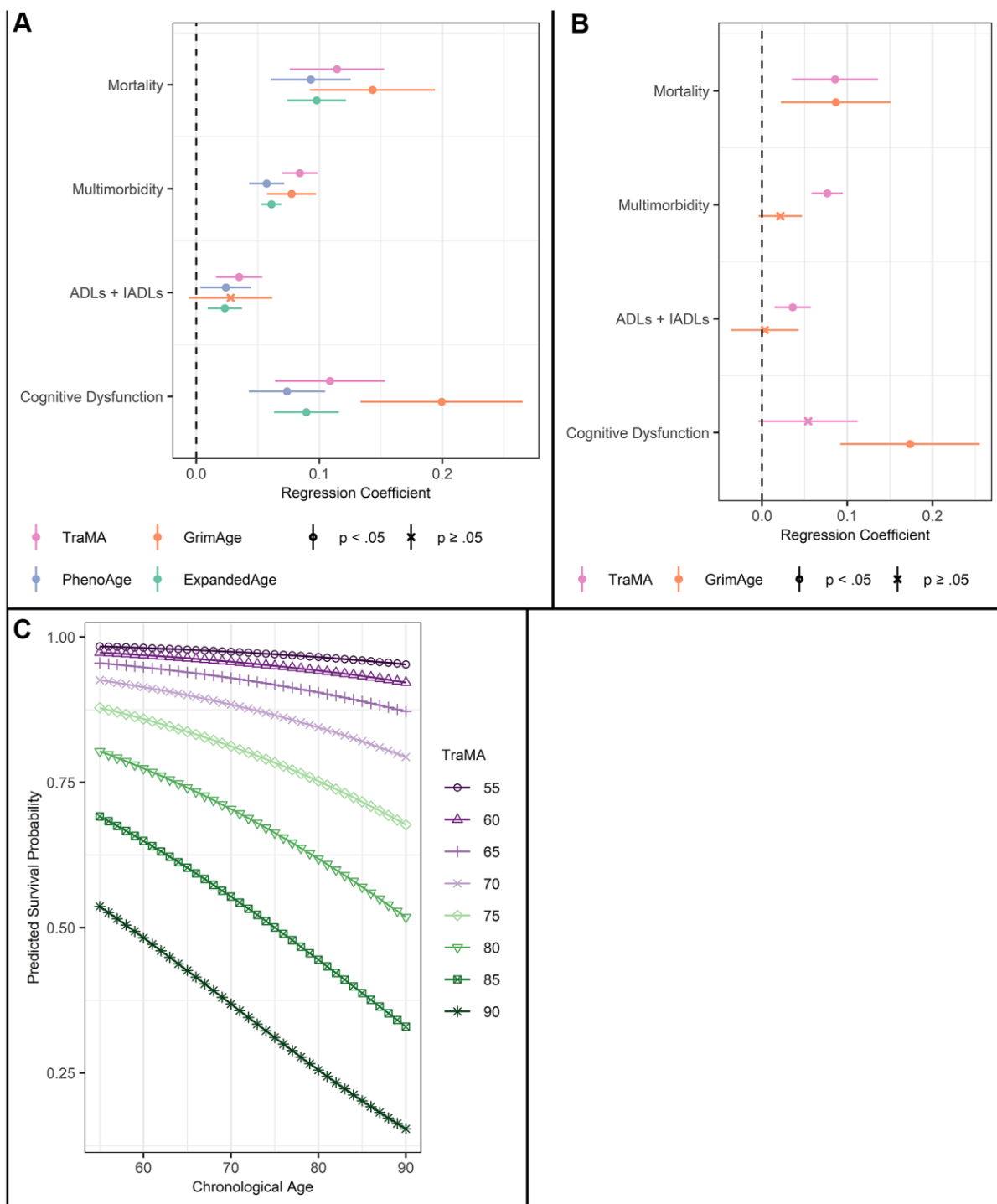
- Epigenetics. 2022; 17:589–611.
<https://doi.org/10.1080/15592294.2021.1939479>
 PMID:34227900
53. Belsky DW, Moffitt TE, Cohen AA, Corcoran DL, Levine ME, Prinz JA, Schaefer J, Sugden K, Williams B, Poulton R, Caspi A. Eleven Telomere, Epigenetic Clock, and Biomarker-Composite Quantifications of Biological Aging: Do They Measure the Same Thing? *Am J Epidemiol*. 2018; 187:1220–30.
<https://doi.org/10.1093/aje/kwx346>
 PMID:29149257
 54. Higgins-Chen AT, Thrush KL, Wang Y, Minter CJ, Kuo PL, Wang M, Niimi P, Sturm G, Lin J, Moore AZ, Bandinelli S, Vinkers CH, Vermetten E, et al. A computational solution for bolstering reliability of epigenetic clocks: Implications for clinical trials and longitudinal tracking. *Nat Aging*. 2022; 2:644–61.
<https://doi.org/10.1038/s43587-022-00248-2>
 PMID:36277076
 55. Bugliari D, Carroll J, Hayden O, Hayes J, Hurd M, Karabatakis A, Main R, Marks J, McCullough C, Meijer E, Moldoff M, Pantoja P, Rohwedder S, St Clair P. RAND HRS Longitudinal File 2016 (V2) Documentation. 2016.
<https://hrsdata.isr.umich.edu/data-products/2016-rand-hrs-fat-file>
 56. R Core Team. R: A language and environment for statistical computing. 2022. <https://www.R-project.org/>
 57. Wickham H, Averick M, Bryan J, Chang W, McGowan LDA, François R, Golemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, et al. Welcome to the Tidyverse. *J Open Source Softw*. 2019; 4:1686.
<https://doi.org/10.21105/joss.01686>
 58. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw*. 2010; 33:1–22.
 PMID:20808728
 59. Simon N, Friedman J, Hastie T, Tibshirani R. Regularization Paths for Cox's Proportional Hazards Model via Coordinate Descent. *J Stat Softw*. 2011; 39:1–13.
<https://doi.org/10.18637/jss.v039.i05>
 PMID:27065756
 60. Golemund G, Wickham H. Dates and Times Made Easy with lubridate. *J Stat Softw*. 2011; 40:1–25.
<https://doi.org/10.18637/jss.v040.i03>
 61. Lumley T. Analysis of Complex Survey Samples. *J Stat Softw*. 2004; 9:1–19.
<https://doi.org/10.18637/jss.v009.i08>
 62. Venables B, Ripley BD. Modern Applied Statistics with S. Springer; 2002. <http://dx.doi.org/10.1007/b97626>

SUPPLEMENTARY MATERIALS

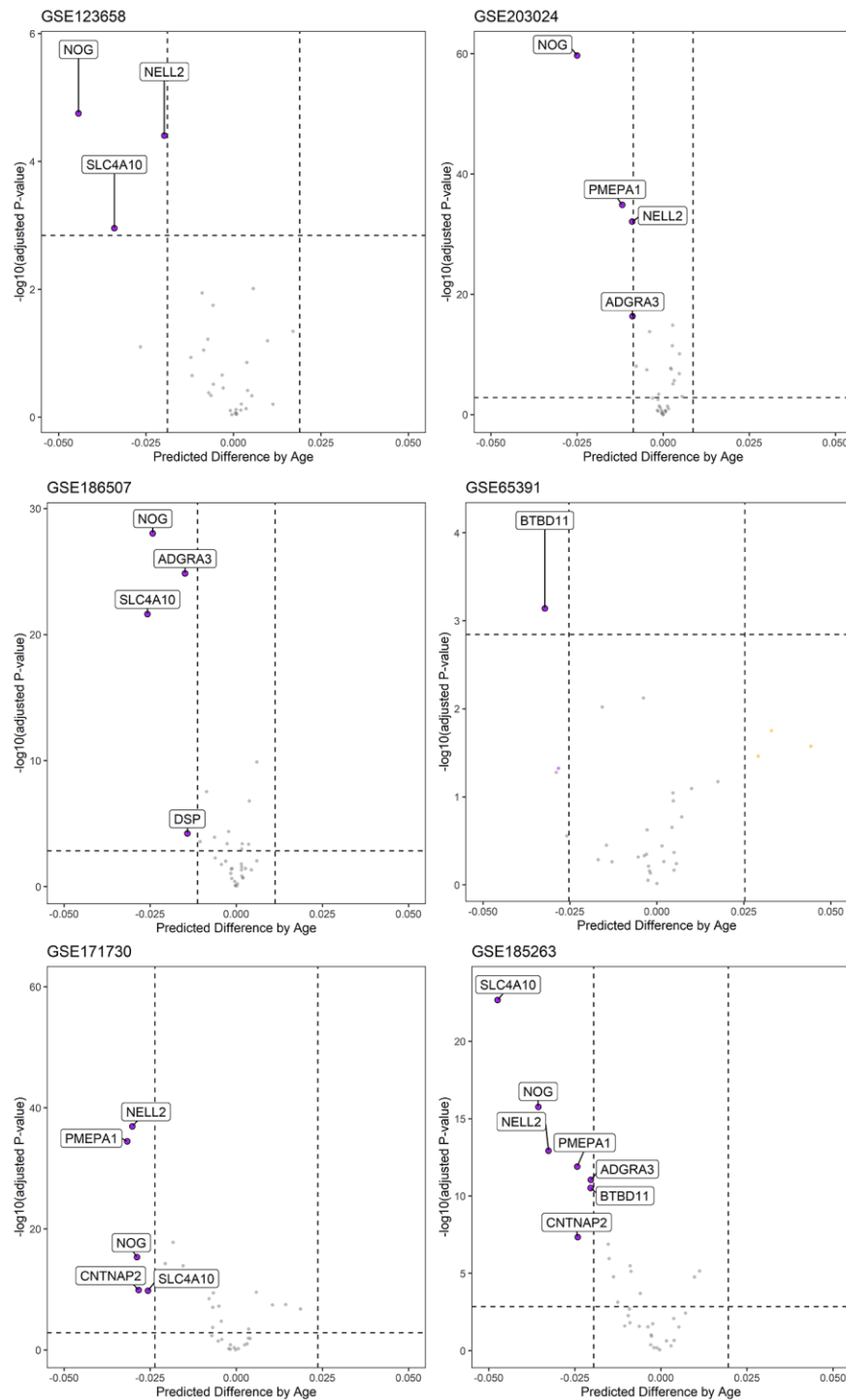
Supplementary Figures



Supplementary Figure 1. Gene networks and functions from GeneMANIA analysis. Line color indicates type of network, and circle colors indicate functions of networks.



Supplementary Figure 3. (A) Regressions of health outcomes on biological aging measures; points represent regression coefficients and bars represent 95% confidence intervals; all models include age, race/ethnicity, sex/gender, and cell type as covariates; each point represents a separate regression equation. (B) Regressions of health outcomes on TraMA and GrimAge controlling for each other; points represent regression coefficients and bars represent 95% confidence intervals; all models include age, race/ethnicity, sex/gender, and cell type as covariates; each row represents a separate regression equation. (C) Predicted survival probabilities for imaginary White, not Hispanic male participants with mean levels of each cell type and varying ages and TraMA scores, using the same regression equation as shown in Figure 1D, Model 3 for HRS.



Supplementary Figure 4. Volcano plots from regressions of each gene in all of the human GEO datasets on age. NOG is significantly, negatively associated with age in all but 1 of these samples, suggesting it is robustly associated with aging. This is also one of the largest coefficients in the TraMA score.

Supplementary Tables

Please browse Full Text version to see the data of Supplementary Tables 1 and 2.

Supplementary Table 1. Genes in TraMA score and their ontologies.

Supplementary Table 2. Genes in the TraMA score and their most associated traits from past.

Supplementary Table 3. Gene ontology terms enriched among the list of genes identified in pathways and FDR corrected Q values from GeneMania analysis.

Function	FDR	Genes in network	Genes in genome
Nephron development	0.034	5	84
Main axon	0.034	4	38
Glomerulus development	0.034	4	42
Cell adhesion mediator activity	0.038	4	56
Regulation of transmembrane receptor protein serine/threonine kinase signaling pathway	0.038	6	201
Regulation of pathway-restricted SMAD protein phosphorylation	0.038	4	58
Kidney vasculature development	0.038	3	18
Transmembrane receptor protein serine/threonine kinase signaling pathway	0.038	7	282
Kidney development	0.038	5	115
Renal system vasculature development	0.038	3	18
Pathway-restricted SMAD protein phosphorylation	0.039	4	60
Regulation of epidermal growth factor-activated receptor activity	0.062	3	24
Endocardial cushion morphogenesis	0.065	3	25
Renal system development	0.081	5	151
Cardiac chamber morphogenesis	0.085	4	79
Urogenital system development	0.088	5	158
Neuron recognition	0.096	3	31

Supplementary Table 4. Correlations among aging measures in HRS.

TraMA	–	0.835	0.838	0.744	0.657	0.329	0.397	0.283
PCGrimAge	0.835	–	0.840	0.719	0.279	0.479	0.273	0.152
PCPhenoAge	0.838	0.840	–	0.749	0.381	0.304	0.632	0.274
ExpandedAge	0.744	0.719	0.749	–	0.349	0.199	0.343	0.692
Age Accelerated TraMA	0.657	0.279	0.381	0.349	–	0.537	0.597	0.449
Age Accelerated PCGrimAge	0.329	0.479	0.304	0.199	0.537	–	0.520	0.303
Age Accelerated PCPhenoAge	0.397	0.273	0.632	0.343	0.597	0.520	–	0.443
Age Accelerated ExpandedAge	0.283	0.152	0.274	0.692	0.449	0.303	0.443	–
	TraMA	PCGrimAge	PCPhenoAge	ExpandedAge	Age Accelerated TraMA	Age Accelerated PCGrimAge	Age Accelerated PCPhenoAge	Age Accelerated ExpandedAge

Age accelerated values are residuals from a regression of the aging measure on chronological age.