

PROTEOMIC AGING CLOCK AND CANCER RISK AND MORTALITY

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE
UNIVERSITY OF MINNESOTA
BY

SHUO WANG

IN PARTIAL FULLFILMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Advisor: Anna Prizment, PhD, MPH

December 2022

Acknowledgements

I would like to express my deepest appreciation to those who supported me through my PhD training and made this achievement possible. First, I would like to thank my advisor, Anna Prizment, for her guidance, mentorship, and support during my PhD training. She was always willing to share her knowledge and ideas on cancer epidemiology and always been available to discuss and answer my questions. I feel incredibly lucky to have Anna as my advisor.

I would also like to thank my other dissertation committee members. Dr. Weihua Guan has shared his biostatistical expertise with me and given me insightful comments which were invaluable to the development of my thesis work. I am also grateful to Dr. Jim Pankow. His expertise in epidemiological methods and insightful comments were invaluable. Thank you also to Dr. Bharat Thyagarajan for always being willing to share his knowledge on aging and biomarkers and providing valuable input on my thesis.

In addition, I would like to acknowledge the students, postdocs, and faculty in the division of Epidemiology and Community Health who helped me along the way. Thank you to all who helped me with statistical analysis coding and advised me on PhD questions. I would particularly like to thank Guillaume Onyeaghala and Elizabeth Polter.

I would also like to thank Dr. Elizabeth Platz, who provided guidance through my thesis and other cancer projects. Dr. Platz always answered my research questions and provided important comments on the draft of the manuscripts.

Finally, I would like to thank my family and friends. Thank you for providing me encouragement and support throughout my PhD program. To my parents, Yingxin Wang and Qingxian Zhao, thank you for encouraging me to pursue a PhD and believing in me.

Dedication

This dissertation is dedicated to my parents – Yingxin Wang and Qingxian Zhao, who have always encouraged me and been my strongest supporters.

Abstract

Proteomic aging clocks (PACs) have been proposed as a measure of biological aging. The deviation of PAC from chronological age is called PAC acceleration. The aims of this dissertation were to construct and validate PACs in a large population-based study -- the Atherosclerosis Risk in Communities (ARIC) study and test their associations with mortality, cancer risk and survival.

Using proteomics data from ARIC, in manuscript 1, we constructed PACs at middle age and older age using healthy participants. Both PACs were strongly correlated with chronological age, and the age acceleration for both PACs was significantly associated with all-cause mortality. Among middle-aged participants, we compared our newly created PAC to three published PACs. The newly created PAC was strongly correlated with all three published PACs and the age acceleration for all these PACs was similarly associated with mortality.

In Manuscript 2, using Visit 2 proteomics data, we constructed a cancer-specific PAC (CaPAC) and compared it to the published PAC created by Lehallier (Lehallier's PAC, this PAC was most strongly correlated with chronological age among published PACs). CaPAC was strongly correlated with chronological age and Lehallier's PAC. Although only the age acceleration for CaPAC but not Lehallier's PAC was significantly associated with risk of overall cancer and colorectal cancer, the magnitude of these associations was similar for both PACs. Acceleration for both PACs was significantly and similarly associated with lung cancer risk.

In Manuscript 3, using Visit 5 proteomics data, we constructed a cancer-specific PAC (CaPAC5) and compared it to Lehallier's PAC in cancer survivors and cancer-free

participants. Both PACs were similarly associated with all-cause mortality among cancer survivors of overall cancer. These associations were similar to those among cancer-free participants.

Overall, the findings from this dissertation showed that PACs that were developed in different populations were similarly associated with all-cause mortality, cancer risk and survival, suggesting the robustness of PACs.

Tables of Contents

List of Tables	ix
List of Figures.....	xiii
Chapter 1 Introduction	1
A. Aging.....	1
B. Biological age.....	2
B.1 Aging clock	3
C. Aging in people with cancer	10
C.1 Aging and risk of cancer	10
C.2 Accelerated aging in cancer survivors	11
Chapter 2 ARIC Study	13
A. Study design and population.....	13
B. Data collection	13
B.1 Proteomics data	14
B.2 Assessment of cancer	14
B.3 Assessment of mortality.....	15
Chapter 3 Manuscript 1: Construction and comparison of proteomic aging clocks in the Atherosclerosis Risk in Communities (ARIC) Study	16
A. Overview.....	16
B. Introduction.....	18
C. Methods.....	20
C.1 Study population	20
C.2 Blood collection	21

C.3 Protein measurement and quality control.....	21
C.4 Identifying healthy participants.....	22
C.5 Assessment of mortality and other characteristics of interest	22
C.6 Statistical analysis	24
D. Results.....	29
D.1 PACs at V2.....	29
D.2 V5 ARIC PAC	31
D.3 Exploratory analysis.....	32
D.4 Sensitivity analysis.....	33
E. Discussion	34
F. Supplemental material	61
Chapter 4 Manuscript 2: Proteomic Aging Clock and Risk of Cancer in the	
Atherosclerosis Risk in Communities (ARIC) Study	66
A. Overview.....	66
B. Introduction.....	67
C. Methods.....	69
C.1 Study population	69
C.2 Ascertainment of cancer cases	70
C.3 Blood collection	70
C.4 Protein measurement and quality control.....	70
C.5 Assessment of other participant characteristics	71
C.6 Statistical analysis	72
D. Results.....	77

D.1 Pearson correlation coefficients between PACs and chronological age	77
D.2 Distributions of participant characteristics across quartiles of PAC acceleration	77
D.3 Associations of PAC acceleration with risk of cancer	77
D.4 Exploratory analysis.....	78
D.5 Sensitivity analysis.....	81
E. Discussion	81
Chapter 5 Manuscript 3: Proteomic Aging Clock and Risk of Mortality among older Cancer Survivors in the Atherosclerosis Risk in Communities (ARIC) Study.....	107
A. Overview.....	107
B. Introduction	109
C. Methods.....	110
C.1 Study population	110
C.2 Assessment of cancer cases.....	111
C.3 Assessment of mortality	111
C.4 Blood collection	112
C.5 Protein measurement and quality control.....	112
C.6 Assessment of other participant characteristics	113
C.7 Statistical analysis	114
C.8 Sensitivity analysis.....	118
C.9 Exploratory analysis.....	118
D. Results.....	118
D.1 Pearson correlation between PAC and chronological age	118

D.2 Distribution of participant characteristics across tertiles of PAC acceleration	119
D.3 Associations between PAC acceleration and mortality	120
D.4 Sensitivity analysis.....	121
D.5 Exploratory analysis.....	122
E. Discussion	122
Chapter 6 Summary	138
Chapter 7 References.....	141

List of Tables

Table 1.1 Results summarized from previous studies that examined association of age acceleration for epigenetic clocks with risk of cancer and mortality	5
Table 1.2 10 examples of common proteins associated with age	7
Table 1.3 Results summarized from previous studies that examined association of PAC or PAC acceleration with all-cause mortality	9
Table 3.1 Pearson correlation and MAE between PACs at V2 and V5 and chronological age	42
Table 3.2 V2 participant characteristics across quartiles of age acceleration for V2 ARIC PAC and Lehallier's PAC	43
Table 3.3 Association between age acceleration for V2 ARIC PAC and Lehallier's PAC and mortality	44
Table 3.4 V5 participant characteristics across quartiles of age acceleration for the V5 ARIC PAC	45
Table 3.5 Association of V5 ARIC PAC acceleration and V2 to V5 change in PAC acceleration with mortality	46
Table 3.6 Association between participant characteristics at V2 and age acceleration for the V5 ARIC PAC	47
Supplemental Table 3.1 Description of published proteomic aging clocks.....	48
Supplemental Table 3.2 Description of V2 ARIC PACs.....	49
Supplemental Table 3.3 Pearson correlation coefficients between V2 ARIC PACs in the V2 test set.....	50

Supplemental Table 3.4 Including/excluding participants with hypertension for healthy participants at V5 to construct PACs using elastic net regression	51
Supplemental Table 3.5 Model R squared after regressing PAC acceleration on covariates at the corresponding visits	52
Supplemental Table 3.6 Top 20 proteins with the largest absolute weight in the V2 ARIC PAC and V5 ARIC PAC.....	53
Supplemental Table 3.7 V2 participant characteristics across quartiles of age acceleration for Tanaka's PAC and Sathyan's PAC	54
Supplemental Table 3.8 Association between age acceleration for Tanaka's and Sathyan's PACs and mortality	55
Supplemental Table 3.9 Association between age acceleration for the V2 ARIC PAC and published PACs and mortality stratified by sex and race	56
Supplemental Table 3.10 Association between age acceleration for the V5 ARIC PAC and mortality stratified by sex and race	57
Supplemental Table 3.11 Top 40 pathways overrepresented in the V2 ARIC PAC and V5 ARIC PAC	58
Supplemental Table 3.12 Association between V2 to V5 change in PAC acceleration and mortality stratified by sex and race.....	60
Table 4.1 Visit 2 characteristics of cancer-free participants at Visit 2 across quartiles of age acceleration for CaPAC and Lehallier's PAC.....	89
Table 4.2 Association between age acceleration for CaPAC and Lehallier's PAC risk of overall cancer and cancer subtypes.....	90

Table 4.3 Association between age acceleration for CaPAC and Lehallier's PAC risk of overall cancer and cancer subtypes stratified by sex and race.....	91
Table 4.4 Association between age acceleration for CaPAC2 (cancer-specific PAC constructed with proteins with the largest within-individual changes between V2 and V5) and V2 ARIC PAC (healthy participants PAC) and cancer risk	93
Table 4.5 Association between CaPAC3 (Visit 3 cancer-specific clock) acceleration and risk of overall cancer and cancer subtypes	94
Supplemental Table 4.1 Description of proteomic aging clocks used in this study	97
Supplemental Table 4.2 Association between age acceleration for CaPAC and Lehallier's PAC risk of overall cancer and cancer subtypes stratified by smoking status.....	98
Supplemental Table 4.3 Visit 2 characteristics of cancer-free participants at Visit 2 across quartiles of age acceleration for CaPAC2 and V2 ARIC PAC.....	100
Supplemental Table 4.4 Association between age acceleration for CaPAC2 and V2 ARIC PAC and cancer risk stratified by sex and race.....	102
Supplemental Table 4.5 Associations of age acceleration for CaPAC2 and V2 ARIC PAC with cancer risk stratified by smoking status	104
Supplemental Table 4.6 Top 40 pathways overrepresented in CaPAC	106
Table 5.1 Participant characteristics of across tertiles of acceleration for PAC in cancer survivors and cancer-free participants; Visit 5	129
Table 5.2 Associations of PAC acceleration with all-cause and CVD mortality among cancer survivors and cancer-free participants, and cancer mortality among cancer survivors.....	131

Table 5.3 Associations of PAC acceleration with all-cause mortality among survivors of postmenopausal breast, prostate, and colorectal cancers in Model 3	132
Supplemental Table 5.1 Associations of PAC acceleration with all-cause and CVD mortality among cancer survivors and cancer-free participants, and cancer mortality among cancer survivors in Models 1-3.....	134
Supplemental Table 5.2 Association of PAC acceleration with all-cause mortality among cancer survivors and cancer-free participants stratified by race and sex.....	135
Supplemental Table 5.3 Associations of PAC acceleration with all-cause mortality among a subset of survivors of postmenopausal breast, prostate, and colorectal cancers with the information on stage of diagnosis.....	136
Supplemental Table 5.4 Top 40 pathways overrepresented in CaPAC5	137

List of Figures

Figure 3.1 Study population at Visit 2	37
Figure 3.2 Person correlation (r) between proteomic aging clocks at Visit 2 and chronological age in healthy participants in the V2 test set.....	38
Figure 3.3 Study population at Visit 5	40
Figure 3.4 Pearson correlation (r) between the V5 ARIC PAC and chronological age in healthy participants in the V5 test set	41
Figure 4.1 Study population.....	87
Figure 4.2 Pearson correlation (r) between CaPAC and Lehallier's PAC and chronological age in the test set	88
Supplemental Figure 4.1 Pearson correlation (r) between CaPAC2 (i.e., cancer-specific PAC constructed using proteins with largest within-individual changes between Visit 2 and Visit 5) and chronological age in the test set2 (r= 0.75)	95
Supplemental Figure 4.2 Pearson correlation (r) between V2 ARIC PAC (healthy participants PAC) and chronological age in participants who remain cancer-free until 2015 (r=-0.77).....	96
Figure 5.1 Pearson correlation (r) between PAC and chronological age in cancer-free participants and cancer survivors, Visit 5	128
Supplemental Figure 5.1 Distribution of PAC acceleration among cancer-free participants (test set) and cancer survivors, Visit 5	133

Chapter 1 Introduction

A. Aging

In the U.S., the average human life expectancy has been increased by 30 years in the 20th century. This has been considered as one of the greatest accomplishments of public health. However, the increased life expectancy in the population has also led to steady growth in the prevalence of age-related health conditions and disabilities. Since the development of those conditions will lead to an increase in the risk of mortality and an increase in the healthcare costs for the elderly in the U.S., understanding the intrinsic biological mechanisms of aging and the development of prevention measures and interventions that prolong the healthy life have captured the attention of greatest scientific minds.^{1,2}

As has been underscored by gerontologists and geriatricians, individuals' aging processes cannot be sufficiently measured by chronological age. For example, some individuals may develop several age-related conditions, appear frail, and need assistance in daily routines at age 70, whereas others may remain free of any health conditions and independent of assistance even if they are much older than 70. Evidence also shows that, compared to non-centenarians, centenarians frequently achieve longer longevity with a lower prevalence of comorbidities.³ With these discoveries, there is an opportunity to apply some anti-aging interventions via changes in lifestyle and anti-aging drugs to delay the incidence of diseases and prolong healthy lifespan.

Several anti-aging drugs have been proposed, in particular the drugs called senolytics, which could reduce the incidence of physical dysfunction and prolong healthy lifespan.² Senolytics were shown to decrease the risk of physical dysfunction in a human

clinical trial,⁴ and shown to decrease the risk of physical dysfunction and mortality in mice.⁵ Besides anti-aging drugs, exercise may be a cost-effective prevention strategy to reduce morbidity and mortality.⁶ In mice model, exercise was shown to prevent certain cellular senescent phenotypes, which could accelerate many age-related pathologies.⁷

Given that the older adult US population is expanding, and as anti-aging drugs are being developed, understanding individuals' aging processes is imperative.

B. Biological age

To better understand individuals' aging processes, researchers introduced a term called "biological age" to capture specific information about how old an individual is biologically, independent of chronological age. Biological age, according to the definition proposed by Baker and Spratt, is characterized by the biological parameter[s] of an organism, either alone or in some multivariate composite that will, in the absence of disease, better predict functional capability at some late age than will chronological age".⁸

Several clinical measures have been proposed to capture biological age, including gait speed, grip strength, 6-minute walk, Fried (the Cardiovascular Health Study (CHS)) Frailty Phenotype, clinical geriatric assessment, maximal oxygen consumption.⁹⁻¹¹ However, these measures are unlikely to account for all aspects of aging and it is time-consuming and difficult to collect them, especially in the elderly. Therefore, there is a need in creating an objective and invasive measure of biological age. To create such a measure, researchers have proposed biological age estimators called aging clocks, using epigenetics, transcriptomics, metabolomics, proteomics, and other biomarkers.¹² This dissertation focused on proteomic aging clock (PAC). Before

discussing PAC, we discussed aging clock in general and the most studied aging clock – epigenetic clocks.

B.1 Aging clock

Aging clock is calculated based on a set of age-associated molecules that are measured in blood or tissue. Aging clock is strongly correlated with chronological age in healthy individuals; however, in individuals with age-related conditions, aging clock will deviate from chronological age, because these conditions would impact the levels of age-associated molecules.^{13,14} The positive deviation of aging clock from chronological age is called age acceleration. Therefore, aging clock could be used to identify individuals with age acceleration. Aging clock could also predict individuals' future risk of age-related conditions. In addition, aging clock could track the effectiveness of anti-aging interventions in clinical trials and may be used as a surrogate measure of the disease if the association is causal.^{13,15,16} Several aging clocks have been developed, including epigenetic clocks and PACs.

B.1.1 Epigenetic clock

Epigenetic clocks, which include a set of DNA methylation-based biomarkers in blood or tissue, are the most studied aging clocks. Among all the epigenetic clocks, Horvath and Hannum epigenetic clocks, which were trained on chronological age, are the first-generation epigenetic clocks. Horvath's clock is calculated based on methylation levels of 353 CpGs sites obtained from a variety of tissues and cell types. The Hannum epigenetic clock is derived from 71 CpGs sites using whole blood samples. Both of these

clocks show high correlations with chronological age (Pearson correlation (r) = 0.96 for Horvath and r = 0.91 for Hannum) and small mean absolute errors from chronological age (3.6 years for Horvath and 4.9 years for Hannum).^{17,18} Besides these two clocks, Levine et al. developed a second-generation epigenetic clock, the DNAm PhenoAge.¹⁹ DNAm PhenoAge was calculated based on 513 CpGs from the whole blood and was highly correlated with chronological age (r = 0.89).¹⁹ The difference between first- and second-generation epigenetic clocks was that the second-generation clocks were trained on health outcomes, rather than chronological age alone. Further, Lu et al. developed GrimAge, another second-generation epigenetic clock. GrimAge included 1030 CpGs and found correlations with chronological age of 0.79-0.95 in different validation cohorts.²⁰ The importance of this epigenetic clock is that it is a composite biomarker based on seven DNAm surrogates and a DNAm-based estimator of smoking pack-years, which was found to be a better predictor of mortality than the actual observed biomarkers.²⁰ A recent study developed a novel estimator – DunedinPACE, which is a DNA methylation-based biomarker of pace of aging.²¹ DunedinPACE measures “years of physiological decline occurring per 12 months of calendar time”,²² not biological age. Other epigenetic clocks include Weidner’s,²³ Zhang’s,²⁴ Lin’s,²⁵ Yang’s,²⁶ Bocklandt’s,²⁷ and ELOVL2²⁸ clocks.

Epigenetic clocks have been applied in epidemiology studies to examine the associations with age-related diseases. This dissertation focuses on aging and risk of mortality, cancer risk and cancer survival. Therefore, I summarized results from previous studies that examined the associations of age acceleration for epigenetic clocks with risk of cancer and mortality (in general population and in cancer survivors) (**Table 1.1**).

Table 1.1 Results summarized from previous studies that examined association of age acceleration for epigenetic clocks with risk of cancer and mortality

Paper	Data	Outcome	Epigenetic clock	Unit for age acceleration	HR/OR
Risk of mortality in general population					
2020 Tanaka ²⁹	InCHIANTI	All-cause mortality	GrimAge	Per 1 SD	1.44 (1.20, 1.74)
			DNAm PhenoAge		1.32 (1.14, 1.54)
			Horvath		1.02 (0.88, 1.18)
2019 Lu ²⁰	A meta-analysis of Women's Health Initiative (WHI), Framingham Heart Study (FHS), and Jackson Heart Study (JHS), and InCHIANTI	All-cause mortality	GrimAge	Per 1 year	1.10 (1.09, 1.12)
			DNAm PhenoAge		1.05 (1.04, 1.05)
			Hannum		1.04 (1.03, 1.05)
			Horvath		1.02 (1.01, 1.03)
2018 Levine ¹⁹	A meta-analysis of WHI, FHS, and JHS	All-cause mortality	DNAm PhenoAge	Per 1 year	1.05 (1.03, 1.05)
			Hannum		1.04 (1.03, 1.05)
			Horvath		1.02 (1.02, 1.03)
2018 Roetker ³⁰	Africa Americans in ARIC	CVD mortality	Horvath	Per 5-year	1.06 (0.98, 1.14)
			Hannum		1.04 (0.94, 1.14)
2016 Chen ³¹	A meta-analysis of 13 population-based cohort (including ARIC)	All-cause mortality	Horvath	Per 1 year	1.01 (1.01, 1.02)
			Hannum		1.03 (1.02, 1.04)
Risk of cancer					
2022 Morales Bernstein ³²	A study using Mendelian randomization method	Prostate cancer risk	GrimAge	Per 1 year	0.93 (0.87, 0.99)
		Colorectal cancer (CRC) risk	GrimAge		1.12 (1.04, 1.20)
2021 Dougue ³³	A Pooled analysis of seven case-control studies nested in the Melbourne Collaborative Cohort Study (MCCS)	Risk of multiple types of cancer combined	DNAm PhenoAge	Per 1 SD	1.11 (1.05, 1.18)
			GrimAge		1.11 (1.03, 1.20)

2020 Chung ³⁴	Pooled analysis: pancreatic cancer cases and matched controls from Nurses' health study and Health Professionals Follow-up study	Pancreatic cancer risk	Hannum	Q4 vs. Q1	1.73 (1.11, 2.71)
			Horvath		1.28 (0.83, 1.97)
			DNAm PhenoAge		1.67 (1.07, 2.62)
2019 Lu ²⁰	A meta-analysis of WHI, FHS, JHS, and InCHIANTI.	Any cancer risk	GrimAge	Per 1 year	1.07 (1.05, 1.08)
			DNAm PhenoAge		1.02 (1.01, 1.03)
			Hannum		1.01 (0.99, 1.02)
			Horvath		1.01 (1.00, 1.03)
2019 Kresovich ³⁵	The Sister Study	Breast cancer risk	Hannum	Per 5-year	1.10 (1.00, 1.21)
			Horvath		1.08 (1.00, 1.17)
			DNAm PhenoAge		1.15 (1.08, 1.23)
2018 Dougue ³⁶	A Pooled analysis of seven case-control studies nested MCC	Risk of multiple types of cancer combined	Hannum	Per 5-year	1.09 (1.04, 1.14)
			Horvath		1.05 (1.01, 1.10)
2018 Levine ³⁷	WHI	Lung cancer risk	DNAm PhenoAge	Per 1 year	1.05 (p = 0.031)
2017 Ambatipudi ³⁸	A case-control study nested in European Prospective Investigation into Cancer and Nutrition (EPIC) cohort	Postmenopausal breast cancer risk	Horvath	Per 1 year	1.07 (1.02, 1.11)
2015 Levine ²⁵	WHI	Lung cancer risk	Horvath	Per 1 year	1.50 (P=0.0034)
Risk of mortality in cancer survivors					
2019 Zheng ³⁹	Cancer Genome Atlas (TCGA)	All-cause mortality among CRC survivors	Horvath	Positive Horvath age acceleration vs. negative	1.97 (1.14, 1.39)
2018 Dougue ³⁶	A Pooled analysis of seven case-control studies nested in MC	All-cause mortality among all types of cancer survivors	Hannum	Per 5-year	1.05 (1.01, 1.10)
			Horvath		1.02 (0.98, 1.05)

In summary, several epigenetic clocks have been developed using tissue or the whole blood. All of these clocks showed high correlations with chronological age. Some clocks were associated with the risk of cancer and mortality (in general population and in cancer survivors), and the magnitude of most of the associations were similar for different epigenetic clocks.

B.1.2 Proteomic aging clock (PAC)

Besides epigenetic clock, a novel measure – PAC, which combines a set of proteomic-based aging-related biomarkers in the blood, has been proposed as a potential biological age estimator. Age-associated protein biomarkers may be promising biomarkers of aging because they, as intermediate phenotypes, can provide more information on aging and age-related pathology.^{13,40} In addition, proteins serve as a target in 96% of FDA approved medications.⁴¹ Therefore, proteins comprising PACs hold promise as targets of anti-aging drugs.

Many proteins have been reported to be associated with age in previous studies. We selected 10 examples of proteins commonly reported to be associated with age in previous studies (those 10 proteins showed the most significant association with age) (**Table 1.2**).² Previous studies have applied pathway analysis for proteins associated with age. For example, two studies performed enrichment analysis using Gene Ontology Biological Process (GO BP) database for proteins that were found to be associated with age and found that pathways relevant to immune system were enriched.^{2,14} A systematic review (2020) did the GLAD4U disease enrichment analysis on 1,128 proteins associated with age and found that the top enriched diseases including cancer and heart diseases.²

Table 1.2 10 examples of common proteins associated with age

Protein	UniProt ID	Expression trend
GDF15	Q99988	Increase
PTN	P21246	Increase
ADAMTS5	Q9UNA0	Increase
NPPB	P16860	Increase
MMP12	P39900	Increase
FLSTL3	Q95633	Increase
CXCL10	P02778	Increase
TNFRSF11B	Q00300	Increase
VEGFA	P15692	Increase
CCDC80	Q76M96	Increase

In addition to individual age-related proteins, several PACs have been proposed. In a systematic review of 36 studies, Johnson et al. proposed two PACs: the 23-protein PAC and the 83-protein PAC.² They developed and validated these two PACs using the INTERVAL cohort (White healthy persons),⁴² which comprised 3,301 healthy individuals aged 18-76 years (two-thirds of samples was used as the training set and the remaining one-third of the samples was used as the test set). In their test set, the 23-protein and 83-protein PACs had Pearson correlations of 0.66 and 0.87 with chronological age, respectively, and mean absolute errors between PAC and chronological age were 8.17 years and 4.88 years, for the 23-protein and 83-protein PACs, respectively.² Recently, using data from INTERVAL cohort, Lehallier et al. proposed another PAC, which included 491 proteins and had a correlation with chronological age of 0.94 in the test set (one-third of the total samples, N =1,123).¹⁴ Additional PACs have been created by Sathyan et al. using the LonGenity cohort (N = 1,025, age range: 65-95 years, White persons), which included two groups of offspring with different survival of parents. One group had parents with exceptional longevity (at least one parent lived to age 95 or older), and the other group had parents with usual survival (having neither parents who lived to age 95).⁴³ They created four PACs that

consisted of 162, 75, 67, and 35 proteins. The 162-protein PAC was selected by elastic net regression from 4,265 proteins. While the 75-, 67-, and 35-protein PACs were selected by elastic net regression from the top significant 200, 100, and 50 proteins, in the 4,265 proteins, associated with chronological age. The 162-, 75-, 67-, and 35-protein PACs had correlations with chronological age of 0.79, 0.80, 0.79, and 0.78, respectively.⁴³ Further, another study of healthy individuals by Tanaka et al. found a correlation of 0.94 between a PAC, including 76 proteins, and chronological age.¹³ To determine the minimum number of proteins required to create a meaningful PAC, Tanaka et al. fitted a series of PACs. They found that a PAC that only contained eight proteins could reach a correlation with chronological age of 0.92.¹³ An even higher correlation of 0.97 was observed between a PAC including 373 proteins and chronological age in the study that combined the INTERVAL and LonGenity cohorts.⁴⁴ Collectively, PACs showed high correlations with chronological age.

Because PAC is a novel measure, studies on PAC and risk of health conditions are scarce. Only a few studies examined the association between PAC or PAC acceleration and all-cause mortality (**Table 1.3**). To the best of our knowledge, no previous studies examined the association between PAC and cancer or tested PAC in cancer survivors.

Table 1.3 Results summarized from previous studies that examined association of PAC or PAC acceleration with all-cause mortality

Paper	Study population	Exposure	Outcome	Unit	Results
2020 Tanaka ²⁹	InCHIANTI	Age acceleration for the 76-protein PAC developed by Tanaka et al	All-cause mortality	Per 1 SD	HR = 1.29 (1.11, 1.50)
2020 Sathyan ⁴³	LonGenity cohort	162-protein PAC developed by Sathyan et al	All-cause mortality	Per 1 year	HR = 1.21 (1.15, 1.27)

In summary, several PACs have been developed, and all of them showed high correlations with chronological age. Only a few studies examined association between PACs and all-cause mortality, and it is unclear if PACs can predict the risk of other age-related conditions, such as cancer. No studies tested PAC in cancer survivors. Thus, studies are needed to examine the associations between PACs and risk of other age-related conditions, such as cancer, and to test PACs in cancer survivors. The importance of examining association between PAC and cancer risk and testing PAC in cancer survivors is further described in the next section (**C. Aging in people with cancer**). In addition, most published PACs were developed in a study population that included White persons. Thus, there is a need to create PACs in more diverse populations.

C. Aging in people with cancer

C.1 Aging and risk of cancer

Growing evidence indicates that people with cancer age faster than those without, and thus they experience accelerated aging. The reason for the accelerated aging is most likely related to cancer-related inflammation, cancer treatment, and lifestyle risk factors such as smoking and obesity. However, it is not clear if the accelerated aging starts even before a diagnosis of cancer and how much the rate of aging differs between people with and without cancer. These issues can be examined by using PACs, which have been proposed to estimate an individual's biological age. Measuring PACs in relation to cancer incidence is especially important because most cancer cases develop over a long period of time and have a long subclinical period, and 30%-50% of all cancer cases are preventable;⁴⁵ thus, measuring PAC may provide a new method for identifying people at

high risk of cancer that need more frequent screening. In addition, proteins included into PACs may serve as targets for novel anti-aging drugs that will eventually slow down aging.⁴⁶ Cancer is an age-related disease, and both cancer and aging result from epigenetic damage accumulation,⁴⁷ inflammation, oxidative stress, and other damage.⁴⁸⁻⁵⁰ Therefore, slowing aging is likely to delay the development of cancer.

C.2 Accelerated aging in cancer survivors

Cancer survivors are a rapidly growing population in the U.S. There are 16.9 million cancer survivors as of January 2019, which is projected to rise to 26.1 million by 2040 in the U.S.⁵¹ This trend is attributed to the improved screening, early detection, advances in cancer treatment, and an increased life expectancy in the U.S. However, the improvement in cancer survival has also led to the realization that many cancer survivors experience accelerated aging, i.e., their biological age is greater than their chronological age.

It is common in clinical practice to observe that, after receiving treatment, cancer survivors age faster, i.e., they experience more age-related health conditions. Clinically, “accelerated aging” phenotypes in cancer survivors are characterized by the development of age-related conditions, including premature mortality and comorbidities – cardiovascular disease (CVD), chronic organ dysfunction, frailty, and cognitive impairment, which will impact their long-term health and quality of life.

Given that many cancer survivors experience accelerated aging, it is vital to examine biological age in this group. Examining biological age may characterize the aging process in routine follow-ups for cancer survivors and inform the need for anti-

aging interventions via changes in lifestyle and anti-aging treatment, thus improving quality of life of cancer survivors.

The aims of this dissertation were to construct de novo PACs in a large population-based prospective cohort of White and Black men and women – the Atherosclerosis Risk in Communities (ARIC) study. We also compared the performance of the newly constructed PACs in ARIC to published PACs by examining their correlation with chronological age and their association with risk of mortality in general population (Manuscript 1), cancer risk (Manuscript 2) and cancer survival (Manuscript 3).

Chapter 2 ARIC Study

This dissertation includes three manuscripts that utilize data from the Atherosclerosis Risk in Communities (ARIC) study.

A. Study design and population

The ARIC study (RRID:SCR_021769) is a prospective cohort initiated in 1987.^{52,53} At Visit 1 (1987-89), 15,792 participants aged 45-64 years were recruited from four US communities - Washington County, Maryland; the northwest suburbs of Minneapolis, Minnesota; Jackson, Mississippi; and Forsyth County, North Carolina. Participants were selected from each of the four communities using community specific probability sampling. In Forsyth County, households were identified by area sampling. In other communities, age-eligible lists were used to identify households. Lists used included: driver's license or state identification cards in Jackson; jury duty eligibility with driver's license, identification cards, or voter registration cards in Minneapolis; and driver's license or inclusion in a 1975 private county health census in Washington County. Participants in Minnesota and Maryland were primarily white, and the recruitment in Mississippi was restricted to Black residents.

B. Data collection

Participants had a clinical exam at baseline (Visit 1) and eight visits have been completed thus far: Visit 2 (1990-92), Visit 3 (1993-95), Visit 4 (1996-98), Visit 5 (2011-13), Visit 6 (2016-17), Visit 7 (2018-19), and Visit 8 (2020-21).⁵² At each visit, each participant reported information on demographics, tobacco use and alcohol intake,

medical history, and use of medications. The clinical exams at each visit included anthropometry and blood pressure. Resting electrocardiograms (ECG) and carotid artery B-mode ultrasound imaging were obtained at some exams. Medical events were also identified through annual (or semi-annual since 2012) follow-up telephone calls to participants.

B.1 Proteomics data

Proteins were measured using frozen plasma samples collected from Visits 2, 3, and 5. Samples were analyzed using a SOMAmer (Slow Off-rate Modified Aptamers)-based capture array called SomaScan® by Somalogic, Inc. (Boulder, CO, USA).⁵⁴⁻⁵⁷ The SomaScan platform uses single-stranded DNA-based aptamers to capture conformational protein epitopes.

Protein analyte measurements underwent the regular SomaScan data standardization and normalization process.^{58,59} Among the 5,284 available aptamers, aptamers with a Bland-Altman coefficient of variation (CVBA) >50% or a variance of <0.01 on the log scale, or aptamers with nonspecific binding to nonproteins were excluded. After the exclusion, 4955 aptamers were included (at Visits 2, 3 and 5) and the CVBA for split samples were 6% at Visit 2, 12% at Visit 3 and 7% at Visit 5.

B.2 Assessment of cancer

ARIC study has been expanded into a full-fledged cancer epidemiology cohort and includes detailed data on cancer diagnosis and long-term follow up of cancer survivors.⁵³ Incident primary cancer was ascertained through 2015 using state Cancer

registries in Minnesota, North Carolina, Maryland, and Mississippi, and supplemented by abstraction of medical records and hospital discharge summaries.⁵³ An expert panel adjudicated all cases of bladder, breast, colorectal, liver, lung, pancreatic, and prostate cancer. For adjudicated cases, when possible, stage at diagnosis was determined from the cancer registry or medical records using the pathologic TNM stage (Tumor extent, lymph Node involvement, presence of Metastasis). When this information was not available, staging was determined from the cancer registry or clinical TNM stage from cancer registry or medical records according to Surveillance, Epidemiology, and End Results (SEER) summary stage.⁵³

B.3 Assessment of mortality

Mortality was identified through annual (semi-annual since 2012) follow-up telephone calls to participants or their proxies, surveillance of local hospitals, state records and linkage to the National Death Index up to December 31, 2019.⁶⁰

Chapter 3 Manuscript 1: Construction and comparison of proteomic aging clocks in the Atherosclerosis Risk in Communities (ARIC) Study

A. Overview

Background: Proteomic aging clock (PAC) has been proposed to estimate individual's biological age, but most of previous studies included mainly White persons or lacked prospective outcomes. Using the Atherosclerosis Risk in Communities (ARIC) study, a population-based prospective cohort of White and Black participants, we constructed several PACs, and compared their performance with published PACs.

Methods: More than 5,000 plasma proteins were measured at Visit 2 (V2, 1990-92) and Visit 5 (V5, 2011-13) using SomaScan, an aptamer-based assay. Using 2,993 randomly selected healthy participants at V2 (age range: 46-70 years), we trained five ARIC PACs at V2 against chronological age: using 1) elastic net regression: V2 ARIC PAC; 2) Lasso regression: V2 ARIC PAC2; and 3&4) applying different transformations to proteins' measurements: V2 ARIC PAC3&4; and 5) accounting for the nonlinear associations between chronological age and proteins: V2 ARIC PAC5. At V2, we also computed three published PACs. We compared the performance of the five V2 ARIC PACs and published PACs by examining their correlation with chronological age in the remaining 1496 healthy participants at V2 (V2 test set). At V5, we constructed V5 ARIC PAC using elastic net regression in 630 healthy participants at V5 (age range: 66-90 years). We internally validated V5 ARIC PAC by calculating its correlation with chronological age in the remaining 315 healthy participants (V5 test set). To capture associations that are independent of chronological age, we created PAC acceleration, which was estimated as

the residual for each participant after regressing the PAC on chronological age. We further examined the associations between PAC acceleration and mortality.

Results: The five V2 ARIC PACs showed similar correlations with chronological age and were highly correlated with each other in the V2 test set. The Pearson correlation coefficient (r) between chronological age and V2 ARIC PAC in the V2 test set was 0.80, and $r = 0.68-0.76$ between chronological age and three published PACs. The V2 ARIC PAC and all published PACs showed similar associations with all-cause, CVD, and cancer mortality. After adjustment for confounders, a one standard deviation (SD) increase in V2 ARIC PAC acceleration was associated with all-cause mortality: HR (95%CI) = 1.38 (1.34, 1.43); and CVD mortality (in the competing risk model): 1.20 (1.14, 1.27); but not cancer mortality (in the competing risk model): 1.05 (0.98, 1.11). In the V5 test set, the V5 ARIC PAC was correlated with chronological age ($r = 0.71$). In adjusted models, a one SD increase in V5 ARIC PAC acceleration was also significantly associated with all-cause mortality: HR (95%CI) = 1.67 (1.56, 1.79); CVD mortality (in the competing risk model): 1.33 (1.18, 1.51); and cancer mortality (in the competing risk model): 1.19 (1.03, 1.37).

Conclusion: PACs constructed using different methods showed similar correlations with chronological age and were highly correlated with each other. PACs developed in different populations showed similar associations with mortality, suggesting the robustness of PAC.

B. Introduction

In the U.S., the average human life expectancy increased by 30 years during the 20th century. However, the increased life expectancy in the population has also led to a rise in the prevalence of age-related diseases and disabilities. Given that the development of those conditions will lead to an increased risk of mortality and increased healthcare costs in the U.S., research is needed to understand the biological mechanisms of aging and develop prevention measures and interventions that prolong healthy life.^{1,61}

Individuals' aging processes cannot be accurately measured by chronological age. For example, some individuals may develop physiological dysregulation at an earlier chronological age than others.^{1,61} To better understand aging processes, researchers introduced a term called “biological age” to capture specific information about how old an individual is biologically, independent of chronological age. Biological age, according to the definition proposed by Baker and Sprott, is characterized by the “biological parameter[s] of an organism, either alone or in some multivariate composite that will, in the absence of disease, better predict functional capability at some late age than will chronological age”.⁶²

To estimate an individual's biological age, researchers have developed metrics called aging clocks, using epigenetics, transcriptomics, metabolomics, proteomics, and other biomarkers.¹² Aging clocks are strongly correlated with chronological age in healthy individuals; however, in individuals with age-related conditions, aging clocks will deviate from chronological age because these conditions would impact levels of age-associated molecules.^{13,14} Aging clocks could be used to identify individuals who have a positive deviation of biological age from their chronological age (called age

acceleration), and predict their future risk of age-related conditions. In addition, aging clocks could also track the effectiveness of anti-aging interventions in clinical trials and may be used as a surrogate measure of the disease if the association is causal.^{13,15,16}

The most studied aging clocks are epigenetic clocks, which include a set of DNA methylation sites, such as Horvath, Hannum, DNAm PhenoAge, and GrimAge clocks.¹⁷⁻²⁰ However, the underlying mechanisms of changes in these DNA methylation sites remain unclear.⁶³ Thus, we examined another novel aging clock, the proteomic aging clock (PAC). PACs are promising because proteomic-based biomarkers, as intermediate phenotype between genetics and diseases, may provide more information on aging and age-related pathology.^{13,40} In addition, proteins serve as a target in 96% of FDA approved medications.⁴¹ Therefore, the proteins comprising PACs hold promise as targets of anti-aging drugs. Targeting aging instead of targeting a single disease is advantageous, because it can simultaneously decrease the incidence of several age-related diseases together and prolong individuals' healthy lifespan. Recently, several PACs have been published, including a PAC developed by Lehallier [2020]¹⁴ (so called Lehallier's PAC), a PAC by Tanaka [2020]¹³ (Tanaka's PAC), and a PAC by Sathyan [2020]⁴³ (Sathyan's PAC). The descriptions of published PACs are presented in **Supplemental Table 3.1**. Although all these published PACs showed high correlations with chronological age, previous studies have been limited by small sample sizes, included mainly White persons, or did not have prospective outcomes, e.g., mortality. Thus, there is a need in creating PAC in a large population-based study that can perform well in diverse populations.

The goal of this study is to construct de novo PACs in a large population-based prospective cohort of White and Black men and women — the Atherosclerosis Risk in Communities (ARIC) study, which recently measured more than 5,000 proteins at several

visits using SomaScan assay (v.4), an aptamer-based technology. In this study, we constructed ARIC PACs at Visit 2 (V2, 1990-92) and Visit 5 (V5, 2011-13) in a group of participants without major diseases or health conditions including cancer, diabetes, hypertension (or uncontrolled hypertension for V5), abnormal kidney function, chronic obstructive pulmonary disease (COPD), and cardiovascular disease (CVD) (called “healthy participants” in this study). We compared the performance of our newly constructed ARIC PACs with the three published PACs: Lehallier’s, Tanaka’s, and Sathyan’s PACs based on their correlation with chronological age and their associations with mortality.

C. Methods

C.1 Study population

This study included White and Black men and women participating in the ARIC study (RRID:SCR_021769). The ARIC study is a prospective cohort initiated in 1987.^{52,53} At Visit 1 (V1, 1987-89), 15,792 participants aged 45-64 years were recruited from four study centers – Maryland, Minnesota, Mississippi, and North Carolina. Participants in Minnesota and Maryland were primarily white, and the recruitment in Mississippi was restricted to Black residents. The ARIC study was approved by institutional review boards at each participating center, and all study participants provided written informed consent. Thus far, eight study visits have been completed.⁵² Additionally, participants received follow-up telephone calls annually between 1987-2012 and semi-annually after 2012, with response rates of 90%-99% for the annual follow-up calls and 83%-90% for semi-annual follow-up calls among living participants who have not withdrawn consent to be contacted.⁵³

C.2 Blood collection

In this study, we used the blood samples collected at V2 from 11,761 White and Black women and men (chronological age range: 46-70 years) and at V5 from 5,183 White and Black women and men (chronological age range: 66-90 years). The blood sample collection, processing, and storage protocol was designed to minimize the spontaneous biochemical reactions after blood collection and is consistent with the recommended practice for proteomics data analysis in epidemiological studies.^{58,64,65} Briefly, after venipuncture, blood samples were put immediately in an ice water bath. Centrifugation was then performed within 10 min after venipuncture at room temperature (15-25 °C). After centrifugation, the aliquots were stored at -80 °C within 90 min from venipuncture and were never thawed before this analysis.

C.3 Protein measurement and quality control

Samples were analyzed using a SOMAmer (Slow Off-rate Modified Aptamers)-based capture array called SomaScan® by Somalogic, Inc. (Boulder, CO, USA).⁵⁴⁻⁵⁷ The SomaScan platform uses single-stranded DNA-based aptamers to capture conformational protein epitopes.

Protein analyte measurements underwent the regular SomaScan data standardization and normalization process.^{58,59} First, hybridization control normalization was applied to each sample based on a set of hybridization control sequences to correct for systematic biases during hybridization. Second, median signal normalization was applied to measures within a plate to remove sample or assay biases that may be because

of pipetting variation, variation in reagent concentrations, assay timing, and other sources of systematic variability within a single plate run. Finally, each plate contained calibrator samples for each SOMAmer reagent, which was used to correct for plate-to-plate variation based on established global reference standards. Among the 5,284 available aptamers, aptamers with a Bland-Altman coefficient of variation (CVBA) >50% or a variance of <0.01 on the log scale, or aptamers nonspecific binding to nonproteins were excluded. After the exclusion, 4955 aptamers were included (at V2 and V5) that corresponded to 4,712 proteins and the CVBA for split samples were 6% at V2 and 7% at V5. Protein measures were reported as relative fluorescent units (RFU) and were log₂ transformed to correct for skewness.

C.4 Identifying healthy participants

In this study, healthy participants were defined as participants without major diseases or health conditions including cancer, diabetes, hypertension (or uncontrolled hypertension for V5), abnormal kidney function, chronic obstructive pulmonary disease (COPD), and cardiovascular disease (CVD). We identified 4,489 healthy participants at V2 and 945 healthy participants at V5 (7,272 participants and 4,238 participants with any major diseases or health conditions at V2 and at V5, respectively) (**Figures 3.1 and 3.3**). Identification of healthy participants at V2 and V5 is described in **F. Supplementary material**.

C.5 Assessment of mortality and other characteristics of interest

Mortality was identified through annual (semi-annual since 2012) follow-up telephone calls to participants or their proxies, surveillance of local hospitals, state records and linkage to the National Death Index up to December 31, 2019.⁶⁰ The date and some causes of death were verified by death certificate review. All-cause mortality was defined as death from any cause. CVD mortality and cancer mortality were defined based on underlying cause of death: *International Classification of Diseases, Ninth Revision*, codes (ICD-9 codes) 390–459 or *International Classification of Diseases, Tenth Revision*, codes (ICD-10 codes) I00–I99 for CVD deaths. ICD-9 code 140-239 or ICD-10 codes C00-C97 for cancer deaths.

Other characteristics of interest included demographic, and lifestyle/medical characteristics, namely chronological age, sex, race, study center, education, smoking status, pack-years of smoking, alcohol intake, body mass index (BMI), physical activity, aspirin use, and estimated glomerular filtration rate (eGFR).⁵² Sex, race, study center, and education attainment were collected at V1, pack-years of smoking was collected at V2, physical activity was collected at V1 (used as physical activity at V2 in this study) and V5, chronological age at V2 and V5 were calculated from date of birth until date of corresponding visits, and all the other variables listed above were collected at both V2 and V5. Another characteristic of interest is cognitive function. Cognitive function at V2 was assessed using three neuropsychological tests: the Delayed Word Recall Test (DWRT), the Digit Symbol Substitution Test (DSST) of the Wechsler Adult Intelligence Scale–Revised, and the Word Fluency Test (WFT). Detailed procedures for assessing these characteristics are described in **F. Supplemental material**.

C.6 Statistical analysis

PACs were constructed using R (version 4.1.2, package “glmnet”) and all the other analyses were performed using SAS 9.4 (SAS Institute Inc, Cary, NC).

C.6.1 Construction of PACs

To construct PACs at V2 and V5, among the healthy participants at the corresponding visits, we randomly selected two-thirds of healthy participants at each visit and used them as the training set and the remaining one-third of healthy participants was used as the test set (**Figure 3.1 and Figure 3.3**). We used the training set to train PACs against chronological age and get appropriate hyperparameter and weight for each protein. We used the test set to internally validate PACs.

Construction of PACs at V2

Using the V2 training set, we constructed five V2 ARIC PACs (**Supplemental Table 3.2**). We constructed two PACs by applying elastic net regression (alpha = 0.5, V2 ARIC PAC) and Lasso regression (V2 ARIC PAC2). Because the V2 ARIC PAC and V2 ARIC PAC2 were similarly correlated with chronological age and perfectly correlated with each other (**Table 3.1 and Supplemental Table 3.3**), we used elastic net regression to construct all the other ARIC PACs. Elastic net regression combines the regularization of both Lasso regression and Ridge regression,⁶⁶ and it was used to construct most previous aging clocks, including PACs and epigenetic clocks. To reduce the effect of outliers, we constructed PACs with proteins winsorized at 4 standard deviations (SDs) (V2 ARIC PAC3) and with proteins after inverse normal transformation (V2 ARIC

PAC4). To account for the potential nonlinear associations between chronological age and proteins, we also constructed a PAC by including square and cubic terms of each protein (each aptamer if there were two or more aptamers per protein, V2 ARIC PAC5). The hyperparameter values for penalized regressions were selected based on 10-fold cross-validation.

In addition to the newly constructed V2 ARIC PACs, we also computed three published PAC at V2: Lehallier's PAC,¹⁴ Tanaka's PAC,¹³ and Sathyan's PAC.⁴³ We estimated ARIC weights for Lehallier's and Tanaka's PACs instead of using the published weights because ARIC did not include all the aptamers reported in Lehallier's and Tanaka's PACs (**Supplemental Table 3.1**). Using the aptamers available in ARIC, we estimated ARIC weights for three published PACs by applying Ridge regression. The hyperparameter value for Ridge regression was selected based on a 10-fold cross-validation. Although ARIC included all the aptamers reported in Sathyan's PAC, we also estimated ARIC weights for this PAC, because the PAC with ARIC weights showed a higher correlation with chronological age than the PAC with the published weights (**Table 3.1 and Supplemental Table 3.1**).

Construction of PAC at V5

Because hypertension is one of the most common conditions in older persons in the U.S.,⁶⁷ to construct PAC at V5, we added participants with controlled hypertension when selecting healthy participants. Adding these participants increased the number of healthy participants by 462 but did not change the PAC's performance as shown by our sensitivity analysis (**Supplementary Table 3.4**).

ARIC PAC at V5 (V5 ARIC PAC) was constructed using elastic net regression and with log2 transformed proteins (the same as V2 ARIC PAC). Using the V5 training set, elastic net regression (alpha = 0.5 and hyperparameter value was selected based on 10-fold cross-validation) selected 135 aptamers for V5 ARIC PAC.

C.6.2 Validation of PACs

We used the V2 and V5 test sets to internally validate PACs at the corresponding visits. We computed the PACs for each of the participant in the test set based on their aptamer levels: $PAC = \beta_0 + \sum_{i=1}^n \beta_i \times aptamer_i$, where $aptamer_i$ is the level of i th aptamer and the intercept (β_0) and weights (β_i) were estimated using the training set. We estimated the performance of each PAC by calculating the Pearson correlation between the PAC and chronological age. In addition, we tested performance by calculating median absolute error (MAE) between the PAC and chronological age in the test set.

To further investigate these PACs, we compared PACs across demographic and lifestyle/medical characteristics and examined their association with mortality. To conduct these analyses, we focused on the V2 ARIC PAC (788 aptamers, without any additional protein transformation) among the five V2 ARIC PACs, because the five PACs were highly correlated with each other (Pearson correlation > 0.96) in the V2 test set (**Supplemental Table 3.3**). Validation of this simplest PAC would facilitate its application in external studies. To capture associations that are independent of chronological age, we calculated PAC acceleration, the residual for each participant after regressing the PAC on chronological age at the corresponding visit.³⁵ Demographic and lifestyle/medical characteristics at V2 and V5 were examined across quartiles of PAC

acceleration. We used Cox proportional hazards regression to estimate hazard ratios (HRs) and 95% confidence intervals (CIs) for all-cause mortality, CVD mortality, and cancer mortality in relation to PAC acceleration calculated at V2 and V5. PAC acceleration was modeled as a continuous variable. For V2 analysis, for each participant, total person-years were determined from date of V2 until death or administrative censoring on December 31, 2019, whichever occurred first. For V5 analysis, the follow-up was started from V5. The proportional hazards assumption was examined by graphical methods (log-log survival curves, dichotomized PAC acceleration at median) and the assumption was not violated in any regression models. The model was adjusted for chronological age, sex, race, study center, education, BMI, smoking status, pack-years of smoking (for V2 analysis only, because it was not available at V5), alcohol intake, physical activity, diabetes, eGFR, hypertension, and history of CVD at the corresponding visits. These variables were selected based on R squared (> 0.0015) for the association between a variable of interest at V2 or V5 (exposure) and PAC acceleration at the corresponding visits (outcome) in a chronological age-adjusted model. The R squared estimates are shown in **Supplemental Table 3.5**. We did not adjust for aspirin use because aspirin use had a small R squared (<0.0015) at both V2 and V5. We also examined if sex or race modified the associations of PAC acceleration with all-cause mortality, CVD mortality, and cancer mortality using a multiplicative term.

C.6.3 Exploratory analyses

Using ToppGene,⁶⁸⁻⁷⁰ we conducted overrepresentation analyses (ORA) for the proteins included in V2 ARIC PAC and V5 ARIC PAC. ORA is a type of pathway

analysis that shows whether proteins from pre-defined pathways, for example those participating in the same molecular process or are regulated by the same transcription factor (i.e., to a specific GO term or KEGG pathways) are present more than would be expected (overrepresented) in our data.⁷¹ UniProt IDs were provided as the inputs, the background was set to all protein-coding genes, and the False Discovery Rate (FDR) significance level was set to 0.05.

In another exploratory analysis, we examined if demographic and lifestyle/medical characteristics and cognitive function at middle age (at V2) were associated with PAC acceleration at older age (at V5). This analysis was conducted using the multivariable linear regression among 4,553 participants at V5; the participants' characteristics -- chronological age, sex, race, education, BMI, smoking status, pack-years of smoking, alcohol intake, physical activity (at V1), hypertension, CVD, diabetes, and three cognitive function tests (DWRT, DSST, and WFT) -- were included as predictors.

In the last exploratory analysis, we examined whether the change in PAC acceleration between V2 and V5 was associated with all-cause mortality, CVD mortality, and cancer mortality after V5. This analysis was performed among 2,707 participants at V5 after excluding those in the V2 and V5 training sets. We calculated the V2 to V5 change in PAC acceleration as PAC acceleration at V5 minus PAC acceleration at V2. We used Cox proportional hazard regression to estimate HRs (95% CIs) for all-cause mortality, CVD mortality, and cancer mortality. For each participant, total person-time was determined from date of V5 until death or administrative censoring on December 31, 2019, whichever occurred first. The model was adjusted for chronological age, sex, race,

study center, education, BMI, smoking status, alcohol intake, physical activity, diabetes, eGFR, hypertension, and CVD at V5. We also examined if sex or race modified the association between the V2 to V5 change in PAC acceleration and mortality using a multiplicative term.

C.6.4 Sensitivity analysis

In a sensitivity analysis, using the Fine and Gray method,^{72,73} deaths from other causes were treated as competing events when estimating associations of V2 ARIC PAC acceleration, V5 ARIC PAC acceleration, and the V2 to V5 change in PAC acceleration with CVD mortality and cancer mortality.

D. Results

D.1 PACs at V2

D.1.1 Pearson correlation coefficients between PACs at V2 and chronological age

All the five V2 ARIC PACs, i.e., PACs that were newly constructed in ARIC, showed similar correlations with chronological age (**Table 3.1 and Figure 3.2a-3.2e**). For all these newly constructed PACs, we did not notice a major overfitting problem, i.e., neither the correlation of PAC with chronological age nor median absolute error (MAE) markedly differ between training and test sets. Pearson correlation coefficients between the V2 ARIC PAC and chronological age in the V2 training and test sets were 0.92 (MAE=1.50 years) and 0.80 (MAE=2.20 years), respectively (**Table 3.1 and Figure 3.2a**). The top 20 proteins with the largest absolute weights in V2 ARIC PAC are shown in **Supplemental Table 3.6**. Of the three published PACs, Lehallier's PAC had a slightly

higher correlation with chronological age than Tanaka's and Sathyan's PACs (**Table 3.1 and Figure 3.2f-3.2h**). Pearson correlation coefficients between Lehallier's PAC and chronological age in the V2 training and test sets were 0.82 (MAE=2.13 years) and 0.76 (MAE=2.39 years), respectively (**Table 3.1 and Figure 3.2f**). Tanaka's and Sathyan's PACs had similar correlations with chronological age in the V2 test set (**Table 3.1 and Figure 3.2g-3.2h**). In the V2 test set, V2 ARIC PAC was strongly correlated with Lehallier's PAC ($r=0.89$), Tanaka's PAC ($r=0.77$), and Sathyan's PAC ($r=0.85$).

D.1.2 Distribution of V2 characteristics across quartiles of PAC acceleration at V2

Distributions of V2 characteristics across quartiles of acceleration for V2 PACs are shown in **Table 3.2** and **Supplemental Table 3.7**. Among 8,768 participants at V2, the range of V2 ARIC PAC acceleration was -13.97 to 24.22 years. Participants with higher V2 ARIC PAC acceleration were more likely to be ever smokers and never drinkers, and have diabetes, hypertension, cardiovascular disease (CVD), a lower education level (less than high school), a lower mean physical activity level, and a lower mean eGFR (**Table 3.2**).

D.1.3 Association between PAC acceleration at V2 and mortality

Among 8,768 participants at V2, 5,294 died by 2019 (median follow-up = 24.88 years). Age acceleration for the V2 ARIC PAC and three published PACs showed similar associations with all-cause mortality, CVD mortality, and cancer mortality (**Table 3.3 and Supplemental Table 3.8**). For V2 ARIC PAC: per 1 SD, HRs (95% CIs) were 1.38 (1.34, 1.43) for all-cause mortality; 1.45 (1.37, 1.52) for CVD mortality; and 1.19 (1.12,

1.27) for cancer mortality (**Table 3.3**). Neither sex nor race modified the associations of accelerations for V2 ARIC PAC or published PACs with all-cause mortality, CVD mortality, and cancer mortality (p-interactions >0.05) (**Supplemental Table 3.9**).

D.2 V5 ARIC PAC

D.2.1 Pearson correlation coefficients between the V5 ARIC PAC and chronological age

Pearson correlation coefficients between the V5 ARIC PAC and chronological age in the V5 training and test sets were 0.84 (MAE=1.47 years) and 0.71 (MAE=2.36 years), respectively (**Table 3.1 and Figure 3.4**). For the newly constructed V5 ARIC PAC, we did not notice a major overfitting problem. The top 20 proteins with the largest absolute weights in the V5 ARIC PAC are shown in **Supplemental Table 3.6**.

D.2.2 Distribution of V5 characteristics across quartiles of V5 ARIC PAC acceleration

Among the 4,553 participants at V5, the range of V5 ARIC PAC acceleration was -7.54 to 16.99 years. Participants with higher V5 ARIC PAC acceleration were more likely to be current smokers and never drinkers, and have hypertension, CVD, less education, a lower mean physical activity level, and a lower mean eGFR (**Table 3.4**).

D.2.3 Association between V5 ARIC PAC acceleration and mortality

Among 4,553 participants at V5, 1,123 died by the end of 2019 (median follow-up=7.43 years). After adjustment for potential confounders, V5 ARIC PAC acceleration was associated with all-cause mortality, CVD mortality, and cancer mortality: per 1 SD, HRs (95% CIs) were 1.67 (1.56, 1.79) for all-cause mortality; 1.54 (1.36, 1.75) for CVD

mortality; and 1.33 (1.15, 1.53) for cancer mortality (**Table 3.5**). Sex statistically modified the association with cancer mortality [p-interaction = 0.01]; the association was stronger in women compared to men. Neither sex nor race modified the associations with all-cause mortality or CVD mortality (p-interactions >0.15) (**Supplemental Table 3.10**).

D.3 Exploratory analysis

D.3.1 Overrepresented pathways in the V2 ARIC PAC and V5 ARIC PAC

Using the Gene Ontology Biological Process database, we found 1,472 overrepresented pathways (FDR p-value < 0.05) among proteins included in the V2 ARIC PAC. The top 40 overrepresented pathways are listed in **Supplemental Table 3.11**. They include “positive regulation of phosphorylation”, “regulation of immune system process”, “positive regulation of cell differentiation”, “regulation of phosphorylation”, “cytokine production”, “regulation of cytokine production”, “cell killing”, “regulation of programmed cell death”, “positive regulation of immune system process”, and “innate immune response”.

For proteins included in the V5 ARIC PAC, we found that 202 pathways were overrepresented (FDR p-value < 0.05). The top 40 pathways are listed in **Supplemental Table 3.11**. They include “circulatory system development”, “inflammatory response”, “learning or memory”, “cognition”, and “regulation of immune system process”.

D.3.2 Association between V2 participant characteristics and V5 ARIC PAC acceleration

White race, current smoking, hypertension, CVD, diabetes, a higher BMI, higher eGFR, and lower scores for the Digit Symbol Substitution Test (DSST) and Word

Fluency Test (WFT) (all measured at V2 but race assessed at baseline) were associated with a higher V5 ARIC PAC acceleration (**Table 3.6**).

D.3.3 Association between V2 to V5 change in PAC acceleration and mortality

Among 2,707 participants, the Person correlation coefficient was 0.69 ($p < 0.0001$) between V2 ARIC PAC and V5 ARIC PAC. The V2 to V5 change in PAC acceleration was associated with all-cause mortality and CVD mortality, but not cancer mortality: per 1 SD, HRs (95% CIs) were 1.30 (1.19, 1.42) for all-cause mortality; 1.20 (1.03, 1.40) for CVD mortality; and 1.17 (0.98, 1.40) for cancer mortality (**Table 3.5**). Neither sex nor race modified the association of V2 to V5 change in PAC acceleration with all-cause mortality, CVD mortality, or cancer mortality (p -interactions > 0.10) (**Supplemental Table 3.12**).

D.4 Sensitivity analysis

In a sensitivity analysis, deaths from other causes were treated as competing events using the Fine and Gray method. After accounting for non-CVD deaths as competing events, the associations of V2 ARIC PAC acceleration, V5 ARIC PAC acceleration, and the V2 to V5 change in acceleration with CVD mortality became weaker or nonsignificant: per 1 SD, HRs (95% CIs) were 1.20 (1.14, 1.27) for V2 ARIC PAC acceleration; 1.33 (1.18, 1.51) for V5 ARIC PAC acceleration; and 1.09 (0.94, 1.27) for the V2 to V5 change in acceleration (**Table 3.3 and Table 3.5**). After accounting for non-cancer deaths as competing events, the associations of V2 ARIC PAC acceleration and V5 ARIC PAC acceleration with cancer mortality became weaker or nonsignificant:

per 1 SD, HRs (95% CIs) were 1.05 (0.98, 1.11) for V2 ARIC PAC; and 1.19 (1.03, 1.37) for V5 ARIC PAC; and the association for V2 to V5 change in acceleration remained insignificant: per 1 SD, HR (95% CI) was (0.91, 1.33) (**Table 3.3 and Table 3.5**).

E. Discussion

In a large prospective population-based study of White and Black individuals, we created and validated de novo PACs in middle age (V2 ARIC PAC) and older age (V5 ARIC PAC) using 4,955 aptamers measured by the SomaScan assay (v.4). Both ARIC PACs were developed in healthy participants and were strongly correlated with chronological age. Age acceleration for both PACs was statistically significantly associated with all-cause mortality and CVD mortality in both Cox proportional hazards and competing risk models after adjustment for potential confounders. In addition, PAC acceleration at later age was significantly associated with cancer mortality in both Cox proportional hazards and competing risk models. We also found that the change in PAC acceleration from middle to late adulthood was also associated with all-cause mortality, but the association was weaker than the association for acceleration for PAC constructed in older adults.

We used different statistical methods and different transformations of proteins to construct five PACs and choose the best among them using ARIC proteomic data measured in middle-aged participants (at V2). These five ARIC PACs showed similar correlation with chronological age and were highly correlated with each other. Thus, PACs trained using different statistical methods and with different transformations of proteins show very similar performance.

In our study, we found that acceleration for V2 ARIC PAC, V5 ARIC PAC, and the three published PAC was inversely associated with education level, which has been shown to impact health outcomes.⁷⁴ We also found that males (compared to females) tended to have higher V5 ARIC PAC acceleration, but not V2 ARIC PAC acceleration. On average, males and females may age similarly up to middle age but males may age faster when they become older.⁷⁵ We found inconsistent associations between race and the V2 ARIC PAC and published PACs. For example, Black participants were likely to have higher V2 ARIC PAC acceleration, on average, but lower age acceleration for Lehallier's and Tanaka's PACs, and the percentages of Black participants were similar in the first and last quartiles of Sathyan's PAC acceleration. The inconsistent associations may be explained by the different study populations used to construct these PACs. The ARIC study includes both White and Black participants. Tanaka et al. constructed their PAC in White individuals, Black individuals, and individuals with other race groups. Lehallier and Sathyan et al. constructed their PACs in individuals from Europe.

Among middle-aged participants, age acceleration for V2 ARIC PAC (PAC of middle age) and published PACs were similarly associated with mortality including all-cause, CVD, and cancer mortality. Our findings are similar to findings for Tanaka's PAC in the InCHIANTI study (N = 459, chronological age: 21-98 years) which reported a significant association between age acceleration and all-cause mortality after adjustment for chronological age, sex, and study site [per 1 SD, HR (95% CI) = 1.29 (1.11, 1.50)].²⁹ In our study, we found stronger associations with mortality for V5 ARIC PAC acceleration than the associations for V2 ARIC PAC acceleration, which may be explained by the older study participants at V5 who had a shorter time of follow-up until

death. Thus, our de novo PACs created in ARIC and published PACs showed very similar correlations with chronological age and similar associations with mortality although they included different proteins and were developed in individuals of different health status and different race-ethnicities. Taken together, these findings suggest that PACs are robust predictors of mortality outcomes.

The strengths of this population-based study include the prospective design with over 20 years of follow-up in males and females, validated information about some causes of death, and detailed demographic and lifestyle information. Another strength is that the cohort includes a large community-based sample of both White and Black individuals, while the populations in most of the previous studies of PACs were primarily White.^{14,43} In addition, the version of SomaScan assay used in ARIC measured more proteins than most of the published PACs. With a larger panel of proteins, we may be able to develop a PAC that more accurately predicts chronological age. Moreover, with the availability of proteomics measured at two visits, we constructed PACs of middle age and older age. Our study also has several limitations. First, the possibility of protein degradation during long-term storage cannot be excluded. However, the blood samples were frozen right after their collection and have never been thawed, thereby limiting the possibility of degradation⁷⁶. Second, the SomaScan assay provides relative quantification instead of absolute quantification.⁷⁷

In conclusion, we developed de novo PACs in a population of White and Black individuals in midlife and older age and showed that these PACs were able to estimate biological age and predict risk of mortality. Future studies are needed to validate our PACs by examining their association with other age-related diseases.

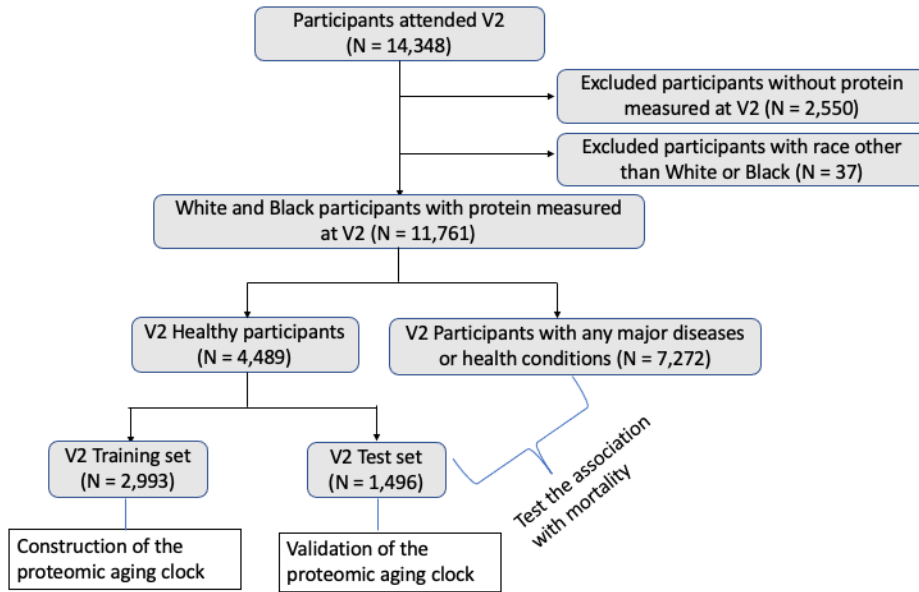
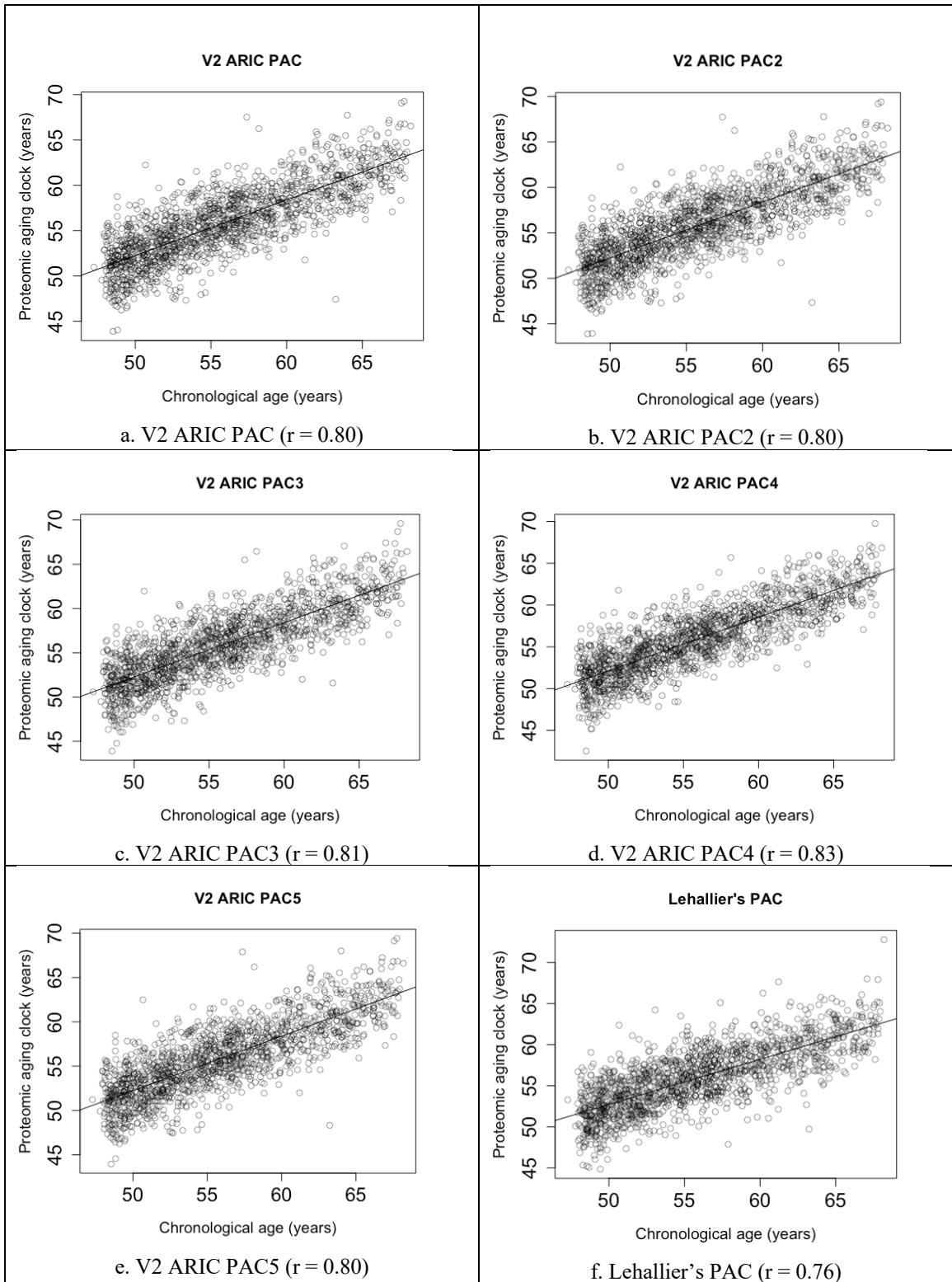


Figure 3.1 Study population at Visit 2



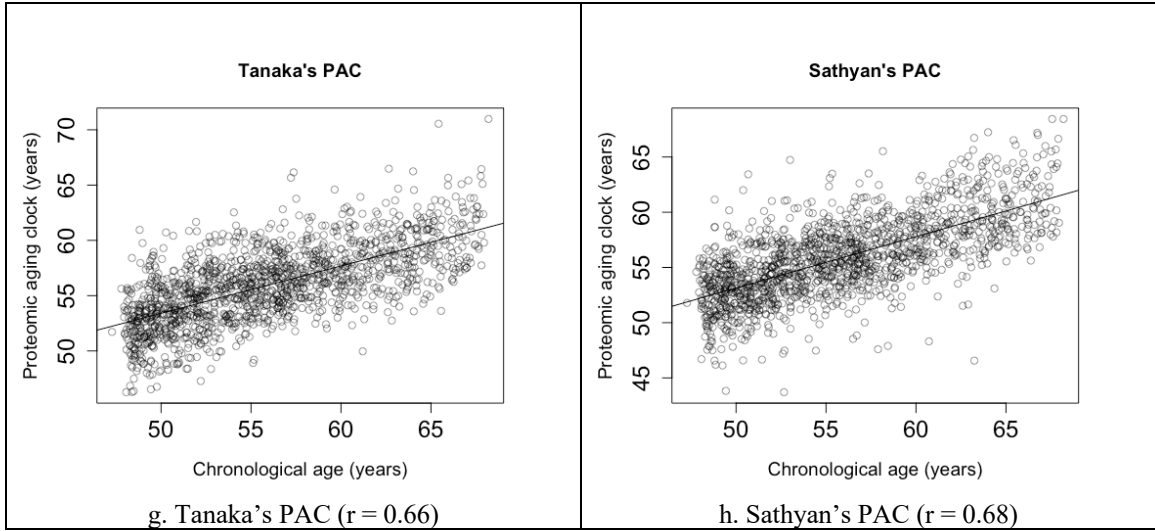


Figure 3.2 Pearson correlation (r) between proteomic aging clocks at Visit 2 and chronological age in healthy participants in the V2 test set

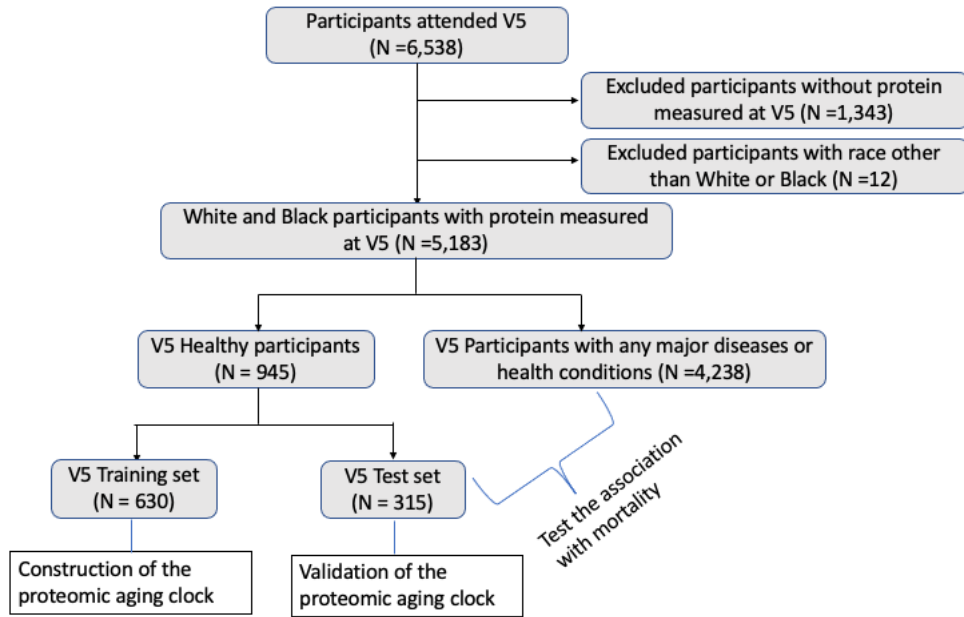


Figure 3.3 Study population at Visit 5

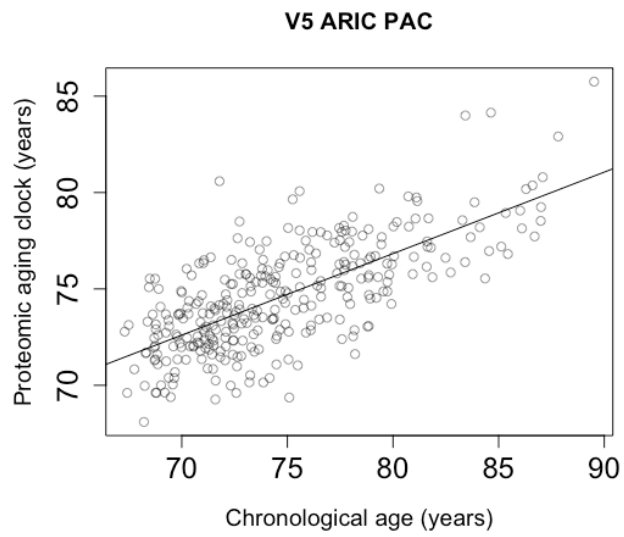


Figure 3.4 Pearson correlation (r) between the V5 ARIC PAC and chronological age in healthy participants in the V5 test set

Table 3.1 Pearson correlation and MAE between PACs at V2 and V5 and chronological age

	V2 ARIC PAC	V2 ARIC PAC2	V2 ARIC PAC3	V2 ARIC PAC4	V2 ARIC PAC5 ^c
Number of aptamers in PAC	788	781	631	694	1071
Hyperparameter value for penalized regression	0.11	0.05	0.12	0.10	0.11
Correlation ^a in the training set ^b	0.92	0.92	0.91	0.92	0.92
Correlation ^a in the test set ^b	0.80	0.80	0.81	0.83	0.80
MAE ^a in the training set ^b	1.50	1.49	1.58	1.48	1.51
MAE ^a in the test set ^b	2.19	2.20	2.16	2.00	2.20

	Lehallier's PAC ^d	Tanaka's PAC ^d	Sathyan's PAC ^d	Sathyan's PAC calculated based on proteins at V2 and with published weights ^d	V5 ARIC PAC
Number of aptamers in PAC	415	68	162	162	135
Hyperparameter value for penalized regression	1.03	0.19	0.71	-	0.46
Correlation ^a in the training set ^b	0.82	0.68	0.77	-	0.84
Correlation ^a in the test set ^b	0.76	0.66	0.68	0.58	0.71
MAE ^a in the training set ^b	2.13	2.74	2.38	-	1.47
MAE ^a in the test set ^b	2.39	2.78	2.67	10.00	2.36

Abbreviations: MAE – median absolute error; PAC – proteomic ageing clock; V2 – Visit 2; V5 – Visit 5.

^a Pearson correlation and MAE between PAC and chronological age.

^b Among the healthy participants at V2 and V5, we randomly selected two-thirds of healthy participants at each visit and used them as training set at the corresponding visits and the remaining one-third of healthy participants at each visit was used as test set at the corresponding visits.

^c V2 ARIC PAC5 included the square and cubic terms of aptamers.

^d Published PACs were calculated using protein levels from V2.

Table 3.2 V2 participant characteristics across quartiles of age acceleration for the V2 ARIC PAC and Lehallier's PAC

	V2 ARIC PAC				P-value	Lehallier's PAC				P-value
	Q1 (N = 2192)	Q2 (N = 2192)	Q3 (N = 2192)	Q4 (N = 2192)		Q1 (N = 2192)	Q2 (N = 2192)	Q3 (N = 2192)	Q4 (N = 2192)	
PAC acceleration	-14.0 to -1.9	-1.8 to -0.2	-0.1 to 1.7	1.8 to 24.2		-15.1 to -2.0	-1.9 to -0.2	-0.1 to 1.8	1.9 to 26.5	
Chronological age, years (SD)	58.3 (5.8)	57.9 (5.8)	57.8 (5.7)	58.3 (5.6)	0.07	58.2 (5.8)	58.1 (5.8)	58.1 (5.6)	58.1 (5.6)	0.88
Female, %	55.6	51.6	55.4	55.8	0.01	53.2	53.2	55.52	56.52	0.06
White, %	72.8	75.7	75.2	67.9	<0.01	69.8	75.6	74.2	72.2	<0.01
Education at V1, %										
Less than high school	21.8	21.8	22.9	30.0		22.0	22.9	24.9	26.6	
High school equivalent	41.1	40.9	43.2	40.6	<0.01	39.8	42.4	40.9	42.6	<0.01
Greater than high school	37.1	37.4	33.9	29.3		38.2	34.6	34.1	30.7	
BMI, kg/m ² (SD)	28.2 (5.1)	28.3 (5.1)	28.3 (5.6)	28.6 (6.3)	0.13	28.5 (5.3)	28.4 (5.4)	28.2 (5.5)	28.4 (5.9)	0.58
Smoking status, %										
Current smoker	22.3	23.2	22.4	24.2		21.6	23.3	23.6	23.4	
Former smoker	36.6	38.4	39.1	39.6	0.06	36.8	38.3	38.5	39.9	0.05
Never smoker	41.1	38.4	38.6	36.2		41.5	38.4	37.8	36.5	
Pack-years of smoking among ever smokers	28.5 (22.8)	30.4 (23.4)	29.9 (22.0)	32.5 (24.1)	<0.01	28.4 (22.3)	29.5 (23.1)	30.3 (23.0)	32.9 (23.7)	<0.01
Alcohol intake, %										
Current drinker	57.4	58.2	54.9	48.3		54.8	56.8	55.9	51.2	
Former drinker	21.6	20.9	20.8	26.5	<0.01	22.7	20.4	21.2	25.3	<0.01
Never drinker	20.9	20.9	24.3	25.2		22.5	22.6	22.7	23.5	
Physical activity at V1, score (SD)	2.5 (0.8)	2.5 (0.8)	2.4 (0.8)	2.3 (0.8)	<0.01	2.4 (0.8)	2.4 (0.8)	2.4 (0.8)	2.4 (0.8)	0.03
Aspirin use in the preceding two weeks, %	51.1	51.2	54.5	52.7	0.07	48.7	53.1	52.4	55.3	<0.01
diabetes, %	15.1	16.9	20.1	29.2	<0.01	17.0	16.4	20.2	27.8	<0.01
Hypertension, %	42.6	46.2	49.4	55.8	<0.01	44.8	48.4	47.4	53.3	<0.01
CVD, %	11.8	13.9	16.4	21.5	<0.01	11.8	13.6	15.5	22.7	<0.01
eGFR, mL/min/1.73 m ² (SD)	100.9 (14.3)	98.9 (15.1)	96.9 (15.8)	89.3 (21.8)	<0.01	100.2 (14.7)	98.7 (14.7)	96.4 (16.6)	90.7 (21.7)	<0.01

Abbreviations: V1 – Visit 1; V2 – Visit 2; PAC – proteomic aging clock; SD – standard deviation; BMI – body mass index; CVD – cardiovascular disease; eGFR – estimated glomerular filtration rate.

Table 3.3 The association between age acceleration for the V2 ARIC PAC and Lehallier's PAC and mortality

	No. of participants	No. of deaths	Total person-years	V2 ARIC PAC HR (95% CI) ^a per 1 SD (SD=2.94 years)	Lehallier's PAC HR (95% CI) ^a per 1 SD (SD=3.00 years)
All-cause mortality	8768	5294	187,784	1.38 (1.34, 1.43)	1.34 (1.30, 1.38)
CVD mortality	8768	1734	187,784	1.45 (1.37, 1.52)	1.39 (1.32, 1.46)
<i>CVD mortality (Fine and Grey model)</i>				<i>1.20 (1.14, 1.27)</i>	<i>1.19 (1.13, 1.25)</i>
Cancer mortality	8768	1516	187,784	1.19 (1.12, 1.27)	1.19 (1.12, 1.26)
<i>Cancer mortality (Fine and Grey model)</i>				<i>1.05 (0.98, 1.11)</i>	<i>1.05 (0.99, 1.12)</i>

Abbreviations: V1 – Visit 1; V2 – Visit 2; PAC – proteomic aging clock; BMI – body mass index; CVD – cardiovascular disease; eGFR - estimated glomerular filtration rate; SD – standard deviation; HR – Hazard ratio; CI – confidence interval.

^aModel adjusted for chronological age, gender, race, center, education (at V1), BMI, smoking status, pack-years of smoking, alcohol intake, physical activity (at V1), diabetes, eGFR, hypertension, and CVD at V2.

Table 3.4 V5 participant characteristics across quartiles of age acceleration for the V5 ARIC PAC

PAC acceleration	Q1 (N = 1138) -7.5 to -1.7	Q2 (N = 1138) -1.6 to -0.2	Q3 (N = 1139) -0.1 to 1.4	Q4 (N = 1138) 1.5 to 16.9	P-value
Chronological age, years (SD)	76.9 (5.1)	75.9 (5.0)	76.1 (5.3)	76.7 (5.4)	<0.01
Female, %	57.7	57.5	57.1	52.9	0.06
White, %	75.9	83.0	82.5	79.6	<0.01
Education at V1, %					
Less than high school	13.7	11.7	15.2	17.0	
High school equivalent	41.3	44.4	42.1	42.6	0.01
Greater than high school	44.9	43.8	42.7	40.4	
BMI, kg/m ² (SD)	29.1 (4.9)	28.8 (5.2)	28.6 (5.7)	28.6 (6.7)	0.15
Smoking status, %					
Current smoker	4.0	5.1	6.8	10.0	
Former smoker	54.5	56.8	50.1	50.8	<0.01
Never smoker	41.4	38.1	43.2	39.1	
Alcohol intake, %					
Current drinker	53.2	50.6	49.1	46.1	
Former drinker	28.3	30.8	28.4	31.0	0.01
Never drinker	18.5	18.6	22.5	22.8	
Physical activity, score (SD)	2.7 (0.8)	2.6 (0.8)	2.6 (0.8)	2.4 (0.8)	<0.01
Aspirin use in the preceding two weeks, %	68.8	69.8	71.0	73.0	0.14
diabetes, %	40.5	36.3	34.0	39.2	0.01
Hypertension, %	75.5	74.9	75.6	82.1	<0.01
CVD, %	19.9	24.7	29.8	38.4	<0.01
eGFR, mL/min/1.73 m ² (SD)	77.1 (15.7)	71.0 (16.0)	66.7 (17.8)	53.7 (19.3)	<0.01

Abbreviations: V1 – Visit 1; V5 – Visit 5; PAC – proteomic aging clock; SD – standard deviation; BMI – body mass index; CVD – cardiovascular disease; eGFR – estimated glomerular filtration rate.

Table 3.5 Association of V5 ARIC PAC acceleration and V2 to V5 change in PAC acceleration with mortality

V5 ARIC PAC acceleration				
	No. of participants	No. of deaths	Total person-years	HR (95%CI) ^a per 1 SD (SD = 2.61 years)
All-cause mortality	4553	1123	31,012	1.67 (1.56, 1.79)
CVD mortality	4553	348	31,012	1.54 (1.36, 1.75)
<i>CVD mortality (Fine and Grey model)</i>				<i>1.33 (1.18, 1.51)</i>
Cancer mortality	4553	278	31,012	1.33 (1.15, 1.53)
<i>Cancer mortality (Fine and Grey model)</i>				<i>1.19 (1.03, 1.37)</i>
V2 to V5 change in PAC acceleration				
	No. of participants	No. of deaths	Total person-years	HR (95%CI) per 1 SD (SD = 2.91 years)
All-cause mortality	2707	736	18,180	1.30 (1.19, 1.42)
CVD mortality	2707	239	18,180	1.20 (1.03, 1.40)
<i>CVD mortality (Fine and Grey model)</i>				<i>1.09 (0.94, 1.27)</i>
Cancer mortality	2707	172	18,180	1.17 (0.98, 1.40)
<i>Cancer mortality (Fine and Grey model)</i>				<i>1.10 (0.91, 1.33)</i>

Abbreviations: V1 – Visit 1; V2 – Visit 2; V5 – Visit 5; PAC – proteomic aging clock; SD – standard deviation; BMI – body mass index; CVD – cardiovascular disease; eGFR – estimated glomerular filtration rate. HR – Hazard ratio; CI – confidence interval.

^a Model adjusted for chronological age, gender, race, center, education (at V1), BMI, smoking status, alcohol intake, physical activity, diabetes, eGFR, hypertension, and CVD at V5.

Table 3.6 Association between participant characteristics at V2 and age acceleration for the V5 ARIC PAC

Characteristics at V2 ^a	Coefficients ^b	P-value or P-trend ^b
Chronological age	-0.04	<0.001
Male	0.07	0.47
Black	-0.77	<0.001
Education at V1		
Less than high school	0	
High school equivalent	-0.05	0.08
Greater than high school	0.16	
BMI	0.04	<0.0001
Smoking status		
Never smoker	0	
Former smoker	-0.27	<0.0001
Current smoker	0.26	
Pack-years of smoking	0.01	<0.0001
Alcohol intake		
Never drinker	0	
Former drinker	-0.16	0.06
Current drinker	-0.27	
Physical activity at V1	-0.06	0.26
Hypertension	0.44	<0.0001
CVD	0.40	0.02
Diabetes	1.06	<0.0001
eGFR	-0.03	<0.0001
Cognitive function		
Delayed Word Recall Test (DWRT)	-0.005	0.87
Digit Symbol Substitution Test (DSST)	-0.01	0.04
Word Fluency Test (WFT)	-0.01	0.001

Abbreviations: V1 – Visit 1; V2 - Visit 2; V5 – Visit 5; PAC – proteomic aging clock; BMI – body mass index; CVD – cardiovascular disease; eGFR – estimated glomerular filtration rate.

^a Characteristics were included into model simultaneously.

^b Coefficients and P-value (P-trend) were calculated from multiple linear regression.

Supplemental Table 3.1 Description of published proteomic aging clocks

Proteomic aging clock (PAC)	Study population	Chronological age	SomaScan assay	Number of aptamers reported in published PACs	Correlation between PAC and Chronological age reported in published papers	Number of aptamers available in ARIC among the aptamers reported in published PACs
Lehallier's PAC ¹⁴	3,301 healthy individuals from INTERVAL cohort (49% female; White individuals)	18-76	SomaScan v.3 assay (measured 2,978 proteins)	491	0.96	415 (85% all 491 aptamers)
Tanaka's PAC ¹³	240 healthy individuals from BLSA and GESTALT studies (50% female; White, Black, and other individuals)	22-93	SomaScan v.2 assay (measured 1,301 proteins)	76	0.94	68 (89% of all 76 aptamers)
Sathyan's PAC ⁴³	1,025 individuals from LonGenity study (55.7% female; White individuals)	65-95	SomaScan v.4 assay (measured 4,265 proteins)	162	0.79	162

Supplemental Table 3.2 Description of V2 ARIC PACs

PAC	Penalized regression	Transformation of proteins	Included square and cubic terms of each aptamer
V2 ARIC PAC	Elastic net regression (alpha=0.5)	Proteins were log2 transformed.	No
V2 ARIC PAC2	Lasso regression	Proteins were log2 transformed.	No
V2 ARIC PAC3	Elastic net regression (alpha=0.5)	Proteins were log2 transformed and were winsorized at 4 SD.	No
V2 ARIC PAC4	Elastic net regression (alpha=0.5)	Proteins were log2 transformed and then we applied inverse normal transformation.	No
V2 ARIC PAC5	Elastic net regression (alpha=0.5)	Proteins were log2 transformed.	Yes

Abbreviations: V2 - Visit 2; PAC – proteomic aging clock.

Supplemental Table 3.3 Pearson correlation coefficients between V2 ARIC PACs in the V2 test set

	V2 ARIC PAC	V2 ARIC PAC2	V2 ARIC PAC3	V2 ARIC PAC4	V2 ARIC PAC5
V2 ARIC PAC	1.00				
V2 ARIC PAC2	1.00	1.00			
V2 ARIC PAC3	0.99	0.99	1.00		
V2 ARIC PAC4	0.97	0.97	0.98	1.00	
V2 ARIC PAC5	0.99	0.99	0.99	0.98	1.00

Supplemental Table 3.4 Including/excluding participants with hypertension for healthy participants at V5 to construct PACs using elastic net regression

Different inclusion/exclusion for participants with hypertension ^a	Participants with controlled hypertension are included	Participants with hypertension are included	Participants with hypertension are excluded
No. of participants	945	1233	483
Training set size	630	822	322
Hyperparameter value (selected using 10-fold cross-validation)	0.47	0.36	0.64
Pearson correlation ^b in the training set	0.84	0.85	0.86
Pearson correlation ^b in the test set	0.71	0.72	0.64
MAE ^b in the training set	1.8	1.85	1.89
MAE ^b in the test set	2.23	2.32	2.47
HR (95% CI) ^c for all-cause mortality until 2019 per 1 SD increase in PAC acceleration ^d	1.67 (1.56, 1.79)	1.68 (1.57, 1.81)	1.66 (1.54, 1.78)
SD for PAC acceleration	2.61	2.67	2.47

Abbreviations: V5 – Visit 5; PAC – proteomic aging clock; BMI – body mass index; COPD – chronic obstructive pulmonary; CVD – cardiovascular disease; MAE – median absolute error; HR – hazard ratio; SD – standard deviation.

^a Hypertension was defined as diastolic blood pressure ≥ 90 mmHg, systolic blood pressure ≥ 140 mmHg, or taking medication for high blood pressure. Hypertension is controlled if the measured diastolic blood pressure is below 90 mmHg and measured systolic blood pressure is below 140 mmHg while the participant is on medication.⁷⁸

^b Pearson correlation and MAE between PAC and chronological age.

^c Model adjusted for chronological age, gender, race, center, education (at V1), BMI, smoking status, alcohol intake, physical activity, diabetes, eGFR, hypertension, and CVD at V5.

^d PAC was calculated as residuals after regressing PAC on chronological age.

Supplemental Table 3.5 Model R squared after regressing PAC acceleration on covariates at the corresponding visits

Regressing acceleration for the V2 ARIC PAC on V2 covariates	
Covariates at V2	R squared
Chronological age	0.0000
Chronological age + gender	0.0000
Chronological age + race, center	0.0039
Chronological age + education at V1	0.0074
Chronological age + BMI	0.0011
Chronological age + smoking, pack-year of smoking	0.0032
Chronological age + alcohol intake	0.0051
Chronological age + physical activity at V1	0.0064
Chronological age + aspirin use	0.0006
Chronological age + diabetes	0.0264
Chronological age + eGFR	0.0778
Chronological age + hypertension	0.0124
Chronological age + CVD	0.0166
Regressing acceleration for the V5 ARIC PAC on V5 covariates	
Covariates at V5	R squared
Chronological age	0.0000
Chronological age + gender	0.0023
Chronological age + race, center	0.0037
Chronological age + education at V1	0.0020
Chronological age + BMI	0.0016
Chronological age + smoking	0.0001
Chronological age + alcohol intake	0.0035
Chronological age + physical activity	0.0292
Chronological age + aspirin use	0.0013
Chronological age + diabetes	0.0001
Chronological age + eGFR	0.2033
Chronological age + hypertension	0.0046
Chronological age + CVD	0.0316

Abbreviations: V1 – Visit 1; V2 – Visit 2; V5 – Visit 5; PAC – proteomic aging clock; BMI – body mass index; CVD – cardiovascular disease; eGFR – estimated glomerular filtration rate.

Supplemental Table 3.6 Top 20 proteins with the largest absolute weights in the V2 ARIC PAC and V5 ARIC PAC

V2 ARIC PAC			
No.	Aptamer ID	Proteins	Weight
1	SeqId_8956_96	Scavenger receptor class F member 2	2.39462492
2	SeqId_14136_234	Complement component C1q receptor	-2.232162
3	SeqId_16890_37	ADAMTS-like protein 1	1.9584385
4	SeqId_15640_54	Transgelin	1.86624409
5	SeqId_3045_72	Pleiotrophin (PTN)	1.79477603
6	SeqId_3362_61	Chordin-like protein 1 (CHRD1)	1.76344784
7	SeqId_6392_7	WNT1-inducible-signaling pathway protein 2	1.71303183
8	SeqId_9793_145	Immunoglobulin superfamily DCC subclass member 4	-1.5215087
9	SeqId_13114_50	Lumican	1.3807798
10	SeqId_3331_8	RGM domain family member B	-1.1463758
11	SeqId_7179_69	Neurofascin	1.13848556
12	SeqId_11196_31	Collagen alpha-3(VI) chain	-1.1215024
13	SeqId_7551_33	Transforming growth factor beta activator LRRC32	1.11417136
14	SeqId_8974_172	Collagen alpha-1(XV) chain	-1.0493308
15	SeqId_3344_60	Antithrombin-III	-1.036693
16	SeqId_12417_46	EKC/KEOPS complex subunit TPRKB	1.02886089
17	SeqId_8841_65	Cartilage intermediate layer protein 2	-1.0021828
18	SeqId_4541_49	Cell adhesion molecule-related/down-regulated by oncogenes	-0.9964351
19	SeqId_9484_75	Desmoglein-2	-0.994248
20	SeqId_4374_45	Growth/differentiation factor 15 (GDF15)	0.97937678
V5 ARIC PAC			
No.	Aptamer ID	Proteins	Weight
1	SeqId_2677_1	Protein flightless-1 homolog	-2.522735
2	SeqId_15640_54	Transgelin	2.17188056
3	SeqId_6392_7	WNT1-inducible-signaling pathway protein 2	2.03908149
4	SeqId_3362_61	Chordin-like protein 1	1.48881636
5	SeqId_3045_72	Pleiotrophin (PTN)	1.39529172
6	SeqId_8974_172	Collagen alpha-1(XV) chain	-1.032661
7	SeqId_14136_234	Complement component C1q receptor	-0.9259469
8	SeqId_6081_52	Aldo-keto reductase family 1 member B10	-0.7204833
9	SeqId_19332_1	MOB-like protein phocein	0.66952444
10	SeqId_15434_5	Prepronociceptin	0.63560224
11	SeqId_11388_75	WAP four-disulfide core domain protein 2	0.60409029
12	SeqId_8480_29	EGF-containing fibulin-like extracellular matrix protein 1	0.59731828
13	SeqId_8304_50	Tumor necrosis factor receptor superfamily member 11B	0.56012362
14	SeqId_3327_27	Netrin-4	-0.5540319
15	SeqId_5731_1	Serine protease inhibitor Kazal-type 6	0.55237444
16	SeqId_9889_42	Actin filament-associated protein 1-like 1	-0.5422456
17	SeqId_5307_12	Coagulation factor IX	0.52424523
18	SeqId_7009_8	B-cell differentiation antigen CD72	-0.5158859
19	SeqId_9266_1	Transmembrane and ubiquitin-like domain-containing protein 2	0.49286769
20	SeqId_4496_60	Macrophage metalloelastase	0.48625879

Abbreviations: V2 – Visit 2; V5 – Visit 5; PAC – proteomic aging clock.

Supplemental Table 3.7 V2 participant characteristics across quartiles of age acceleration for Tanaka's PAC and Sathyan's PAC

	Tanaka's PAC					Sathyan's PAC				
	Q1 (N = 2192)	Q2 (N = 2192)	Q3 (N = 2192)	Q4 (N = 2192)	P- value	Q1 (N = 2192)	Q2 (N = 2192)	Q3 (N = 2192)	Q4 (N = 2192)	P- value
PAC acceleration	-16.7 to -2.0	-1.9 to -0.2	-0.1 to 1.7	1.8 to 28.4		- 14.0 to -2.2	2.1 to -0.2	-0.1 to 1.7	1.8 to 21.6	
Chronological age, years (SD)	58.3 (5.8)	58.1 (5.7)	57.6 (5.6)	58.2 (5.6)	<0.01	58.5 (5.6)	57.8 (5.7)	57.7 (5.8)	58.4 (5.7)	<0.01
Female, %	56.4	53.7	53.7	54.6	0.23	54.6	52.7	55.0	56.2	0.14
White, %	64.8	74.8	76.0	76.1	<0.01	69.6	75.4	75.0	71.7	<0.01
Education at V1, %										
Less than high school	23.6	23.2	22.7	26.9		21.7	22.3	24.2	28.3	
High school equivalent	39.2	42.0	42.7	41.8	<0.01	41.7	41.4	42.0	40.7	<0.01
Greater than high school	37.2	34.7	34.5	31.2		36.5	36.3	33.8	31.0	
BMI, kg/m ² (SD)	29.0 (5.7)	28.5 (5.3)	28.3 (5.4)	27.8 (5.6)	<0.01	28.3 (5.0)	28.4 (5.2)	28.4 (5.5)	28.4 (5.4)	0.86
Smoking status, %										
Current smoker	18.3	21.1	24.8	27.8		22.6	22.1	23.6	23.8	
Former smoker	36.5	40.2	38.5	38.5	<0.01	36.9	39.5	39.1	38.3	0.32
Never smoker	45.2	38.6	36.7	33.6		40.5	38.4	37.4	38.0	
Pack-years of smoking among ever smokers	27.8 (22.9)	28.7 (22.6)	31.1 (22.5)	33.2 (23.9)	<0.01	28.2 (21.8)	29.4 (22.2)	29.8 (23.4)	34.0 (24.6)	<0.01
Alcohol intake, %										
Current drinker	54.2	56.4	55.4	52.9		57.4	58.3	54.7	48.6	
Former drinker	22.1	21.7	21.6	24.3	0.17	21.0	21.9	22.3	24.5	<0.01
Never drinker	23.7	21.9	22.9	22.7		21.8	19.8	23.0	26.8	
Physical activity at V1, score (SD)	2.4 (0.8)	2.4 (0.8)	2.4 (0.8)	2.4 (0.8)	0.20	2.5 (0.8)	2.5 (0.8)	2.4 (0.8)	2.3 (0.8)	<0.01
Aspirin use in the preceding two weeks, %	47.6	52.2	52.1	57.7	<0.01	48.1	52.3	53.8	55.3	<0.01
diabetes, %	22.4	19.8	18.6	20.7	0.02	19.1	17.8	19.1	25.4	<0.01
Hypertension, %	47.9	47.6	47.4	51.2	0.04	43.9	46.7	48.9	54.6	<0.01
CVD, %	11.5	13.3	15.2	23.7	<0.01	10.7	15.1	15.4	22.6	<0.01
eGFR, mL/min/1.73 m ² (SD)	102.9 (14.1)	99.2 (14.9)	96.9 (15.1)	87.2 (21.4)	<0.01	101.7 (13.9)	99.4 (15.2)	96.5 (15.8)	88.5 (21.6)	<0.01

Abbreviations: V1 – Visit 1; V2 – Visit 2; PAC – proteomic aging clock; SD – standard deviation; BMI – body mass index; CVD – cardiovascular disease; eGFR – estimated glomerular filtration rate.

Supplemental Table 3.8 Association between age acceleration for Tanaka’s and Sathyan’s PACs and mortality

	No. of participants	No. of deaths	Total person-years	Tanaka’s PAC HR (95% CI) ^a per 1 SD (SD=3.15 years)	Sathyan’s PAC HR (95% CI) ^a per 1 SD (SD=3.25 years)
All-cause mortality	8768	5294	187,784	1.31 (1.27, 1.36)	1.38 (1.34, 1.43)
CVD mortality	8768	1734	187,784	1.34 (1.27, 1.41)	1.41 (1.34, 1.49)
<i>CVD mortality (Fine and Grey model)</i>				<i>1.14 (1.08, 1.21)</i>	<i>1.18 (1.12, 1.25)</i>
Cancer mortality	8768	1516	187,784	1.17 (1.10, 1.24)	1.23 (1.16, 1.31)
<i>Cancer mortality (Fine and Grey model)</i>				<i>1.05 (0.99, 1.12)</i>	<i>1.07 (1.01, 1.14)</i>

Abbreviations: V1 – Visit 1; V2 – Visit 2; PAC – proteomic aging clock; BMI – body mass index; CVD – cardiovascular disease; eGFR - estimated glomerular filtration rate; SD – standard deviation; HR – Hazard ratio; CI – confidence interval.

^aModel adjusted for chronological age, gender, race, study center, education (at V1), BMI, smoking status, pack-years of smoking, alcohol intake, physical activity (at V1), diabetes, eGFR, hypertension, and CVD at V2.

Supplemental Table 3.9 Association between acceleration for the V2 ARIC PAC and published PACs and mortality stratified by sex and race

<i>Stratified by sex^a</i>								
	Sex	No. of participants	No. of events	Total person-years	V2 ARIC PAC HR (95% CI) ^b per 1 SD ^c	Lehallier's PAC HR (95% CI) ^b per 1 SD ^c	Tanaka's PAC HR (95% CI) ^b per 1 SD ^c	Sathyan's PAC HR (95% CI) ^b per 1 SD ^c
All-cause mortality	Female	4789	2625	107,285	1.38 (1.32, 1.44)	1.35 (1.29, 1.40)	1.30 (1.25, 1.36)	1.40 (1.35, 1.46)
	Male	3979	2669	80,500	1.39 (1.33, 1.46)	1.35 (1.29, 1.41)	1.33 (1.27, 1.39)	1.38 (1.32, 1.44)
	P-interaction				0.61	0.98	0.79	0.38
CVD mortality	Female	4789	843	107,285	1.45 (1.34, 1.57)	1.37 (1.27, 1.47)	1.31 (1.21, 1.41)	1.42 (1.32, 1.53)
	Male	3979	891	80,500	1.46 (1.35, 1.57)	1.43 (1.33, 1.55)	1.38 (1.28, 1.49)	1.43 (1.33, 1.54)
	P-interaction				0.14	0.06	0.06	0.18
cancer mortality	Female	4789	721	107,285	1.15 (1.06, 1.25)	1.20 (1.10, 1.30)	1.15 (1.05, 1.25)	1.19 (1.09, 1.30)
	Male	3979	795	80,500	1.23 (1.13, 1.34)	1.17 (1.08, 1.27)	1.19 (1.10, 1.30)	1.26 (1.16, 1.37)
	P-interaction				0.67	0.26	0.83	0.92
<i>Stratified by race^a</i>								
	Race	No. of participants	No. of events	Total person-years	V2 ARIC PAC HR (95% CI) ^b per 1 SD ^d	Lehallier's PAC HR (95% CI) ^b per 1 SD ^d	Tanaka's PAC HR (95% CI) ^b per 1 SD ^d	Sathyan's PAC HR (95% CI) ^b per 1 SD ^d
All-cause mortality	White	6395	3859	139,615	1.37 (1.32, 1.41)	1.35 (1.30, 1.40)	1.31 (1.26, 1.36)	1.39 (1.34, 1.44)
	Black	2373	1435	48,169	1.41 (1.32, 1.50)	1.32 (1.24, 1.40)	1.31 (1.23, 1.39)	1.36 (1.28, 1.45)
	P-interaction				0.27	0.94	0.19	0.99
CVD mortality	White	6395	1168	139,615	1.48 (1.39, 1.57)	1.42 (1.34, 1.52)	1.37(1.28, 1.46)	1.45 (1.36, 1.54)
	Black	2373	566	48,169	1.39 (1.26, 1.53)	1.33 (1.22, 1.47)	1.28 (1.16, 1.41)	1.34 (1.22, 1.48)
	P-interaction				0.49	0.56	0.95	0.51
cancer mortality	White	6395	1138	139,615	1.19 (1.11, 1.27)	1.20 (1.13, 1.29)	1.18 (1.10, 1.26)	1.24 (1.16, 1.33)
	Black	2373	378	48,169	1.22 (1.07, 1.39)	1.14 (1.00, 1.29)	1.14 (0.99, 1.30)	1.20 (1.05, 1.36)
	P-interaction				0.86	0.43	0.90	0.53

Abbreviations: V1 – Visit 1; V2 - Visit 2; PAC – proteomic aging clock; SD – standard deviation; BMI – body mass index; CVD – cardiovascular disease; eGFR – estimated glomerular filtration rate; HR – hazard ratio; CI – confidence interval.

^a Interactions with sex and race were examined using a multiplicative term.

^b Model adjusted for chronological age, gender, race, study center, education (at V1), BMI, smoking status, pack-years of smoking, alcohol intake, physical activity (at V1), diabetes, eGFR, hypertension, and CVD at V2.

^c SDs for PAC acceleration across sex were: V2 ARIC PAC=3.02 and 2.85 years for females and males, respectively; Lehallier's PAC=3.04 and 2.95 years for females and males, respectively; Tanaka's PAC=3.17 and 3.13 years for females and males, respectively; and Sathyan's PAC=3.30 and 3.19 years for females and males, respectively.

^d SDs for PAC acceleration across race were: V2 ARIC PAC=2.75 and 3.40 years for White and Black participants, respectively; Lehallier's PAC=2.84 and 3.40 years for White and Black participants, respectively; Tanaka's PAC=2.95 and 3.61 years for White and Black participants, respectively; and Sathyan's PAC = 3.03 and 3.79 years for White and Black participants, respectively.

Supplemental Table 3.10 Association between age acceleration for the V5 ARIC PAC and mortality, stratified by sex and race

<i>Stratified by sex</i>				
	No. of participants	No. of events	Total person-years	HR (95% CI) ^b per 1 SD ^c
All-cause mortality				
Female	2565	536	17,809	1.72 (1.55, 1.91)
Male	1988	587	13,204	1.65 (1.50, 1.82)
P-interaction				0.16
CVD mortality				
Female	2565	163	17,809	1.44 (1.20, 1.74)
Male	1988	185	13,205	1.59 (1.34, 1.90)
P-interaction				0.58
Cancer mortality				
Female	2565	129	17,809	1.68 (1.38, 2.05)
Male	1988	149	13,204	1.06 (0.86, 1.30)
P-interaction				0.01
<i>Stratified by race</i>				
	No. of participants	No. of events	Total person-years	HR (95% CI) ^b per 1 SD ^d
All-cause mortality				
White	3655	935	24,949	1.68 (1.56, 1.82)
Black	898	188	6063	1.56 (1.29, 1.90)
P-interaction				0.78
CVD mortality				
White	3655	274	24,949	1.50 (1.30, 1.73)
Black	898	74	6063	1.83 (1.36, 2.46)
P-interaction				0.64
Cancer mortality				
White	3655	233	24,949	1.36 (1.17, 1.59)
Black	898	45	6063	0.99 (0.65, 1.49)
P-interaction				0.79

Abbreviations: V1 – Visit 1; V5 – Visit 5; PAC – proteomic aging clock; BMI – body mass index; CVD – cardiovascular disease; eGFR - estimated glomerular filtration rate; HR – hazard ratio; CI – confidence interval.

^a Interactions with sex and race were examined using a multiplicative term.

^b Model adjusted for chronological age, gender, race, study center, education (at V1), BMI, smoking status, alcohol intake, physical activity, diabetes, eGFR, hypertension, and CVD at V5.

^c SDs for V5 ARIC PAC acceleration across sex were 2.56 and 2.68 years for females and males, respectively.

^d SDs for V5 ARIC PAC acceleration across race were 2.50 and 3.03 years for White and Black participants, respectively.

Supplemental Table 3.11 Top 40 pathways overrepresented in the V2 ARIC PAC and V5 ARIC PAC

V2 ARIC PAC					
No.	Pathway	FDR P-value	No.	Pathway	FDR P-value
1	cell adhesion	7.575E-24	21	locomotion	1.732E-13
2	regulation of response to external stimulus	4.352E-19	22	protein phosphorylation	1.875E-13
3	positive regulation of phosphorylation	3.344E-16	23	positive regulation of cell population proliferation	1.875E-13
4	positive regulation of phosphorus metabolic process	3.344E-16	24	regulation of MAPK cascade	2.436E-13
5	positive regulation of phosphate metabolic process	3.344E-16	25	regulation of response to stress	3.434E-13
6	regulation of immune system process	5.875E-16	26	transmembrane receptor protein tyrosine kinase signaling pathway	3.434E-13
7	response to wounding	6.211E-16	27	cytokine production	4.74E-13
8	positive regulation of protein phosphorylation	6.448E-16	28	peptidyl-tyrosine phosphorylation	5.516E-13
9	regulation of phosphate metabolic process	1.928E-15	29	positive regulation of intracellular signal transduction	1.173E-12
10	regulation of phosphorus metabolic process	1.928E-15	30	regulation of cytokine production	1.173E-12
11	regulation of protein phosphorylation	2.26E-15	31	MAPK cascade	1.227E-12
12	positive regulation of MAPK cascade	2.26E-15	32	positive regulation of developmental process	1.292E-12
13	regulation of phosphorylation	3.115E-15	33	defense response to other organism	2.276E-12
14	positive regulation of signal transduction	4.365E-15	34	cell killing	3.395E-12
15	cell-cell adhesion	1.861E-14	35	regulation of programmed cell death	4.725E-12
16	negative regulation of multicellular organismal process	3.615E-14	36	positive regulation of immune system process	5.983E-12
17	peptidyl-tyrosine modification	6.535E-14	37	cell activation	6.623E-12
18	regulation of defense response	6.535E-14	38	regulation of locomotion	1.065E-11
19	enzyme-linked receptor protein signaling pathway	6.747E-14	39	regulation of multicellular organismal development	1.121E-11
20	positive regulation of multicellular organismal process	1.675E-13	40	innate immune response	1.152E-11
V5 ARIC PAC					
No.	Pathway	FDR P-value	No.	Pathway	FDR P-value
1	cell adhesion	5.137E-07	21	mononuclear cell migration	4.436E-04
2	vasculature development	2.848E-05	22	regulation of multicellular	4.447E-04

3	regulation of response to external stimulus	3.737E-05	23	organismal development cognition	5.401E-04
4	blood vessel development	5.79E-05	24	leukocyte chemotaxis	5.89E-04
5	leukocyte migration	7.459E-05	25	regulation of cell-substrate adhesion	9.524E-04
6	blood vessel morphogenesis	8.36E-05	26	cell chemotaxis	1.064E-03
7	circulatory system development	9.691E-05	27	granulocyte migration	1.2193E-03
8	regulation of chemotaxis	1.308E-04	28	regulation of immune system process	1.24E-03
9	negative regulation of multicellular organismal process	2.084E-04	29	cell-matrix adhesion	1.301E-03
10	tube development	2.099E-04	30	locomotion	1.322E-03
11	cell-substrate adhesion	2.099E-04	31	positive regulation of multicellular organismal process	1.418E-03
12	inflammatory response	2.347E-04	32	behavior	2.244E-03
13	learning or memory	2.432E-04	33	positive regulation of macrophage chemotaxis	2.244E-03
14	myeloid leukocyte migration	2.432E-04	34	positive regulation of chemotaxis	2.244E-03
15	response to wounding	3.016E-04	35	regulation of locomotion	2.242E-03
16	cellular extravasation	3.218E-04	36	response to inorganic substance	2.63E-03
17	tube morphogenesis	3.218E-04	37	cell-cell adhesion	2.702E-03
18	taxis	3.218E-04	38	post-embryonic eye morphogenesis	3.753E-03
19	chemotaxis	3.218E-04	39	regulation of cell-matrix adhesion	3.753E-03
20	negative regulation of response to external stimulus	3.241E-04	40	angiogenesis	3.753E-03

Abbreviations: V2 - Visit 2; V5 – Visit 5; PAC – proteomic aging clock.

Supplemental Table 3.12 Association between V2 to V5 change in PAC age acceleration and mortality, stratified by sex and race

Stratified by sex ^a				
	No. of participants	No. of events	Total person-years	HR (95% CI) ^b per 1 SD ^c
All-cause mortality				
Female	1531	372	10437	1.20 (1.06, 1.37)
Male	1176	365	7744	1.46 (1.25, 1.60)
P-interaction				0.13
CVD mortality				
Female	1531	121	10437	1.14 (0.92, 1.43)
Male	1176	121	7744	1.26 (1.02, 1.56)
P-interaction				0.96
Cancer mortality				
Female	1531	87	10437	1.25 (0.98, 1.52)
Male	1176	85	7744	1.18 (0.91, 1.52)
P-interaction				0.87
Stratified by race ^a				
	No. of participants	No. of events	Total person-years	HR (95% CI) ^b per 1 SD ^d
All-cause mortality				
White	2103	600	14158	1.33 (1.20, 1.47)
Black	604	136	4022	1.15 (0.93, 1.41)
P-interaction				0.68
CVD mortality				
White	2103	188	14158	1.26 (1.05, 1.51)
Black	604	54	4022	1.02 (0.73, 1.42)
P-interaction				0.30
Cancer mortality				
White	2103	141	14158	1.21 (0.99, 1.46)
Black	604	31	4022	0.98 (0.61, 1.57)
P-interaction				0.75

Abbreviations: V2 - Visit 2; V5 – Visit 5; PAC – proteomic aging clock; BMI – body mass index; CVD – cardiovascular disease; eGFR – estimated glomerular filtration rate; SD – standard deviation; HR – hazard ratio; CI – confidence interval.

^a Interaction with sex and race were examined using a multiplicative term.

^b Model adjusted for chronological age, gender, race, center, education (at Visit 1), BMI, smoking status, alcohol intake, physical activity, diabetes, eGFR, hypertension, and CVD at Visit 5.

^c SDs for the V2 to V5 change in acceleration across sex were 2.90 and 2.91 years for females and males, respectively.

^d SDs for the V2 to V5 change in acceleration across race were 2.81 and 3.23 years for White and Black participants, respectively.

F. Supplemental material

The assessment of cancer, diabetes, hypertension, abnormal kidney function, chronic obstructive pulmonary disease (COPD), and cardiovascular diseases (CVD) events, and other characteristics of interest and the procedures for identifying healthy participants are described below.

Ascertainment of cancer cases

Prevalent cancers at Visit 1 were defined as a self-reported previous diagnosis of cancer. Incident cancers developed after Visit 1 were ascertained through 2015 via linkage with state cancer registries in Minnesota, North Carolina, Maryland, and Mississippi. These records were supplemented by the abstraction of medical records for hospitalizations and review of hospital discharge codes.⁵³

Ascertainment of diabetes, hypertension, and abnormal kidney function

Diabetes mellitus was defined as fasting glucose ≥ 126 mg/dL, non-fasting glucose ≥ 200 mg/dL, self-reported medication to treat diabetes, or self-reported physician diagnosis of diabetes.

Hypertension was defined as diastolic blood pressure ≥ 90 mmHg, systolic blood pressure ≥ 140 mmHg, or taking medication for high blood pressure. Hypertension was classified as controlled if the measured diastolic blood pressure was below 90 mmHg and measured systolic blood pressure was below 140 mmHg while the participant was on medication.⁷⁸

Abnormal kidney function was defined as an estimated glomerular filtration rate (eGFR) < 60 mL/min/1.73 m². eGFR was calculated based equations that utilized serum creatinine and incorporated age and sex.⁷⁹

Ascertainment of COPD

COPD was defined as having a forced expiratory volume in 1 second (FEV₁) <80% of the predicted value and a FEV₁/forced vital capacity (FVC) ratio of <70% based on spirometry. Participants meeting only one of the two criteria for COPD were categorized in the no COPD referent group for analysis.

FEV₁, which is the volume of gas exhaled in the first second of expiration, and FVC, the total volume of gas exhaled, were measured according to the American Thoracic Society criteria.⁸⁰ Collins Survey II water-seal spirometers (Collins Medical, Inc., Braintree, MA) driven by IBM PC/XT computers and under the control of Pulmo-Screen software (PDS Healthcare Products, Inc., Louisville, CO) were used to assist the technicians with quality control, calculation of pulmonary function variables, and compilation of results for transmission to the ARIC Pulmonary Function Reading Center. Participants performed the FVC maneuver until there were two error-free reproducible maneuvers (FEV₁ and FVC within 5%) out of three acceptable maneuvers, with the maneuvers repeated up to eight times if necessary.

FEV₁ predicted was calculated as follows for men and women:⁸¹

$$\textbf{Men: } FEV_1 \text{ predicted} = 0.0414(\text{height in cm}) - 0.0244(\text{age}) - 2.190$$

$$\textbf{Women: } FEV_1 \text{ predicted} = 0.0342(\text{height in cm}) - 0.0255(\text{age}) - 1.578$$

Because Black individuals have lower lung function than White individuals,⁸² the FEV₁ predictions were multiplied by 0.88 for Black participants in order to adjust the above lung function equations (developed for White individuals) for use in Black individuals.^{83,84} The percentage of predicted values for FEV₁ for a person was calculated by dividing measured FEV₁ by predicted FEV₁.

Ascertainment of CVD events

CVD events included 1) heart failure (HF), 2) definite or probable stroke, or 3) coronary heart disease (CHD), defined as definite or probable myocardial infarction (MI) or definite fatal

CHD.^{85,86}

Pre-existing HF at Visit 1 was defined as 1) an affirmative response to “Were any of the medications you took during the last 2 weeks for heart failure?” or 2) stage 3 or “manifest heart failure” by Gothenburg criteria.^{87,88} Pre-existing CHD at Visit 1 was defined as 1) a self-reported previous physician diagnosis of MI or coronary revascularization or 2) as prevalent MI by 12-lead electrocardiography. Pre-existing stroke at Visit 1 was defined by any self-reported previous physician diagnosis of stroke.^{85,86}

CVD events that developed after Visit 1 were identified via annual interviews, study visits, and community-wide surveillance of hospitalization discharge listings and validated via physician review. Incident CHD was defined as definite or probable MI, fatal CHD, or coronary revascularization. Definite or probable HF was defined as the first occurrence of hospitalization of heart failure via hospital discharge code.⁸⁹ Definite or probable stroke was identified via a computer algorithm based on criteria adapted from the National Survey of Stroke,⁹⁰ which utilized classification, signs, symptoms, neuroimaging, and other diagnostic reports.⁹¹

Assessment of other characteristics of interest

At each visit, participants reported information on smoking, alcohol intake, and use of medications, and underwent a physical exam that included measurements of height and weight. Body mass index (BMI) was calculated as weight (kg) divided by height (in meters) squared. Pack-years of smoking at Visit 1 was derived by multiplying the average cigarettes smoked per day times the number of years smoked. Pack-years smoked at Visit 2 was calculated by combining pack-year estimates from Visit 1 with smoking status estimates from Visit 2 (e.g., for current smokers at Visits 1 and 2, the average time between visits (3 years) was multiplied by the number of daily cigarettes reported at Visit 1 and added to Visit 1 pack-years).⁹² Physical activity at Visit 1 and Visit 5 was assessed using a modified version of the Baecke questionnaire from which leisure time physical activity was derived as a function of reported intensity and time spent

on sports and exercise. Cognitive function was assessed using three neuropsychological tests: the Delayed Word Recall Test (DWRT), the Digit Symbol Substitution Test (DSST) of the Wechsler Adult Intelligence Scale–Revised, and the Word Fluency Test (WFT). Protocols for the tests were standardized, and trained examiners administered the tests in a fixed order during one session in a quiet room.⁹³ The DWRT evaluates verbal learning and short-term memory.⁹⁴ Participants learn 10 nouns, use them in sentences, and, after 5 minutes, are asked to recall them. The score of DWRT is the number of nouns recalled (maximum of 10). The DSST evaluates executive function and processing speed.⁹⁵ Participants use a key to write symbols corresponding to numbers in 90 seconds. The score of DSST, ranging from 0 to 93, is the number of correctly written symbols. The WFT evaluates executive function and expressive language.⁹⁶ Participants generate as many words as possible within 60 seconds, starting with F, A, and S, with 1 trial per letter. The score of WFT is the sum of all correct words generated.

Identifying healthy participants at Visit 2

Among 14,348 participants who attended Visit 2, we excluded participants without protein measures at Visit 2 (N=2,550) and participants whose race was other than White or Black (N=37), resulting in 11,761 participants. To restrict to healthy participants, we additionally excluded those who reported a history of cancer (N = 927), met the definition of diabetes (N = 1,857), met the definition of hypertension (N = 3,311), had abnormal kidney function (eGFR < 60 mL/min/1.73 m²) (N = 43), met the definition of COPD (N = 841), or had a history of CVD (N = 293), leaving 4,489 healthy participants. There were 7,272 participants with any major diseases or health conditions.

Identifying healthy participants at Visit 5

Among 6,538 participants who attended Visit 5, we excluded participants without protein measures at Visit 5 (N=1,343) and participants whose race was other than White or Black

(N=12), resulting in 5,183 participants⁵. To restrict to healthy participants, we additionally excluded those with cancer (N = 1,211), diabetes (N = 1,247), uncontrolled hypertension (N = 1,595), abnormal kidney function (N = 385), COPD (N = 537), and CVD (N = 154), leaving in 945 healthy participants. There were 4,238 participants with any major diseases or health conditions.

Chapter 4 Manuscript 2: Proteomic Aging Clock and Risk of Cancer in the Atherosclerosis Risk in Communities (ARIC) Study

A. Overview

Background: Individuals with and without cancer may have a different pace of physiological dysregulation, i.e., they have different biological aging. Aging clocks have been proposed as a measure of biological aging, including a novel proteomic aging clock (PAC). The deviation of a PAC from chronological age is called PAC acceleration. PACs have been linked to age-related diseases but have not been examined in relation to cancer risk. We examined the association between PAC acceleration for a newly constructed cancer-specific PAC (CaPAC) and a published PAC proposed by Lehallier [2020] (so called Lehallier's PAC) and cancer risk in the Atherosclerosis Risk in Communities (ARIC) study.

Methods: ARIC is a prospective cohort of White and Black women and men initiated in 1987 and followed for cancer until 2015. An aptamer-based platform was used to measure more than 5000 proteins in plasma collected at Visit 2 (V2, 1990-92). Of 10,834 participants, 3,347 developed cancer and 7,487 remained cancer-free through 2015. Using elastic net regression, we constructed the CaPAC in a training set randomly selected from participants who remained cancer-free ($N = 4,991$). The CaPAC included 1,282 aptamers and was internally validated in the remaining participants who stayed cancer-free (test set). In addition, we also calculated Lehallier's PAC. In the test set, we calculated PAC acceleration as the residual for each participant after regressing PAC on chronological age. We used Cox proportional hazards regression to estimate hazard ratios

(HRs) and 95% confidence intervals (CIs) for incident cancer overall, and for the most common cancers (prostate, lung, breast, colorectal, bladder, kidney, and pancreatic cancers).

Results: In the test set, both the CaPAC and Lehallier's PAC were correlated with chronological age [Pearson correlation coefficient (r) = 0.82 for CaPAC; and r = 0.76 for Lehallier's PAC] and with each other [r = 0.90]. Age acceleration for CaPAC [HR (95% CI) per 1 SD = 1.04 (1.00, 1.08)], but not for Lehallier's PAC, was significantly associated with overall cancer risk. Age acceleration for both PACs was similarly associated with lung cancer risk [HR (95% CI) per 1 SD = 1.23 (1.12, 1.36) for CaPAC acceleration], and these associations remained positive among never smokers. In addition, CaPAC acceleration was associated with colorectal cancer risk [HR (95% CI) per 1 SD = 1.15 (1.02, 1.30)]. Neither PAC was associated with breast, prostate, kidney, bladder, or pancreatic cancer risk.

Conclusion: PACs hold promise as potential biomarkers for cancer risk.

B. Introduction

Aging plays a critical role in cancer, as indicated by the fact that the incidence of most cancer types markedly increases with age.^{97,98} Growing evidence shows that individuals with cancer develop physiological dysregulation at an earlier chronological age than those without cancer,⁹⁹ suggesting that the biological age of individuals with cancer is higher than their chronological age. Biological age, according to the definition

proposed by Baker and Sprott, is characterized by the “biological parameter[s] of an organism, either alone or in some multivariate composite that will, in the absence of disease, better predict functional capability at some late age than will chronological age”.⁶²

To estimate individual’s biological age, researchers have proposed biological age estimators called aging clocks, using epigenetics, transcriptomics, metabolomics, proteomics, and other biomarkers.¹² A positive difference between an individual’s biological age estimated by aging clocks and chronological age is called age acceleration. The most studied aging clocks were developed using DNA methylation data, including Horvath, Hannum, DNAm PhenoAge, and GrimAge clocks.¹⁷⁻²⁰ Greater age acceleration for these epigenetic clocks has been associated with an increased risk of different types of cancer, although the associations for colorectal, kidney, and pancreatic cancers were inconsistent when using different epigenetic clocks.^{19,20,25,32,33,35,36,38,100,101} However, the underlying mechanisms of changes in these DNA methylation sites remain unclear.⁶³ Proteomic aging clocks (PACs) also hold promise because proteomic-based biomarkers may be intermediate phenotypes between genetics and disease, and may provide more information on aging and age-related pathology.^{13,40} In addition, proteins serve as a target in 96% of FDA approved medications.⁴¹ Therefore, the proteins comprising PACs hold promise as targets of anti-aging drugs.

Several PACs have been proposed, including PACs developed by Lehallier,¹⁴ Tanaka,¹³ and Sathyan.⁴³ Previous findings showed that PACs could identify individuals with age acceleration and predict their future risk of age-related conditions.^{14,29,43}

However, to our knowledge, no studies examined PACs in relation to cancer risk. In addition, previous published PACs were not cancer-specific.

In this study, we constructed a cancer-specific PAC (CaPAC) using data from the Atherosclerosis Risk in Communities (ARIC) study, which recently measured more than 5000 proteins using SomaScan assay (v.4). We hypothesized that the CaPAC would be positively associated with risk of cancer. We also tested if the PAC developed by Lehallier [2020] (so called Lehallier's PAC) and the CaPAC predict risk of cancer in similar way.

C. Methods

C.1 Study population

The ARIC study is a prospective cohort initiated in 1987.^{52,53} In 1987-1989 (Visit 1, V1), 15,792 participants aged 45-64 years were recruited from four study centers - Maryland; Minnesota; Mississippi; and North Carolina. Participants in Minnesota and Maryland were primarily White, and the recruitment of Mississippi was restricted to Black residents. The ARIC study was approved by institutional review boards at each participating center, and all study participants provided written informed consent. Thus far, eight study visits have been completed.⁵² Additionally, participants underwent follow-up via telephone calls annually between 1987-2012 and semi-annually after 2012, with response rates of 90%-99% for the annual follow-up calls and 83%-90% for semi-annual follow-up calls among living participants who have not withdrawn consent to be contacted.⁵³

C.2 Ascertainment of cancer cases

Incident cancer cases were ascertained through 2015 via linkage with state cancer registries in Maryland, Minnesota, Mississippi, and North Carolina. These records were supplemented by the abstraction of medical records and hospital discharge codes.⁵³

C.3 Blood collection

The ARIC protocol for blood sample collection, processing, and storage protocol was designed to minimize the spontaneous biochemical reactions after blood collection, and is consistent with recommended practice for proteomics data analysis in epidemiological studies.^{58,64,65} Briefly, after venipuncture, blood samples were put immediately in an ice water bath. Centrifugation was then performed within 10 min after venipuncture at room temperature (15-25 °C). After centrifugation, the aliquots were stored at -80 °C within 90 min from venipuncture and were never thawed before this analysis.

C.4 Protein measurement and quality control

For this study, we used proteins measured in EDTA-plasma samples collected at Visit 2 (V2, 1990-92). Plasma samples collected at Visit 3 (V3, 1993-95) and Visit 5 (V5, 2011-13) were used for sensitivity and exploratory analyses. Samples were analyzed using a Slow Off-rate Modified Aptamers (SOMAmer)--based capture array called SomaScan by Somalogic, Inc. (Boulder, CO, USA).⁵⁴⁻⁵⁷ The SomaScan platform uses single-stranded DNA-based aptamers to capture conformational protein epitopes.

Protein analyte measurements underwent the regular SomaScan data

standardization and normalization process.^{58,59} First, hybridization control normalization was applied to each sample based on a set of hybridization control sequences to correct for systematic biases during hybridization. Second, median signal normalization was applied to measures within a plate to remove sample or assay biases that may be because of pipetting variation, variation in reagent concentrations, assay timing, and other sources of systematic variability within a single plate run. Finally, each plate contained calibrator samples for each SOMAmer reagent, which was used to correct for plate-to-plate variation based on established global reference standards. The median Bland-Altman coefficient of variation (CVBA) for split samples was 6% at V2 (12% at V3 and 7% at V5) after excluding aptamers with a CVBA >50% or a variance of <0.01 on the log scale, or aptamers nonspecific binding to nonproteins. After the exclusion, 4955 aptamers (at V2, V3, and V5) were included that corresponded to 4,712 proteins. Protein measures were reported as relative fluorescent units (RFU) and were log₂ transformed to correct for skewness.

C.5 Assessment of other participant characteristics

Other characteristics of interest included demographic and lifestyle/medical characteristics, namely chronological age, sex, race, study center, education, smoking status, pack-years of smoking, alcohol intake, body mass index (BMI), hormone replacement therapy (for females only), aspirin use, diabetes status, and estimated glomerular filtration rate (eGFR). Sex, race, study center, and education was collected at V1, chronological age was calculated from date of birth until date of V2, and all the other variables listed above were collected at V1 - V3.⁷⁶ At each visit, participants reported

information on smoking history, alcohol intake, and use of medications and underwent a physical exam that included measurement of height and weight. BMI was calculated as weight (kg) divided by height (in meters) squared. Diabetes mellitus was defined as fasting glucose ≥ 126 mg/dL, non-fasting glucose ≥ 200 mg/dL, self-reported use of medication to treat for diabetes, or self-reported physician diagnosis of diabetes. Detailed procedures for estimation of pack-years of smoking have been published.⁹² eGFR at V2 was calculated based on equations that utilized serum creatinine and cystatin C and incorporated age and sex.⁷⁹

C.6 Statistical analysis

Proteomic aging clocks (PACs) were constructed using R (version 4.1.2, package “glmnet”) and all the other analyses were performed using SAS version 9.4 (SAS Institute Inc, Cary, NC).

Among 14,348 participants who attended V2 (chronological age: 46-70 years), we excluded participants without protein measures (N=2,550), participants with race other than White or Black (N=37), and participants who reported a history of cancer at V2 (N=927), leaving 10,834 participants for analysis. Among them, 3,347 developed cancer by the end of 2015 (median follow-up = 17.19 years), and 7,487 remained cancer-free (**Figure 4.1**).

C.6.1 Construction and validation of PACs

Among the 7,487 participants who remained cancer-free, we randomly selected two-thirds (N=4,991) and used them as the training set. The remaining one-third were

used as the test set (N=2,496) (**Figure 4.1**). Using the training set, we applied elastic net regression to train the CaPAC (**Supplemental Table 4.1**) against chronological age. The elastic net regression (alpha = 0.5 and hyperparameter value was selected based on 10-fold cross-validation) selected 1,282 aptamers out of the 4,955 aptamers.

In addition to the newly constructed CaPAC, we also computed Lehallier's PAC,¹⁴ one of the published PACs. We chose Lehallier's PAC rather than PACs developed by Tanaka and Sathyan,^{13,43} because Lehallier's PAC was more strongly correlated with chronological age ($r=0.96$) compared to other published PACs. We estimated ARIC weights for Lehallier's PAC instead of using the published weights because ARIC did not include all the aptamers reported in Lehallier's PAC. Using the 415 aptamers available in ARIC out of the 491 aptamers reported by Lehallier et al [2020],¹⁴ we estimated ARIC weights by applying Ridge regression (hyperparameter value was selected based on 10-fold cross-validation).

In the test set, we calculated PACs for each of the participant using the weighted sum of their aptamer levels: $PAC = \beta_0 + \sum_{i=1}^n \beta_i \times aptamer_i$, where $aptamer_i$ is the level of i th aptamer and the intercept (β_0) and weights (β_i) were estimated using the training set. We estimated the performance of each PAC by measuring the Pearson correlation between PAC and chronological age in the test set. In addition, we also calculated the median absolute error (MAE) between each PAC and chronological age in the test set.

C.6.2 Association between participant characteristics and PACs

To examine the cross-sectional association between participant characteristics at

V2 and the PACs, we computed the CaPAC and Lehallier's PAC in 5,843 White and Black participants without a history of cancer at V2. To capture associations that are independent of chronological age, we examined associations with PAC acceleration (for CaPAC and Lehallier's PAC).³⁵ PAC acceleration was estimated as the residual for each participant after regressing the PAC on chronological age. Demographic and lifestyle/medical characteristics at V2 were examined across quartiles of PAC acceleration.

C.6.3 Association with risk of cancer

We used Cox proportional hazards regression to estimate hazard ratios (HRs) and 95% confidence intervals (CIs) for the association between PAC acceleration (modeled as a continuous variable) and risk of cancer overall, as well as the most common cancers (prostate, lung, breast, colorectal, bladder, kidney, and pancreatic cancers). For each participant, total person-years was determined from date of V2 until diagnosis of cancer, death, loss to follow-up, or administrative censoring on December 31, 2015, whichever occurred first. The proportional hazards assumption was examined by modeling an interaction between PAC acceleration and follow-up time, and the assumption was not violated in any regression models. We constructed three models. Model 1 was adjusted for chronological age, sex, race, and study center. Model 2 was additionally adjusted for education, BMI, smoking status, pack-years of smoking, alcohol intake, and eGFR. Model 3 (fully adjusted model) was additionally adjusted for hormone replacement therapy (in women), aspirin use, and diabetes status.

We also examined whether sex, race, or smoking status at V2 modified the

association between PAC acceleration and overall cancer risk and risk of most common cancers, because males⁷⁵ and smokers²⁰ might age faster and smokers and Black individuals had a higher incidence of several cancer types.¹⁰² Interaction of PAC acceleration with sex, race, and smoking status were examined using a multiplicative term.

C.6.4 Exploratory analysis

All the previous studies of PACs trained their PACs against chronological age using a cross-sectional design. Cross-sectional studies measure protein levels and chronological age in different individuals, typically across a wide range of chronological age. However, pace of aging could be different between individuals and may be more accurately reflected by within-individual changes in protein levels captured in longitudinal designs.²² In an exploratory analysis, we constructed another cancer-specific PAC -- CaPAC2 (**Supplemental Table 4.1**) based on the top 2000 proteins with the largest within-individual changes in blood levels between V2 and V5 (about 20 years). Among 3,220 who remained cancer-free until 2015 and had proteomics data at both V2 and V5, we randomly selected two-thirds of participants (N=2,147) and used them as the training set² and the remaining one-third of participants were used as the test set² (N=1,073). Using training set², we calculated the mean and SD for the within-individual changes for all the 4,955 aptamers. We ranked aptamers based on their mean/SD for within-individual changes and selected the top 2,000 aptamers. Using the selected top 2000 aptamers as predictors, we applied elastic net regression to train CaPAC2 against chronological age at V2. Elastic net regression selected 420 aptamers out of the top 2000

aptamers with largest within-individual changes. We used the test set2 to validate the performance of CaPAC2. We also examined the association between CaPAC2 acceleration and cancer incidence. In another exploratory analysis, we compared CaPAC and CaPAC2 to the V2 ARIC PAC (developed in healthy participants) (**Supplemental Table 4.1**), which we described in Manuscript 1. We tested if these PACs showed similar association with risk of cancer. In the last exploratory analysis, we used ToppGene⁶⁸⁻⁷⁰ to conduct an overrepresentation analysis (ORA) for the proteins included in the CaPAC. ORA is a type of pathway analysis that determines whether proteins from pre-defined pathways, for example, those participating in the same molecular process or regulated by the same transcription factor (i.e., to a specific GO term or KEGG pathways), are present more than would be expected (overrepresented) in our data.⁷¹ UniProt IDs were provided as the inputs, the background was set to all protein-coding genes, and the False Discovery Rate (FDR) significance level was set to 0.05.

C.6.5 Sensitivity analysis

In a sensitivity analysis, we constructed another PAC (CaPAC3) using proteomic data measured at V3 (**Supplemental Table 4.1**). Among the 10,396 participants who were cancer-free at V3, 3,092 participants developed cancer by the end of 2015 and 7,304 participants remained cancer-free. Using a set randomly selected from participants who remained cancer-free until 2015 (N = 5,000), we constructed CaPAC3 by applying elastic net regression. We then applied CaPAC3 to the 5,396 participants who were cancer-free at V3 (after excluding the 5000 participants who used to constructed CaPAC3) and repeated the main analyses using CaPAC3 acceleration instead of CaPAC acceleration.

For this sensitivity analysis, follow of participants started at V3.

D. Results

D.1 Pearson correlation coefficients between PACs and chronological age

The Pearson correlation coefficients between the CaPAC and chronological age were 0.92 (median absolute error (MAE) = 1.57 years) and 0.82 (MAE = 2.14 years), respectively, in the training set and test set (**Figure 4.2a**). For Lehallier's PAC, the correlations with chronological age were slightly lower: 0.81 (MAE = 2.33 years) and 0.76 (MAE = 2.52 years), respectively, in the training and test sets (**Figure 4.2b**). The correlation between CaPAC and Lehallier's PAC was 0.90 in the test set.

D.2 Distributions of participant characteristics across quartiles of PAC acceleration

Among the 5,843 White and Black participants without a history of cancer at V2, those with higher age acceleration for CaPAC and Lehallier's PAC were more likely to be never drinkers, have diabetes, less education, and a lower mean eGFR (**Table 4.1**). In addition, participants with higher age acceleration for Lehallier's PAC were more likely to be ever smokers (**Table 4.1**).

D.3 Associations of PAC acceleration with risk of cancer

We found that CaPAC acceleration was statistically significantly associated with risk of overall cancer. In the Model 3, HR (95% CI) per 1 SD were 1.04 (1.00, 1.08) (Model 3, **Table 4.2**). This association was not modified by sex, race, or smoking status (**Table 4.3 and Supplemental Table 4.2**). By contrast, the association between

Lehallier's PAC acceleration and overall cancer was not significant in the fully adjusted model (Model 3) [HR (95% CI) per 1 SD=1.02 (0.98, 1.06)] (**Table 4.2**). However, age acceleration for Lehallier's PAC appeared to be associated with higher overall cancer risk in Black participants, but not in White participants and there was a suggestion of interaction with race [p-interaction=0.08] (**Table 4.3**).

We found that higher age acceleration for CaPAC and Lehallier's PAC was statistically significantly associated with an increased lung cancer risk [HRs (95% CIs) per 1 SD = 1.24 (1.12, 1.36) for CaPAC and 1.20 (1.09, 1.32) for Lehallier's PAC] (Model 3, **Table 4.2**). These associations remained positive among never smokers [HRs (95% CIs) per 1 SD = 1.42 (0.96, 2.10) for CaPAC and 1.40 (0.95, 2.01) for Lehallier's PAC] (**Supplemental Table 4.2**). There was no association between age acceleration for CaPAC or Lehallier's PAC and kidney cancer risk (**Table 4.2**). However, higher age acceleration for CaPAC and Lehallier's PAC appeared to be associated with kidney cancer risk in Black participants, but not in White participants [p-interactions = 0.05 and 0.07, for CaPAC and Lehallier's PAC, respectively] (**Table 4.3**). We also found that CaPAC acceleration was significantly associated with colorectal cancer (CRC) risk [HR (95% CI) per 1 SD = 1.15 (1.02, 1.30) (Model 3, **Table 4.2**). This association was not statistically modified by sex or race (**Table 4.3**), but there was a suggestion of interaction with smoking status for this association (p-interaction = 0.09) (**Supplemental Table 4.2**). Age acceleration for CaPAC and Lehallier's PAC was not associated with risk of prostate, postmenopausal breast, bladder, or pancreatic cancer (**Table 4.2**).

D.4 Exploratory analysis

For CaPAC2, Pearson correlation coefficients with chronological age were 0.85 (MAE = 1.85 years) and 0.75 (MAE = 2.31 years), respectively, in the training set2 and test set2 (**Supplemental Figure 4.1**). In participants who remained cancer-free until 2015 (N=1,775, after excluding training sets for CaPAC and CaPAC2), the Pearson correlation coefficient between CaPAC and CaPAC2 was 0.90.

Among the 8,687 White and Black participants without a history of cancer at V2, those with higher CaPAC2 acceleration were more likely to be never drinkers, aspirin users, have diabetes, less education, and a lower mean eGFR (**Supplemental Table 4.3**). CaPAC2 acceleration was positively associated with risk of lung cancer [HR (95% CI) per 1 SD=1.24 (1.12, 1.38)], but negatively associated with risk of postmenopausal breast cancer [HR (95% CI) per 1 SD=0.83 (0.76, 0.94)] in the fully adjusted model (**Table 4.4**). These associations remained significant among never smokers [HR (95% CI) per 1 SD=1.71 (1.25, 2.54) for lung cancer risk and 0.77 (0.67, 0.89) for postmenopausal breast cancer risk] (**Supplemental Table 4.5**). In addition, CaPAC2 acceleration was negatively associated with risk of bladder cancer [HR (95% CI) per 1 SD=0.82 (0.68, 0.99)] (**Table 4.4**). There were no associations of CaPAC2 acceleration with overall cancer or kidney cancer risk overall (**Table 4.4**). However, CaPAC2 acceleration appeared to be associated with risk of overall cancer and kidney cancer in Black participants, but not White participants [p-interactions<0.05] (**Supplemental Table 4.4**). CaPAC2 acceleration was not associated with risk of prostate, colorectal, or pancreatic cancers.

D.4.1 V2 ARIC PAC (developed in healthy participants)

The Pearson correlation between the V2 ARIC PAC and chronological age was 0.77 among 5,361 participants who remain cancer-free (after excluding the training set for V2 ARIC PAC) (**Supplemental Figure 4.2**). The correlation between the CaPAC and the V2 ARIC PAC was 0.92 (N=1,763 after excluding the training sets for the CaPAC and V2 ARIC PAC).

Among the 7,842 White and Black participants without a history of cancer at V2, those with higher V2 ARIC PAC acceleration were more likely to be ever smokers, never drinkers, have diabetes, less education, and a lower mean eGFR (**Supplemental Table 4.3**). Similar to the results for the CaPAC, V2 ARIC PAC acceleration was statistically significantly associated with risk of lung cancer [HR (95% CI) per 1 SD = 1.31 (1.16, 1.47)], and risk of CRC [HR (95% CI) per 1 SD = 1.17 (1.00, 1.37)] (**Table 4.4**). These associations were not modified by sex, race, or smoking status (**Supplemental Tables 4.4 and 4.5**). V2 ARIC PAC acceleration was not significantly associated with overall cancer in the fully adjusted model [HR (95% CI) per 1 SD = 1.04 (0.99, 1.10)] (**Table 4.4**). V2 ARIC PAC acceleration was not associated with risk of prostate, postmenopausal breast, bladder, kidney, or pancreatic cancers.

D.4.2 Overrepresentation analysis

Using the Gene Ontology Biological Process database, we found that 1963 pathways were overrepresented (FDR p-value < 0.05) for the proteins included in CaPAC. The Top 40 overrepresented pathways are listed in **Supplemental Table 4.6**. They included “positive regulation of phosphorylation”, “positive regulation of cell population proliferation”, “regulation of immune system process”, “cytokine production”,

“regulation of programmed cell death”, “regulation of phosphorylation”, and “regulation of cytokine production”.

D.5 Sensitivity analysis

The associations between age acceleration for CaPAC3, i.e., cancer-specific PAC constructed using proteomic data at V3, and risk of overall cancer and cancer subtypes (**Table 4.5**) were similar to those for CaPAC.

E. Discussion

In a large prospective study of White and Black individuals, we constructed a cancer-specific PAC, using data on more than 5,000 proteins. In participants who remained cancer-free through 2015, the CaPAC was correlated with chronological age: $r=0.82$. We found that CaPAC acceleration was associated with risk of overall cancer, CRC, and lung cancer. CaPAC acceleration was associated with an increased kidney cancer risk in Black but not White participants.

In our study, we also constructed another cancer-specific PAC – CaPAC2, i.e., PAC based on proteins with largest within-individual changes between V2 and V5 and we compared it to CaPAC, Lehallier’s PAC, and the V2 ARIC PAC. We found that all of these PACs were similarly correlated with chronological age. The magnitude of the associations for overall cancer risk and risk of prostate, lung, colorectal, kidney, and pancreatic cancers were similar for these PACs, and the direction of the associations for risk of postmenopausal breast and bladder cancers were the same.

In our study, we found that CaPAC acceleration was significantly associated with risk of overall cancer [HR (95% CI) per 1 year = 1.02 (1.00, 1.03)]. Of note, here we report HR per one year of age acceleration, the same units as in published studies, which allowed us to compare our findings to the findings in published studies. Previous study also found significant associations between age acceleration for epigenetic clocks and risk of overall cancer.³⁶ A meta-analysis of five studies reported that a one-year increase in acceleration for GrimAge [HR (95% CI)=1.07 (1.05, 1.08)], DNAm PhenoAge [HR (95% CI) = 1.02 (1.01, 1.03)], and Horvath [HR (95% CI)=1.01 (1.00, 1.03)] were significantly associated with risk of overall cancer, but there was no association with Hannum age acceleration.²⁰

In our study, we also found that CaPAC acceleration was statistically significantly associated with risk of lung cancer [HR (95% CI) per 1 year = 1.08 (1.04, 1.12)]; and after stratified by smoking status, the associations were significant among current and former smokers and the association among never smokers were positive although not significant. Previous studies also reported significant associations between acceleration for epigenetic clocks and lung cancer risk.^{19,25} The Women's Health Initiative (WHI) study found a one-year increase in age acceleration of Horvath and DNAm PhenoAge clocks were statistically significantly associated with a 50% ($p = 3.4 \times 10^{-3}$)²⁵ and 5% ($p = 0.031$)¹⁹ increase in risk of lung cancer, respectively; and after stratified by smoking status, they found significant association among current smokers^{19,25} but not among former or never smokers.²⁵

In our study, we also found that CaPAC acceleration was associated with an increased risk of CRC [HR (95% CI) per 1 SD = 1.15 (1.02, 1.30)]. Several previous

studies examined the risk of CRC using epigenetic clocks.^{33,36,101} Using the Mendelian randomization method, a meta-analysis of three studies found a significant association between acceleration for GrimAge and CRC risk.³² Likewise, two case-control studies nested within the Melbourne Collaborative Cohort Study (MCCS) found increased CRC risk associated with acceleration for DNAm PhenoAge [rate ratio (95% CI) per 1 SD = 1.22 (1.09, 1.36)] and GrimAge [rate ratio (95% CI) per 1 SD = 1.19 (1.03, 1.36)], but not the Horvath or Hannum clocks.^{33,36} In addition, the EPIC-Italy study reported a significantly higher acceleration in males with CRC compared to male controls for Horvath's clock ($p=0.042$), but not Hannum.¹⁰¹

Although CaPAC acceleration was not associated with kidney cancer risk overall, we found that it was associated with an increased kidney cancer risk in Black participants [HR (95% CI) per 5-year = 1.94 (1.06, 3.55)] but not White participants (p -interaction = 0.07). A few studies examined the risk of kidney cancer using epigenetic clocks, however, findings were inconsistent across different epigenetic clocks.^{33,36} Two case-control studies nested within MCCS found that Hannum age acceleration was associated with increased kidney cancer risk [per 5-year: OR (95% CI)=1.46 (1.10, 1.94)],³⁶ but there were no associations with the Horvath,³⁶ DNAm PhenoAge, or GrimAge clocks.³³ These two case-control studies included individuals who were born in Australia/New-Zealand, Southern Europe, and Northern Europe, and they did not examine association with kidney cancer in non-White race groups.

In our study, we did not find any associations between CaPAC acceleration and risk of postmenopausal breast cancer, but we found that CaPAC2 acceleration was inversely associated with risk of postmenopausal breast cancer. Findings from previous

studies for the association between epigenetic clocks and breast cancer have been inconsistent. Using the Mendelian randomization method, a meta-analysis of three studies did not find associations between age acceleration for the Hannum, Horvath, DNAm PhenoAge, or GrimAge clocks and risk of breast cancer.³² However, in the Sister Study (N = 2764), Kresovich et al. found that age acceleration for the Hannum, Horvath, and DNAm PhenoAge clocks was statistically significantly associated with an increased breast cancer risk, with the strongest association observed for DNAm PhenoAge: HR (95% CI) per 5-years = 1.15 (1.07, 1.23).³⁵ Another case-control study nested in the EPIC study also found that a one-year age acceleration in the Horvath clock was associated with increased risk of postmenopausal breast cancer [OR (95% CI) = 1.07 (1.02, 1.11)].³⁸ The exact reason for the inconsistency between studies is not clear, more studies are needed to investigate the association between age acceleration and risk of breast cancer.

We did not find any associations of CaPAC acceleration with risk of pancreatic cancer. Only a few studies have examined the risk of pancreatic cancer using epigenetic clocks and findings were inconsistent when using different epigenetic clocks.^{34,101,103} For example, a pooled analysis of Nurses' Health Study, Physician's Health Study, and the Health Professionals Follow-up Study found positive dose-response trends of acceleration for Hannum and PhenoAge with pancreatic cancer risk, however, for Horvath, the highest OR was found in the third quartile compared to the lowest quartile.³⁴

In our study, we did not find any associations between CaPAC acceleration and risk of prostate cancer. Our results are similar to those in previous studies that did not detect associations between epigenetic clocks and prostate cancer risk.^{33,36} However, using the Mendelian randomization method, a meta-analysis of three studies found that

acceleration for GrimAge was inversely associated with prostate cancer risk [per one-year: meta-analysis OR (95% CI) = 0.93 (0.87, 0.99)].³²

In addition, we did not find any associations between CaPAC acceleration and risk of bladder cancer, but we found that CaPAC2 acceleration was inversely associated with risk of bladder cancer. No studies have examined the association for risk of bladder cancer with PACs or epigenetic clocks.

To the best of our knowledge, our study is the first to construct a cancer-specific PAC and document significant associations of PAC acceleration with overall cancer risk, lung cancer risk, CRC risk, and kidney cancer risk (in Black individuals only).

Examining a PAC in cancer is especially important because it could be biomarker for future cancer risk. Most cancer cases develop over a long period of time and have a long subclinical period, and 30%-50% of all cancer cases are preventable;⁴⁵ thus, the PAC could provide a new method for identifying people at high risk of cancer that need more frequent screening. In addition, proteins included into PAC may serve as targets for novel anti-aging drugs that will eventually slow down aging.⁴⁶ Cancer is an age-related disease, and both cancer and aging result from epigenetic damage accumulation,⁴⁷ inflammation, oxidative stress, and other damage.⁴⁸⁻⁵⁰ Therefore, slowing aging is likely to delay the development of cancer.

The strengths of this population-based study include the prospective design with over 20 years of follow-up in males and females, a large community-based sample of both White and Black individuals, and adjudicated cancer incidence. Another strength is that the version of SomaScan assay used in ARIC measured more than 5000 proteins, larger than the panels of proteins used in most previous studies. With a larger panel of

proteins, we may be able to develop a PAC that more accurately predicts chronological age. In addition, we found that PACs that included different set of proteins and were developed in populations of different health status – healthy and cancer-free were similarly associated with cancer risk, suggesting the robustness of PACs. Our study also has limitations. First, the possibility of protein degradation during long-term storage cannot be excluded. However, the blood samples were frozen right after their collection and have never been thawed limiting the possibility of degradation.⁷⁶ Second, the SomaScan assay provides relative quantification instead of absolute quantification.¹⁰⁴ Third, we have limited number of cases for uncommon cancers.

In conclusion, we constructed a CaPAC in a population of White and Black individuals and showed that this CaPAC was able to predict risk of cancer independently of known cancer risk factors. The proteins included in this CaPAC hold promise as anti-aging drug targets for cancer. Future studies are needed to validate this CaPAC in other populations.

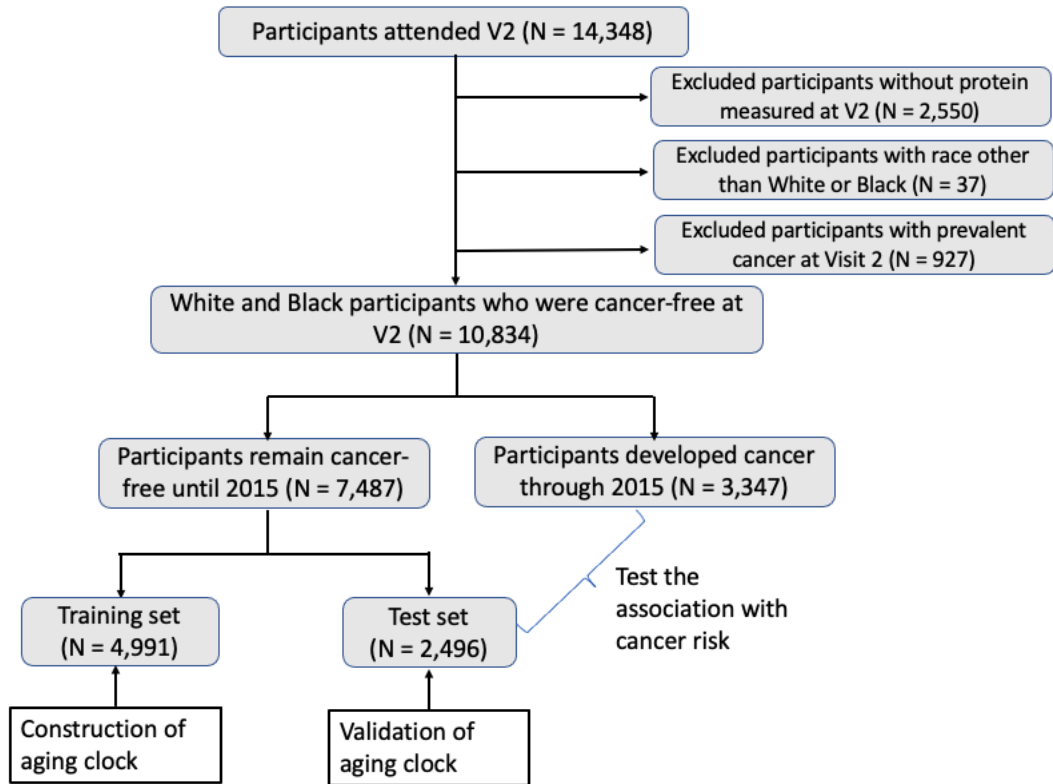


Figure 4.1 Study population

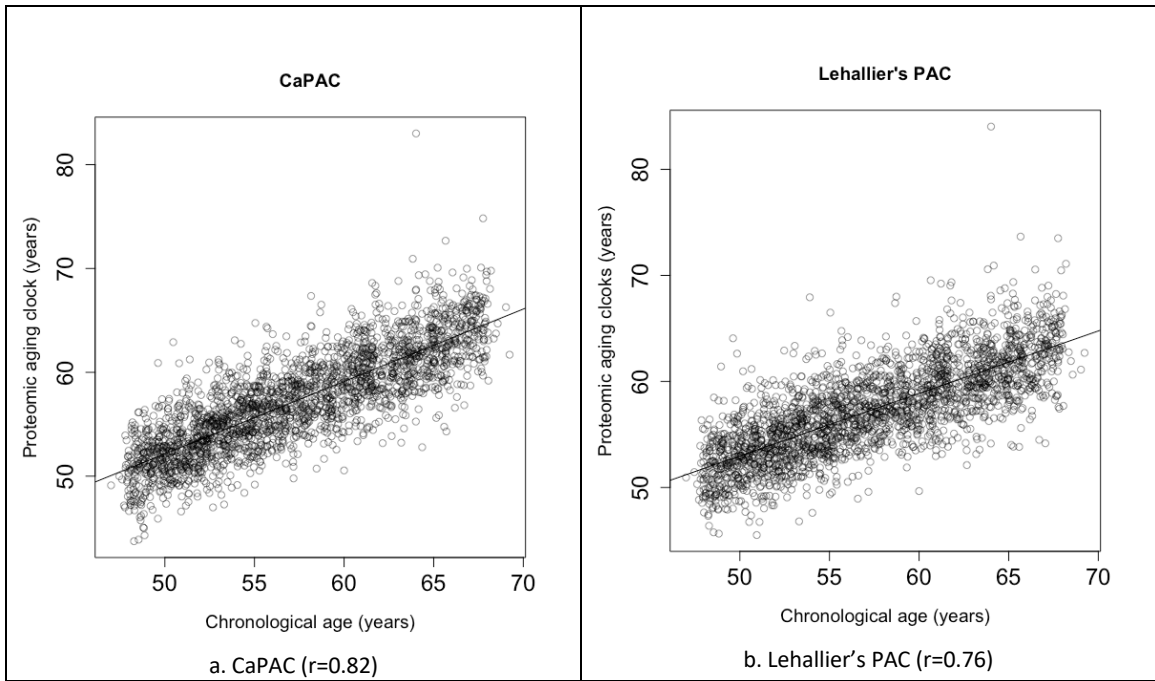


Figure 4.2 Pearson correlation (r) between CaPAC and Lehallier's PAC and chronological age in the test set

Table 4.1 Visit 2 characteristics of cancer-free participants at Visit 2 across quartiles of age acceleration for CaPAC and Lehallier's PAC

	CaPAC				P-value	Lehallier's PAC				P-value
	Q1 (N = 1,460)	Q2 (N = 1,461)	Q3 (N = 1,461)	Q4 (N = 1,461)		Q1 (N = 1,460)	Q2 (N = 1,461)	Q3 (N = 1,461)	Q4 (N = 1,461)	
Range of age acceleration (years)	-12.0 to -1.8	-1.7 to -0.0	0.01 to 1.7	1.8 to 21.3		-10.5 to -1.8	-1.7 to -0.1	0.0 to 1.7	1.8 to 22.9	
Demographic										
Age, years (SD)	57.9 (5.7)	57.7 (5.7)	57.6 (5.7)	57.9 (5.6)	0.30	57.8 (5.8)	57.7 (5.6)	57.8 (5.6)	57.9 (5.6)	0.69
Male, %	47.5	47.0	50.2	47.4	0.28	46.6	47.8	50.3	47.4	0.22
White, %	71.8	78.3	77.8	75.5	<0.01	69.8	77.9	78.3	77.5	<0.01
Education, %										
Less than high school	21.8	21.5	22.3	25.8		22.1	21.7	22.4	25.2	<0.01
High school equivalent	40.1	41.5	42.1	43.1	<0.01	40.3	42.1	40.4	44.0	
Greater than high school	38.0	37.0	35.6	31.1		37.7	36.2	37.2	30.8	
Lifestyle/Medical factors										
BMI, kg/m ² (SD)	28.1 (5.0)	27.8 (5.0)	27.9 (5.5)	28.3 (6.0)	0.03	28.3 (5.3)	27.9 (5.0)	27.9 (5.3)	28.1 (5.8)	0.16
Ever smoker, %	61.5	61.8	62.5	63.4	0.73	58.6	61.7	62.7	66.2	<0.01
Pack-years of smoking among current and former smokers, pack-years (SD)	29.2 (24.4)	29.4 (20.9)	29.91 (21.5)	30.5 (22.4)	0.64	28.0 (21.1)	29.0 (22.9)	29.9 (22.1)	31.9 (22.9)	<0.01
Alcohol intake, %										
Current drinker	59.5	59.8	58.7	53.6		55.7	61.1	59.6	55.2	
Former drinker	20.5	18.1	19.7	22.4	<0.01	21.8	18.0	19.4	21.6	0.01
Never drinker	20.0	22.1	21.5	24.0		22.5	20.9	21.0	23.3	
Diabetes, %	14.0	12.5	14.5	19.3	<0.01	14.2	13.0	13.4	19.7	<0.01
Aspirin use in the preceding two weeks, %	46.3	53.6	51.1	54.0	<0.01	46.7	51.8	49.8	56.8	<0.01
Ever users of hormone replacement therapy among women, %	45.8	47.2	46.2	42.4	0.27	45.0	48.1	44.8	43.6	0.33
eGFR, mL/min/1.73 m ² (SD)	100.3 (15.6)	99.5 (15.4)	98.4 (15.6)	94.5 (18.3)	<0.01	101.1 (14.4)	99.5 (15.1)	97.7 (15.8)	94.3 (19.3)	<0.01

Abbreviations: PAC – proteomic aging clock; CaPAC – cancer-specific proteomic aging clock; BMI – body mass index; eGFR – estimated glomerular filtration rate; SD – standard deviation.

Table 4.2 Association between age acceleration for CaPAC and Lehallier's PAC risk of overall cancer and cancer subtypes

		No. of incident cancer cases	Total person-years	CaPAC HR (95% CI) per 1 SD (SD = 2.67 years)	Lehallier's PAC HR (95% CI) per 1 SD (SD = 2.79 years)
Overall cancer	Model 1 ^a	3,347	93,894	1.04 (1.01, 1.08)	1.04 (1.01, 1.08)
	Model 2 ^b			1.04 (1.00, 1.08)	1.02 (0.98, 1.06)
	Model 3 ^c			1.04 (1.00, 1.08)	1.02 (0.98, 1.06)
Prostate cancer ^e	Model 1	671	41,771	0.98 (0.91, 1.06)	0.96 (0.89, 1.04)
	Model 2			0.98 (0.90, 1.07)	0.97 (0.89, 1.05)
	Model 3			0.98 (0.90, 1.07)	0.96 (0.88, 1.05)
Lung cancer	Model 1	479	93,894	1.24 (1.14, 1.36)	1.32 (1.21, 1.44)
	Model 2			1.24 (1.12, 1.36)	1.20 (1.09, 1.32)
	Model 3			1.23 (1.12, 1.36)	1.20 (1.09, 1.32)
Postmenopausal breast cancer ^e	Model 1	472	52,099	0.97 (0.87, 1.06)	0.97 (0.89, 1.07)
	Model 2			0.94 (0.85, 1.03)	0.94 (0.85, 1.04)
	Model 3			0.93 (0.84, 1.03)	0.94 (0.85, 1.05)
Colorectal cancer	Model 1	312	93,894	1.15 (1.03, 1.29)	1.07 (0.95, 1.19)
	Model 2			1.15 (1.02, 1.30)	1.08 (0.95, 1.22)
	Model 3			1.15 (1.02, 1.30)	1.08 (0.95, 1.22)
Bladder cancer	Model 1	154	93,894	0.96 (0.81, 1.13)	1.01 (0.85, 1.19)
	Model 2			0.96 (0.81, 1.14)	0.93 (0.77, 1.11)
	Model 3			0.97 (0.82, 1.15)	0.93 (0.78, 1.12)
Kidney cancer	Model 1	114	93,894	1.12 (0.93, 1.35)	1.19 (0.99, 1.44)
	Model 2			1.11 (0.91, 1.35)	1.17 (0.96, 1.41)
	Model 3			1.09 (0.90, 1.33)	1.14 (0.94, 1.38)
Pancreatic cancer	Model 1	101	93,894	1.01 (0.83, 1.24)	1.02 (0.83, 1.25)
	Model 2			1.00 (0.81, 1.23)	0.98 (0.79, 1.21)
	Model 3			1.00 (0.81, 1.23)	0.98 (0.79, 1.22)

Abbreviations: PAC – proteomic aging clock; CaPAC – cancer-specific proteomic aging clock; BMI – body mass index; eGFR – estimated glomerular filtration rate; SD – standard deviation; HR – hazard ratio; CI – confidence interval.

^a Model 1 was adjusted for chronological age, sex, race, and study center.

^b Model 2 = Model 1 + education + BMI + smoking status + pack-years of smoking + alcohol intake + eGFR.

^c Model 3 = Model 2 + aspirin use + hormone replacement therapy + diabetes.

^e For analysis involved risk of prostate cancer, SDs were 2.61 and 2.74 years for CaPAC and Lehallier's PAC, respectively.

^f For analysis involved risk of breast cancer, SDs were 2.73 and 2.83 years for CaPAC and Lehallier's PAC, respectively.

Table 4.3 Association between age acceleration for CaPAC and Lehallier's PAC risk of overall cancer and cancer subtypes stratified by sex and race

Stratified by sex^a					
Cancer type	Sex	N of incident cancer cases	Total person-years	CaPAC	Lehallier's PAC
				HR (95% CI) ^b per 1 SD (SDs = 2.73 and 2.61 years for females and males, respectively)	HR (95% CI) ^b per 1 SD (SDs = 2.83 and 2.74 years for females and males, respectively)
Overall cancer	Female	1537	52,123	1.03 (0.98, 1.09)	1.03 (0.97, 1.08)
	Male	1810	41,771	1.05 (0.99, 1.10)	1.02 (0.97, 1.08)
	P-interaction			0.95	0.35
Lung cancer	Female	208	52,123	1.27 (1.10, 1.48)	1.24 (1.07, 1.45)
	Male	271	41,771	1.20 (1.06, 1.37)	1.18 (1.04, 1.34)
	P-interaction			0.99	0.56
Colorectal cancer	Female	151	52,123	1.18 (0.99, 1.40)	1.12 (0.95, 1.34)
	Male	161	41,771	1.13 (0.95, 1.35)	1.03 (0.86, 1.24)
	P-interaction			0.86	0.49
Bladder cancer	Female	38	52,123	1.25 (0.91, 1.70)	1.25 (0.92, 1.70)
	Male	116	41,771	0.87 (0.71, 1.06)	0.81 (0.65, 1.00)
	P-interaction			0.04	0.01
Kidney cancer	Female	57	52,123	1.23 (0.94, 1.59)	1.33 (1.04, 1.71)
	Male	57	41,771	0.94 (0.71, 1.25)	0.91 (0.68, 1.22)
	P-interaction			0.14	0.04
Pancreatic cancer	Female	45	52,123	1.10 (0.80, 1.52)	1.24 (0.91, 1.70)
	Male	56	41,771	0.90 (0.67, 1.19)	0.79 (0.59, 1.06)
	P-interaction			0.42	0.07
Stratified by race^a					
Cancer type	Race	N of incident cancer cases	Total person-years	CaPAC	Lehallier's PAC
				HR (95% CI) ^b per 1 SD (SDs = 2.60 and 2.89 years for White and Black participants, respectively)	HR (95% CI) ^b per 1 SD (SDs = 2.69 and 3.06 years for White and Black participants, respectively)
Overall cancer	White	2,550	72,101	1.02 (0.98, 1.07)	0.99 (0.96, 1.04)
	Black	797	21,793	1.10 (1.01, 1.19)	1.09 (1.01, 1.19)
	P-interaction			0.20	0.08
Prostate cancer ^c	White	473	33,863	1.00 (0.91, 1.10)	0.95 (0.86, 1.05)
	Black	198	7,907	0.92 (0.78, 1.10)	0.98 (0.83, 1.17)
	P-interaction			0.61	0.53
Lung cancer	White	378	72,101	1.21 (1.09, 1.35)	1.19 (1.06, 1.32)
	Black	101	21,793	1.28 (1.01, 1.63)	1.24 (0.98, 1.56)
	P-interaction			0.53	0.93
	White	358	38,213	0.92 (0.82, 1.30)	0.92 (0.82, 1.03)

Postmenopausal breast cancer ^d	Black	114	13,886	1.00 (0.81, 1.24)	1.02 (0.82, 1.26)
	P-interaction			0.33	0.23
Colorectal cancer	White	221	72,101	1.12 (0.97, 1.29)	1.10 (0.95, 1.27)
	Black	91	21,793	1.23 (0.96, 1.60)	1.03 (0.80, 1.31)
	P-interaction			0.68	0.54
Bladder cancer	White	139	72,101	0.99 (0.83, 1.19)	0.94 (0.78, 1.14)
	Black	15	21,793	0.75 (0.40, 1.40)	0.87 (0.48, 1.60)
	P-interaction			0.47	0.81
Kidney cancer	White	78	72,101	0.95 (0.75, 1.21)	0.99 (0.78, 1.26)
	Black	36	21,793	1.46 (1.03, 2.08)	1.49 (1.07, 2.07)
	P-interaction			0.05	0.07
Pancreatic cancer	White	70	72,101	1.04 (0.81, 1.33)	0.95 (0.73, 1.22)
	Black	31	21,793	0.91 (0.59, 1.40)	1.03 (0.67, 1.57)
	P-interaction			0.56	0.62

Abbreviations: PAC – proteomic aging clock; CaPAC – cancer-specific proteomic aging clock; BMI – body mass index; eGFR – estimated glomerular filtration rate; SD – standard deviation; HR – hazard ratio; CI – confidence interval.

^a Interactions with sex and race were examined using a multiplicative term.

^b Model was adjusted for chronological age, sex, race, study center, education, BMI, smoking status, pack-years of smoking, alcohol intake, eGFR, aspirin use, hormone replacement therapy, and diabetes.

^c For analysis involved prostate cancer risk, SDs were: CaPAC = 2.55 and 2.82 years for White and Black participants, respectively; and Lehallier's PAC = 2.63 and 3.01 years for White and Black participants, respectively.

^d For analysis involved postmenopausal breast cancer risk, SDs were: CaPAC = 2.67 and 2.93 years for White and Black participants, respectively; and Lehallier's PAC = 2.74 and 3.08 years for White and Black participants, respectively.

Table 4.4 Association between age acceleration for CaPAC2 (cancer-specific PAC constructed with proteins with the largest within-individual changes between V2 and V5) and V2 ARIC PAC (healthy participants PAC) and cancer risk

	CaPAC2			V2 ARIC PAC		
	N of incident cancer cases	Total person-years	HR (95% CI) ^a per 1 SD (SD= 2.65 years)	N of incident cancer cases	Total person-years	HR (95% CI) ^a per 1 SD (SD = 2.95 years)
Any cancer	3,347	147,578	0.99 (0.95, 1.03)	2,480	139,188	1.04 (0.99, 1.10)
Prostate cancer ^b	671	62,192	1.03 (0.94, 1.13)	500	60,076	0.97 (0.87, 1.08)
Lung cancer	479	147,578	1.24 (1.12, 1.38)	387	139,188	1.31 (1.16, 1.47)
Postmenopausal breast cancer ^c	472	85,338	0.84 (0.76, 0.94)	324	79,064	0.95 (0.83, 1.08)
Colorectal cancer	312	147,578	1.09 (0.96, 1.24)	231	139,188	1.17 (1.00, 1.37)
Bladder cancer	154	147,578	0.82 (0.68, 0.99)	108	139,188	0.81 (0.63, 1.02)
Kidney cancer	114	147,578	1.04 (0.85, 1.28)	96	139,188	1.08 (0.86, 1.37)
Pancreatic cancer	101	147,578	1.09 (0.96, 1.24)	78	139,188	0.99 (0.75, 1.30)

Abbreviations: PAC – proteomic aging clock; BMI – body mass index; eGFR – estimated glomerular filtration rate; SD – standard deviation; HR – hazard ratio; CI – confidence interval.

^a Model was adjusted for chronological age, sex, race, study center, education, BMI, smoking status, pack-years of smoking, alcohol intake, eGFR, aspirin use, hormone replacement therapy, and diabetes.

^b For analysis involved prostate cancer risk, SDs were 2.54, and 2.83 years for CaPAC2, and V2 ARIC PAC, respectively.

^c For analysis involved postmenopausal breast cancer risk, SDs were 2.74, and 3.05 years for CaPAC2, and V2 ARIC PAC, respectively.

Table 4.5 Association between CaPAC3 (Visit 3 cancer-specific clock) acceleration and risk of overall cancer and cancer subtypes

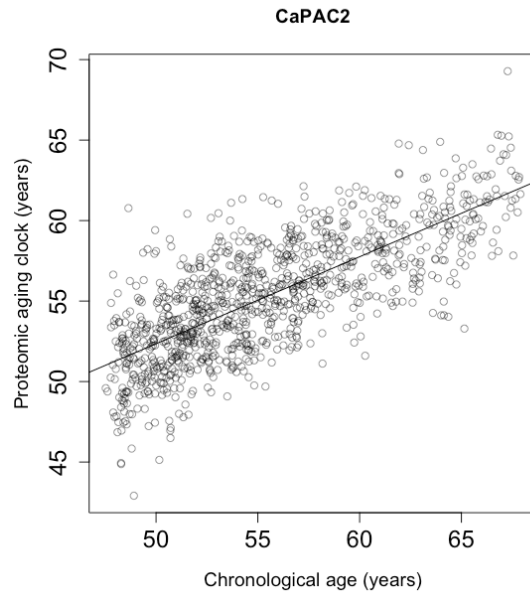
	N of incident cancer cases	Total person-years	HR (95% CI) ^a per 1 SD (SD=2.41 years)
Any cancer	3,092	75,060	1.05 (1.01, 1.09)
Prostate cancer ^b	634	34,217	0.97 (0.89, 1.05)
Lung cancer	430	75,060	1.23 (1.11, 1.37)
Postmenopausal breast cancer ^c	411	40,842	0.97 (0.87, 1.08)
Colorectal cancer	287	75,060	1.13 (1.00, 1.29)
Bladder cancer	160	75,060	1.02 (0.96, 1.09)
Kidney cancer	109	75,060	0.99 (0.81, 1.21)
Pancreatic cancer	94	75,060	1.04 (0.95, 1.13)

Abbreviations: BMI – body mass index; eGFR – estimated glomerular filtration rate; SD – standard deviation; HR – hazard ratio; CI – confidence interval.

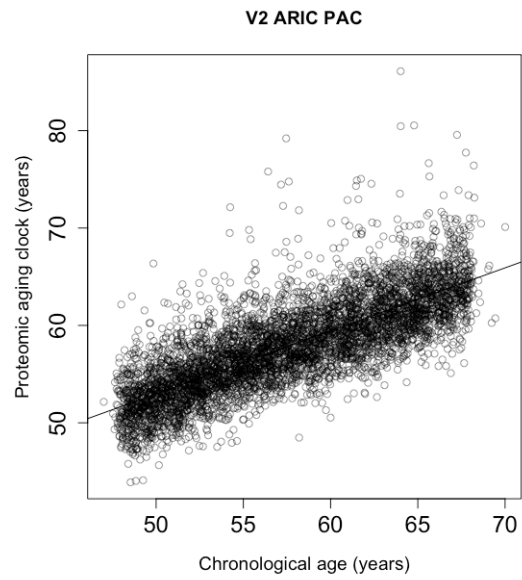
^a Model was adjusted for chronological age, sex, race, study center, education, BMI, smoking status, pack-years of smoking, alcohol intake, eGFR, aspirin use, hormone replacement therapy, and diabetes at Visit 3.

^b For analysis involved prostate cancer risk, SD = 2.35 years.

^c For analysis involved postmenopausal breast cancer risk, SD = 2.45 years.



Supplemental Figure 4.1 Pearson correlation (r) between CaPAC2 (i.e., cancer-specific PAC constructed using proteins with largest within-individual changes between Visit 2 and Visit 5) and chronological age in test set2 ($r= 0.75$)



Supplemental Figure 4.2 Pearson correlation (r) between the V2 ARIC PAC (developed in healthy participants) and chronological age in participants who remained cancer-free through 2015 ($r= 0.77$)

Supplemental Table 4.1 Description of proteomic aging clocks used in this study

Proteomic aging clock (PAC)	Description
CaPAC	Cancer-specific PAC constructed using elastic net regression and with all the 4,955 aptamers available at Visit 2
CaPAC2	Cancer-specific PAC constructed using elastic net regression and with the top 2000 proteins with the largest within-individual changes between Visit 2 and Visit 5
CaPAC3	Cancer-specific PAC constructed using elastic net regression and with all the 4,955 aptamers available at Visit 3
V2 ARIC PAC	Healthy participant's PAC constructed using elastic net regression and with all the 4,955 aptamers available at Visit 2 (PAC developed in previous manuscript)
Lehallier's PAC	Healthy people's PAC proposed by Lehallier [2020] ¹⁴

Supplemental Table 4.2 Association between acceleration for the CaPAC and Lehallier's PAC and risk of overall cancer and cancer subtypes, stratified by smoking status

Cancer type	Smoking status ^a	N of incident cancer cases	Total person-years	CaPAC	Lehallier's PAC
				HR (95% CI) ^b per 1 SD (SDs = 2.58, 2.76, and 2.64 years for current, former, and never smokers, respectively)	HR (95% CI) ^b per 1 SD (SDs = 2.67, 2.86, and 2.76 years, for current, former, and never smokers, respectively)
Overall cancer	Current smoker	940	20,790	1.04 (0.97, 1.12)	1.03 (0.96, 1.10)
	Former smoker	1267	34,996	1.06 (0.99, 1.13)	1.01 (0.95, 1.08)
	Never smoker	1137	37,996	1.01 (0.95, 1.07)	1.01 (0.96, 1.08)
	P-interaction			0.85	0.73
Prostate cancer ^c	Current smoker	137	9858	1.10 (0.91, 1.33)	1.07 (0.89, 1.28)
	Former smoker	330	20,391	0.95 (0.83, 1.08)	0.92 (0.80, 1.05)
	Never smoker	203	11,445	0.96 (0.83, 1.11)	0.98 (0.84, 1.13)
	P-interaction			0.28	0.26
Lung cancer	Current smoker	296	20,790	1.17 (1.04, 1.32)	1.16 (1.03, 1.31)
	Former smoker	150	34,996	1.29 (1.09, 1.54)	1.19 (0.99, 1.43)
	Never smoker	31	37,996	1.42 (0.96, 2.10)	1.40 (0.95, 2.01)
	P-interaction			0.58	0.67
Postmenopausal breast cancer ^d	Current smoker	95	10,932	0.99 (0.79, 1.23)	1.02 (0.82, 1.27)
	Former smoker	128	14,604	0.96 (0.78, 1.18)	1.01 (0.82, 1.24)
	Never smoker	249	26,526	0.90 (0.78, 1.03)	0.89 (0.78, 1.02)
	P-interaction			0.82	0.68
Colorectal cancer	Current smoker	68	20,790	1.01 (0.77, 1.31)	0.92 (0.71, 1.18)
	Former smoker	125	34,996	1.40 (1.15, 1.71)	1.18 (0.96, 1.44)
	Never smoker	119	37,996	1.01 (0.83, 1.22)	1.05 (0.88, 1.26)
	P-interaction			0.09	0.54
Bladder cancer	Current smoker	43	20,790	1.03 (0.75, 1.42)	0.93 (0.68, 1.28)
	Former smoker	90	34,996	0.90 (0.72, 1.14)	0.84 (0.66, 1.06)
	Never smoker	21	37,996	1.09 (0.69, 1.72)	1.29 (0.83, 2.01)

	P-interaction			0.82	0.28
Kidney cancer	Current smoker	32	20,790	0.70 (0.48, 1.01)	1.02 (0.73, 1.44)
	Former smoker	33	34,996	1.44 (0.99, 2.10)	1.23 (0.85, 1.79)
	Never smoker	49	37,996	1.19 (0.89, 1.58)	1.14 (0.87, 1.49)
	P-interaction			0.02	0.79
Pancreatic cancer	Current smoker	26	20,790	0.79 (0.52, 1.21)	0.94 (0.63, 1.39)
	Former smoker	36	34,996	0.81 (0.55, 1.19)	0.82 (0.56, 1.22)
	Never smoker	39	37,996	1.32 (0.94, 1.86)	1.12 (0.81, 1.55)
	P-interaction			0.12	0.52

Abbreviations: PAC – proteomic aging clock; CaPAC – cancer-specific proteomic aging clock; BMI – body mass index; eGFR – estimated glomerular filtration rate; SD – standard deviation; HR – hazard ratio; CI – confidence interval.

^a Interaction with smoking status was examined using a multiplicative term.

^b Model was adjusted for chronological age, sex, race, study center, education, BMI, pack-years of smoking, alcohol intake, eGFR, aspirin use, hormone replacement therapy, and diabetes.

^c For analysis involved risk of prostate cancer, SDs were: CaPAC = 2.45, 2.67, and 2.58 years for current, former, and never smokers, respectively; and Lehallier's PAC = 2.67, 2.77, and 2.72 years for current, former, and never smokers, respectively.

^d For analysis involved risk of postmenopausal breast cancer risk, SDs were: CaPAC = 2.69, 2.87, and 2.67 years for current, former, and never smokers, respectively; and Lehallier's PAC = 2.67, 3.02, and 2.79 years for current, former, and never smokers, respectively.

Supplemental Table 4.3 Visit 2 characteristics of cancer-free participants at Visit 2 across quartiles of acceleration for the CaPAC2 and V2 ARIC PAC

	CaPAC2 (cancer-specific PAC constructed with proteins with the largest within-individual changes between V2 and V5)				P-value
	Q1 (N = 2,171)	Q2 (N = 2,172)	Q3 (N = 2,172)	Q4 (N = 2,172)	
Range of age acceleration (years)	-13.8 to -1.7	-1.6 to -1.1	-1.0 to 1.6	1.7 to 21.5	
Age, years (SD)	57.9 (5.8)	58.0 (5.8)	58.0 (5.7)	58.0 (5.6)	0.91
Male, %	46.2	47.1	46.59	44.43	0.31
White, %	70.6	75.9	76.01	74.95	<0.01
Education, %					
Less than high school	23.4	23.9	22.5	26.9	
High school equivalent	41.2	40.4	43.5	42.1	<0.01
Greater than high school	35.4	35.7	33.9	31.0	
BMI, kg/m ² (SD)	28.2 (5.2)	28.3 (5.3)	27.9 (5.4)	28.0 (6.0)	0.08
Ever smoker, %	58.3	62.8	61.8	64.6	<0.01
Pack-years of smoking among current and former smokers, pack-years (SD)	28.7 (23.5)	29.1 (22.7)	29.1 (20.7)	32.3 (23.4)	<0.01
Alcohol intake, %					
Current drinker	57.3	57.6	56.1	51.5	
Former drinker	21.2	20.0	21.1	25.1	<0.01
Never drinker	21.5	22.5	22.8	23.4	
Diabetes, %	15.3	16.6	15.3	19.9	<0.01
Aspirin use in the preceding two weeks, %	46.8	50.7	53.1	54.6	<0.01
Ever users of hormone replacement therapy among women, %	46.4	46.1	43.8	39.7	<0.01
eGFR, mL/min/1.73 m ² (SD)	101.7 (14.6)	98.6 (15.5)	96.7 (16.0)	92.0 (20.3)	<0.01
	V2 ARIC PAC (healthy participants PAC)				
	Q1 (N = 1,960)	Q2 (N = 1,960)	Q3 (N = 1,961)	Q4 (N = 1,961)	P-value
Range of age acceleration (years)	-13.9 to -1.9	-1.8 to -0.2	-0.1 to 1.7	1.8 to 24.2	
Age, years (SD)	58.2 (5.8)	57.9 (5.8)	57.7 (5.7)	58.2 (5.6)	0.01
Male, %	45.4	50.0	45.5	45.5	<0.01
White, %	71.6	74.5	73.8	66.0	<0.01
Education, %					
Less than high school	22.8	22.6	23.2	30.4	
High school equivalent	41.2	40.8	42.9	40.3	<0.01
Greater than high school	36.1	36.6	33.9	29.3	
BMI, kg/m ² (SD)	28.3 (5.0)	28.5 (5.3)	28.4 (5.5)	28.8 (6.3)	0.06
Ever smoker, %	58.7	61.7	62.0	63.8	0.01
Pack-years of smoking among current and former smokers, pack-years (SD)	28.4 (22.6)	30.3 (23.7)	29.4 (21.8)	32.4 (24.0)	<0.01

Alcohol intake, %					
Current drinker	56.9	57.6	54.9	48.1	
Former drinker	21.9	21.0	20.9	26.4	<0.01
Never drinker	21.3	21.4	24.2	25.5	
Diabetes, %	15.6	17.7	20.5	30.1	<0.01
Aspirin use in the preceding two weeks, %	50.8	51.1	54.1	52.6	0.15
Ever users of hormone replacement therapy among women, %	49.3	46.9	40.7	35.6	<0.01
eGFR, mL/min/1.73 m ² (SD)	101.1 (14.2)	99.0 (15.1)	97.2 (15.8)	89.7 (21.9)	<0.01

Abbreviations: PAC – proteomic aging clock; BMI – body mass index; eGFR – estimated glomerular filtration rate; SD – standard deviation.

^a cancer-free participants at Visit 2 after excluding participants in the training set for corresponding PACs.

Supplemental Table 4.4 Association between age acceleration for the CaPAC2 and V2 ARIC PAC and cancer risk, stratified by sex and race

		Stratified by sex					
Cancer type	Sex ^a	CaPAC2			V2 ARIC PAC		
		N of incident cancer cases	Total person-years	HR (95% CI) ^b per 1 SD (SDs = 2.74 and 2.54 years for females and males, respectively)	N of incident cancer cases	Total person-years	HR (95% CI) ^b per 1 SD (SDs = 3.05 and 2.83 years for females and males, respectively)
Overall cancer	Female	1,537	85,386	0.95 (0.90, 1.01)	1,105	79,112	1.00 (0.93, 1.08)
	Male	1,810	62,192	1.03 (0.98, 1.09)	1,375	60,076	1.08 (1.02, 1.15)
	P-interaction			0.18			0.32
Lung cancer	Female	208	85,386	1.28 (1.10, 1.49)	153	79,112	1.27 (1.04, 1.54)
	Male	271	62,192	1.23 (1.08, 1.40)	234	60,076	1.30 (1.12, 1.52)
	P-interaction			0.85			0.44
Colorectal cancer	Female	151	85,386	1.16 (0.97, 1.39)	108	79,112	1.14 (0.91, 1.43)
	Male	161	62,192	1.02 (0.85, 1.22)	123	60,076	1.20 (0.97, 1.48)
	P-interaction			0.36			0.83
Bladder cancer	Female	38	85,386	0.94 (0.65, 1.34)	22	79,112	0.91 (0.55, 1.52)
	Male	116	62,192	0.78 (0.62, 0.96)	86	60,076	0.77 (0.59, 1.01)
	P-interaction			0.18			0.34
Kidney cancer	Female	57	85,386	1.09 (0.81, 1.45)	50	79,112	1.15 (0.83, 1.58)
	Male	57	62,192	1.01 (0.75, 1.35)	46	60,076	1.01 (0.72, 1.32)
	P-interaction			0.59			0.34
Pancreatic cancer	Female	45	85,386	1.18 (0.85, 1.64)	39	79,112	1.01 (0.69, 1.48)
	Male	56	62,192	0.86 (0.64, 1.17)	39	60,076	0.93 (0.64, 1.37)
	P-interaction			0.25			0.78

		Stratified by race					
Cancer type	Race ^a	CaPAC2			V2 ARIC PAC		
		N of incident cancer cases	Total person-years	HR (95% CI) ^b per 1 SD (SDs = 2.53 and 2.96 years for White and Black	N of incident cancer cases	Total person-years	HR (95% CI) ^b per 1 SD (SDs = 2.73 and 3.43 years for Whites and Black

				participants, respectively)			participants, respectively)
Overall cancer	White	2,550	111,669	0.96 (0.92, 1.01)	1,803	101,320	1.03 (0.98, 1.09)
	Black	797	35,909	1.11 (1.01, 1.21)	677	37,868	1.09 (0.98, 1.21)
	P-interaction			0.01			0.61
Prostate cancer ^c	White	473	49,816	0.99 (0.89, 1.10)	341	12,476	0.92 (0.81, 1.05)
	Black	198	12,376	1.16 (0.97, 1.39)	159	47,600	1.09 (0.88, 1.34)
	P-interaction			0.12			0.16
Lung cancer	White	378	111,669	1.20 (1.08, 1.35)	299	101,320	1.29 (1.13, 1.46)
	Black	101	35,909	1.40 (1.09, 1.78)	88	37,868	1.39 (1.05, 1.85)
	P-interaction			0.28			0.80
Postmenopausal breast cancer ^d	White	358	61,805	0.83 (0.73, 0.93)	222	101,320	0.99 (0.85, 1.15)
	Black	114	23,533	0.91 (0.72, 1.16)	102	37,868	0.88 (0.67, 1.15)
	P-interaction			0.25			0.57
Colorectal cancer	White	221	111,669	1.10 (0.95, 1.27)	156	101,320	1.23 (1.03, 1.46)
	Black	91	35,909	1.09 (0.83, 1.43)	75	37,868	1.05 (0.75, 1.46)
	P-interaction			0.70			0.11
Bladder cancer	White	139	111,669	0.87 (0.72, 1.05)	95	101,320	0.84 (0.66, 1.07)
	Black	15	35,909	0.50 (0.26, 0.95)	13	37,868	0.59 (0.26, 1.35)
	P-interaction			0.15			0.46
Kidney cancer	White	78	111,669	0.90 (0.70, 1.16)	63	101,320	0.97 (0.74, 1.28)
	Black	36	35,909	1.47 (1.02, 2.12)	33	37,868	1.40 (0.90, 2.16)
	P-interaction			0.04			0.23
Pancreatic cancer	White	70	111,669	0.96 (0.74, 1.25)	52	101,320	1.00 (0.73, 1.37)
	Black	31	35,909	1.12 (0.73, 1.74)	26	37,868	1.00 (0.59, 1.69)
	P-interaction			0.48			0.95

Abbreviations: PAC – proteomic aging clock; CaPAC – cancer-specific proteomic aging clock; BMI – body mass index; eGFR – estimated glomerular filtration rate; SD – standard deviation; HR – hazard ratio; CI – confidence interval.

^a Interactions with sex and race were examined using a multiplicative term.

^b Model was adjusted for chronological age, sex, race, study center, education, BMI, smoking status, pack-years of smoking, alcohol intake, eGFR, aspirin use, hormone replacement therapy, and diabetes.

^c For analysis involved prostate cancer risk, SDs were: CaPAC2 = 2.45 and 2.83 years for White participants and Black participants, respectively; and V2 ARIC PAC = 2.66 and 3.31 years for White participants and Black participants, respectively.

^d For analysis involved postmenopausal breast cancer risk, SDs were: CaPAC2 = 2.60 and 3.05 years for White participants and Black participants, respectively; and V2 ARIC PAC = 2.78 and 3.50 years for White participants and Black participants, respectively.

Supplemental Table 4.5 Associations of age acceleration for CaPAC2 and V2 ARIC PAC with cancer risk, stratified by smoking status

Cancer type	Smoking status ^a	Proteomic age acceleration for CaPAC2			Proteomic age acceleration for V2 ARIC PAC		
		N of incident cancer cases	Total person-years	HR (95% CI) ^b per 1 SD (SDs = 2.63, 2.68, and 2.62 years for current, former, and never smokers, respectively)	N of incident cancer cases	Total person-years	HR (95% CI) ^b per 1 SD (SDs = 2.77, 3.02, and 2.96 years for current, former, and never smokers, respectively)
Overall cancer	Current smoker	940	32,492	1.06 (0.99, 1.14)	715	28,418	1.09 (0.99, 1.18)
	Former smoker	1,267	54,706	0.98 (0.91, 1.05)	965	52,970	1.01 (0.93, 1.10)
	Never smoker	1,137	60,171	0.95 (0.89, 1.01)	797	57,543	1.01 (0.92, 1.10)
	P-interaction			0.08			0.34
Prostate cancer ^c	Current smoker	137	14,589	1.20 (0.99, 1.45)	102	13,072	1.03 (0.82, 1.30)
	Former smoker	330	30,466	0.95 (0.83, 1.08)	258	30,767	0.93 (0.79, 1.09)
	Never smoker	203	17,042	1.07 (0.92, 1.24)	139	16,119	0.98 (0.81, 1.19)
	P-interaction			0.08			0.44
Lung cancer	Current smoker	296	32,492	1.22 (1.08, 1.39)	237	28,418	1.20 (1.04, 1.38)
	Former smoker	150	54,706	1.20 (0.99, 1.45)	126	52,970	1.39 (1.12, 1.73)
	Never smoker	31	60,171	1.71 (1.25, 2.54)	22	57,543	1.70 (1.05, 2.76)
	P-interaction			0.29			0.22
Postmenopausal breast cancer ^d	Current smoker	95	17,903	1.03 (0.82, 1.31)	69	15,346	1.08 (0.82, 1.42)
	Former smoker	128	24,240	0.84 (0.67, 1.06)	84	22,203	1.03 (0.77, 1.37)
	Never smoker	249	43,081	0.77 (0.67, 0.89)	171	41,377	0.88 (0.73, 1.05)
	P-interaction			0.18			0.49
Colorectal cancer	Current smoker	68	32,492	1.09 (0.83, 1.42)	52	28,418	1.21 (0.90, 1.64)
	Former smoker	125	54,706	1.18 (0.96, 1.46)	97	52,970	1.47 (1.13, 1.90)
	Never smoker	119	60,171	0.99 (0.81, 1.22)	82	57,543	0.91 (0.71, 1.17)
	P-interaction			0.65			0.10

Bladder cancer	Current smoker	43	32,492	0.80 (0.55, 1.15)	31	28,418	0.88 (0.56, 1.37)
	Former smoker	90	54,706	0.83 (0.64, 1.06)	66	52,970	0.76 (0.55, 1.06)
	Never smoker	21	60,171	0.83 (0.51, 1.33)	11	57,543	0.84 (0.43, 1.62)
	P-interaction			0.91			0.98
Kidney cancer	Current smoker	32	32,492	0.97 (0.65, 1.43)	23	28,418	0.66 (0.40, 1.08)
	Former smoker	33	54,706	1.27 (0.85, 1.88)	29	52,970	1.16 (0.71, 1.88)
	Never smoker	49	60,171	1.00 (0.74, 1.36)	44	57,543	1.30 (0.94, 1.80)
	P-interaction			0.77			0.07
Pancreatic cancer	Current smoker	26	32,492	0.83 (0.53, 1.30)	17	28,418	1.05 (0.62, 1.80)
	Former smoker	36	54,706	0.74 (0.49, 1.11)	29	52,970	0.66 (0.39, 1.10)
	Never smoker	39	60,171	1.40 (0.99, 1.98)	32	57,543	1.28 (0.85, 1.92)
	P-interaction			0.04			0.21

Abbreviations: PAC – proteomic aging clock; CaPAC – cancer-specific proteomic aging clock; BMI – body mass index; eGFR – estimated glomerular filtration rate; SD – standard deviation; HR – hazard ratio; CI – confidence interval.

^a Interaction with smoking status was examined using a multiplicative term.

^b Model was adjusted for chronological age, sex, race, study center, education, BMI, pack-years of smoking, alcohol intake, eGFR, aspirin use, hormone replacement therapy, and diabetes.

^c For analysis involved risk of prostate cancer, SDs were: CaPAC2 = 2.51, 2.55, and 2.52 years for current, former, and never smokers, respectively; and V2 ARIC PAC = 2.67, 2.89, and 2.82 years for current, former, and never smokers, respectively.

^d For analysis involved risk of postmenopausal breast cancer, SDs were: CaPAC2 = 2.73, 2.86, and 2.66 years for current, former, and never smokers, respectively; and V2 ARIC PAC = 2.86, 3.23, and 3.02 years for current, former, and never smokers, respectively.

Supplemental Table 4.6 Top 40 pathways overrepresented in the CaPAC

No.	Pathways	FDR p-value	No.	Pathways	FDR p-value
1	cell adhesion	1.203E-46	21	positive regulation of intracellular signal transduction	4.564E-22
2	negative regulation of multicellular organismal process	1.097E-26	22	regulation of protein phosphorylation	5.514E-22
3	positive regulation of multicellular organismal process	1.097E-26	23	enzyme-linked receptor protein signaling pathway	8.251E-22
4	positive regulation of signal transduction	1.347E-26	24	regulation of phosphorylation	2.858E-21
5	positive regulation of phosphorus metabolic process	3.177E-26	25	regulation of cytokine production	3.033E-21
6	positive regulation of phosphate metabolic process	3.177E-26	26	positive regulation of developmental process	3.21E-21
7	cell-cell adhesion	5.722E-26	27	regulation of phosphate metabolic process	5.812E-21
8	positive regulation of protein phosphorylation	5.722E-26	28	regulation of phosphorus metabolic process	6.032E-21
9	positive regulation of phosphorylation	6.045E-26	29	cell activation	8.299E-21
10	regulation of multicellular organismal development	8.643E-26	30	regulation of apoptotic process	9.095E-21
11	regulation of response to external stimulus	3.438E-25	31	regulation of MAPK cascade	1.61E-20
12	positive regulation of gene expression	7.673E-25	32	positive regulation of immune system process	1.645E-20
13	positive regulation of cell population proliferation	2.39E-24	33	response to wounding	3.095E-20
14	regulation of immune system process	4.894E-24	34	defense response to other organism	9.875E-20
15	positive regulation of MAPK cascade	2.438E-23	35	positive regulation of protein modification process	1.123E-19
16	locomotion	2.836E-23	36	protein phosphorylation	1.362E-19
17	cytokine production	3.591E-23	37	positive regulation of protein metabolic process	1.768E-19
18	chemotaxis	1.203E-46	38	MAPK cascade	6.575E-19
19	taxis	1.097E-26	39	regulation of cell migration	1.028E-18
20	regulation of programmed cell death	1.097E-26	40	regulation of cell adhesion	1.314E-18

Chapter 5 Manuscript 3: Proteomic Aging Clock and Risk of Mortality among older Cancer Survivors in the Atherosclerosis Risk in Communities (ARIC) Study

A. Overview

Background: Longer lifespan and improved cancer treatment led to a rapid rise in the number of cancer survivors. However, many cancer survivors have physiological dysregulations at an earlier chronological age than those without cancer, suggesting that cancer survivors' biological age is higher than their chronological age, i.e., they have accelerated aging. Cancer survivors' biological age may be estimated by a novel proteomic aging clock (PAC). The deviation of PAC from chronological age is called PAC acceleration. To our knowledge, no studies examined PACs in cancer survivors. We studied age acceleration of two PACs – a newly created cancer-specific PAC (CaPAC5) in ARIC and the published PAC developed by Lehallier [2020] (so called Lehallier's PAC) and their association with mortality in cancer survivors.

Methods: ARIC is a prospective cohort of White and Black women and men, followed in 1987-2019. In 2011-13 (Visit 5, V5), more than 5,000 proteins were measured using SomaScan, an aptamer-based assay in 806 cancer survivors and 3,699 cancer-free participants (aged 66-90 years). In a training set of two-thirds randomly selected cancer-free participants (N = 2,466), elastic net regression selected 619 aptamers for CaPAC5. We also computed Lehallier's PAC using weights estimated in ARIC. We internally validated these two PACs in the remaining 1,233 cancer-free participants (test set). PAC acceleration was calculated as residuals after regressing PAC on chronological age. We used Cox proportional hazards regression to estimate hazard ratios (HRs) and 95%

confidence intervals (CIs) for all-cause, cardiovascular disease (CVD), and cancer mortality associated with PAC acceleration in 806 survivors of all cancer types, and survivors of breast, prostate, and colorectal cancers. Also, we compared these associations to the associations in 1,233 cancer-free participants.

Results: In the test set, both PACs were correlated with chronological age [Pearson correlation coefficient (r) = 0.75 for CaPAC5 and $r = 0.70$ for Lehallier's PAC and with each other [$r=0.89$]. Age acceleration for both PACs showed similar association with all-cause mortality in cancer survivors [CaPAC5: HR (95% CI) per 1 SD = 1.42 (1.24, 1.62)]. These associations were not modified by sex or race. The associations with all-cause mortality for both PACs were similar in cancer survivors and cancer-free participants. In the Fine and Grey competing risk model, the age acceleration for CaPAC but not for Lehallier's PAC was significantly associated with cancer mortality in all cancer survivors [HR (95% CI) per 1 SD = 1.34 (1.09, 1.64)]. In addition, age acceleration for both PACs was significantly associated with all-cause mortality in breast cancer survivors, and the association remained significant after additional adjustment for stage [CaPAC5: HR (95% CI) per 1 SD = 2.68 (1.44, 4.99)]. HRs for all-cause mortality in colorectal cancer survivors were increased for age acceleration for both PACs but were unstable due to the small sample size. Neither of age acceleration was associated with all-cause mortality in prostate cancer survivors.

Conclusion: PACs hold the promise as a potential biomarker for premature mortality in cancer survivors.

B. Introduction

The increased life expectancy in the US coupled with improved cancer survival rates has led to a rapid rise in the number of cancer survivors: there are an estimated 16.9 million cancer survivors as of January 2019, which is projected to rise to 26.1 million by 2040 in the US.⁵¹ However, cancer survivors experience pre-mature mortality and multiple comorbidities.^{99,105} In general, clinicians have noticed that cancer survivors are facing accelerated aging – conceptualized as their biological age is greater than their chronological age – and this manifests as experiencing signs of aging earlier in life than what is expected.^{1,10} It has been shown that accelerated aging in cancer survivors, in first turn, may be caused by cancer treatment.^{10,106} A biological mechanism underlying the accelerated aging in cancer survivors involves stress induced by cancer therapies. This stress can induce hallmarks of aging: cellular senescence, telomere attrition, stem cell exhaustion, DNA damage, and epigenetic alterations.¹⁰⁷ In addition to cancer treatment, accelerated aging in cancer survivors may be caused by cancer-related inflammation and unhealthy lifestyle, such as smoking and lack of physical activity.

The observations of accelerated aging in cancer survivors have led to the search for measures of biological aging, which can be used to screen the aging process and predict the risk of mortality in cancer survivors. Recently, researchers have proposed biological age estimators called aging clocks, using epigenetics, transcriptomics, metabolomics, proteomics, and other biomarkers.¹² The most studied aging clocks are epigenetic clocks, which are a set of DNA methylation sites, including Horvath, Hannum, DNAm PhenoAge, and GrimAge clocks.¹⁷⁻²⁰ However, the underlying mechanisms of changes in these DNA methylation sites remain unclear.⁶³ Recently, a growing body of

research supports associations between levels of proteins and chronological age.² A proteomic aging clock (PAC) is advantageous because protein biomarkers link genotype to phenotype and can provide more information on aging and age-related pathology.^{108,109} Such PACs were shown to predict the risk of all-cause mortality in a general population in the Manuscript 1 in this dissertation and in published studies.^{29,43} However, to our knowledge, no studies tested PACs in cancer survivors.

To test PACs in cancer survivors, we constructed a cancer-specific PAC -- CaPAC5, using proteins measured in Visit 5 (V5, 2011-13) plasma samples from the Atherosclerosis Risk in Communities (ARIC) study. Recently more than 5,000 proteins were measured in these samples using SomaScan assay (v.4). We also calculated the published PAC that had been developed by Lehallier [2020]¹⁴ (so called Lehallier's PAC). We hypothesized that these two PACs were positively associated with risk of mortality in cancer survivors. We examined associations with the risk of all-cause, cardiovascular (CVD), and cancer mortality. We also compared the association with mortality in cancer survivors to the association in cancer-free participants.

C. Methods

C.1 Study population

The ARIC study is a prospective population-based cohort of White and Black men and women initiated in 1987.^{52,53} In 1987-89 (Visit 1, V1), 15,792 participants aged 45-64 years were recruited from four study centers - Maryland; Minnesota; Mississippi; and North Carolina. Participants in Minnesota and Maryland were primarily White, and the recruitment of Mississippi was restricted to Black residents. The ARIC study was

approved by institutional review boards at each participating center, and all study participants provided written informed consent. Thus far, eight study visits have been completed.⁵² Additionally, participants received follow-ups via telephone calls annually between 1987-2012 and semi-annually after 2012, with response rates of 90%-99% for the annual follow-up calls and 83%-90% for semi-annual follow-up calls among living participants who have not withdrawn consent to be contacted.⁵³

C.2 Assessment of cancer cases

Incident primary cancer was ascertained through 2015 using state Cancer registries in Minnesota, North Carolina, Maryland, and Mississippi, and supplemented by abstraction of medical records and hospital discharge summaries.⁵³ An expert panel adjudicated all cases of bladder, breast, colorectal, liver, lung, pancreatic, and prostate cancers. For adjudicated cases, when possible, stage at diagnosis was determined from the cancer registry or medical records using the pathologic TNM stage (**T**umor extent, **N**ymph **N**ode involvement, presence of **M**etastasis). When this information was not available, stage at diagnosis was determined from the cancer registry or clinical TNM stage from cancer registry or medical records according to Surveillance, Epidemiology, and End Results (SEER) summary stage.⁵³

C.3 Assessment of mortality

Mortality was identified through annual (semi-annual since 2012) follow-up telephone calls to participants or their proxies, state records and linkage to the National Death Index up to December 31, 2019.⁶⁰ The date and some causes of death were verified

by death certificate review. All-cause mortality was defined as death from any cause. CVD mortality and cancer mortality were defined based on underlying cause of death: *International Classification of Diseases, Ninth Revision*, codes (ICD-9 codes) 390–459 or *International Classification of Diseases, Tenth Revision*, codes (ICD-10 codes) I00–I99 for CVD deaths. ICD-9 code 140-239 or ICD-10 codes C00-C97 for cancer deaths.

C.4 Blood collection

The ARIC protocol for blood sample collection, processing, and storage protocol was designed to minimize the spontaneous biochemical reactions after blood collection, and is consistent with recommended practice for proteomics data analysis in epidemiological studies.^{58,64,65} Briefly, after venipuncture, blood samples were put immediately in an ice water bath. Centrifugation was then performed within 10 min after venipuncture at room temperature (15-25 °C). After centrifugation, the aliquots were stored at –80 °C within 90 min from venipuncture and were never thawed before this analysis.

C.5 Protein measurement and quality control

Samples were analyzed using a SOMAmer (Slow Off-rate Modified Aptamers)-based capture array called SomaScan® by Somalogic, Inc. (Boulder, CO, USA).⁵⁴⁻⁵⁷ The SomaScan platform uses single-stranded DNA-based aptamers to capture conformational protein epitopes.

Protein analyte measurements underwent the regular SomaScan data standardization and normalization process.^{58,59} First, hybridization control normalization

was applied to each sample based on a set of hybridization control sequences to correct for systematic biases during hybridization. Second, median signal normalization was applied to measures within a plate to remove sample or assay biases that may be because of pipetting variation, variation in reagent concentrations, assay timing, and other sources of systematic variability within a single plate run. Finally, each plate contained calibrator samples for each SOMAmer reagent, which was used to correct for plate-to-plate variation based on established global reference standards. The median Bland-Altman coefficient of variation (CVBA) for split samples was 7% at V5 after excluding aptamers with a CVBA >50% or a variance of <0.01 on the log scale, or aptamers nonspecific binding to nonproteins. After all quality control measures were completed, 4,955 aptamer measurements were included. Protein measures were reported as relative fluorescent units (RFU) and were log₂ transformed to correct for skewness.

C.6 Assessment of other participant characteristics

Other characteristics of interest included demographic and lifestyle/medical characteristics, namely chronological age, sex, race, study center, education, smoking status, alcohol intake, body mass index (BMI), aspirin use, diabetes status, and history of CVD. Sex, race, study center, and education was collected at V1, chronological age was calculated from date of birth until date of V5 and all the other variables listed above were collected at each visit.⁷⁶ At each visit, each participant reported information on smoking, alcohol intake, medical history, and use of medications and underwent a physical exam that included measurement of height and weight. BMI was calculated as weight (kg) divided by height (in meters) squared. Diabetes mellitus was defined as fasting glucose

≥ 126 mg/dL, non-fasting glucose ≥ 200 mg/dL, self-reported use of medication to treat for diabetes, or self-reported physician diagnosis of diabetes. CVD events included 1) heart failure (HF), 2) definite or probable stroke, or 3) coronary heart disease (CHD), defined as definite or probable myocardial infarction (MI) or definite fatal CHD.^{85,86} Detailed procedures for assessing HF, stroke, and CHD in ARIC have been published.⁸⁹⁻⁹¹ Another variable of interest was estimated glomerular filtration rate (eGFR) because it associated with different proteins levels.^{110,111} eGFR at V5 was calculated based on equations that utilized serum creatinine and cystatin C and incorporated age and sex.⁷⁹

C.7 Statistical analysis

PAC was constructed using R (version 4.1.2, package “glmnet”) and all the other analyses were performed using SAS version 9.4 (SAS Institute Inc, Cary, NC).

Among 6,538 participants who attended V5 (chronological age: 66-90 years), we excluded participants without protein measures at V5 (N=1,343), participants with race other than White or Black (N=12), participants who diagnosed with a cancer before V1 or within two years before V5 (N=678), resulting in 4,505 White and Black participants (806 cancer survivors and 3,699 participants who were cancer-free at the end of 2015 (called “cancer-free participants” in this analysis)). We excluded participants who had been diagnosed with a cancer within two years before V5 to minimize the number of cancer survivors undergoing active treatment.

C.7.1 Construction and validation of PAC

Among the 3,699 cancer-free participants, we randomly selected two-thirds of participants (N=2,466) and used them as the training set and the remaining one-third of

participants were used as the test set (N=1,233). Using the training set, we applied elastic net regression to train the CaPAC5 against chronological age. The elastic net regression (alpha = 0.5 and hyperparameter value was selected based on 10-fold cross-validation) selected 619 aptamers out of the 4,955 aptamers for CaPAC5.

In addition to the newly constructed CaPAC5, we also computed Lehallier's PAC,¹⁴ one of the published PAC. We chose Lehallier's PAC rather than the other two published PACs (Tanaka's and Sathyan's PACs),^{13,43} because Lehallier's PAC was more strongly correlated with chronological age (r=0.96) compared to other published PACs. We estimated ARIC weights for Lehallier's PAC instead of using the published weights, because ARIC only included 415 aptamers among the 491 aptamers reported in Lehallier's PAC.¹⁴ Using the 415 aptamers available in ARIC, we estimated ARIC weights by applying Ridge regression (hyperparameter value was selected based on 10-fold cross-validation).

We internally tested these two PACs in cancer-free participants in the test set. PAC for each participant was calculated using the weighted sum of their aptamer levels: $PAC = \beta_0 + \sum_{i=1}^n \beta_i \times aptamer_i$, where $aptamer_i$ is the level of i th aptamer and the intercept (β_0) and weights (β_i) were estimated using the training set. We estimated the performance of each PAC by measuring the Pearson correlation between PAC and chronological age in cancer-free participants in the test set. In addition, we also calculated the median absolute error (MAE) between PAC and chronological age.

C.7.2 Association between participant characteristics and PAC

To examine the cross-sectional association between participant characteristics at

V5 and PAC in cancer survivors and in cancer-free participants (test set), we calculated CaPAC5 and Lehallier's PAC in 806 cancer survivors and 1,233 cancer-free participants (test set). To capture associations that are independent of chronological age, we created PAC acceleration,³⁵ estimated as the residual for each participant after regressing each PAC (CaPAC5 and Lehallier's PAC) on chronological age in a group that combined 806 cancer survivors and 1,233 cancer-free participants (test set). Demographic and lifestyle/medical characteristics at V5 were examined across tertiles of PAC acceleration as the mean (standard deviation, SD) or percentage (%) in cancer survivors and in cancer-free participants (test set).

C.7.3 Associations of PAC acceleration with risk of all-cause, CVD and cancer mortality

We used Cox proportional hazards regression to calculate hazard ratios (HRs) and 95% confidence intervals (CIs) for all-cause mortality and CVD mortality in cancer survivors and in cancer-free participants, and cancer mortality in cancer survivors in relation to PAC acceleration (modeled as a continuous variable). For each participant, total person-years was determined from date of V5 until death or administrative censoring on December 31, 2019, whichever occurred first. The proportional hazards assumption was examined by modeling an interaction between PAC acceleration and follow-up time. The proportional hazards assumption was not violated in any regression models.

We constructed three models. Model 1 was a demographic model adjusted for chronological age, sex, race, and study center. Model 2 was additionally adjusted for education, BMI, smoking status, alcohol intake, time since cancer diagnosis (for cancer

survivors only) and eGFR. Model 3 (fully adjusted model) was additionally adjusted for history of CVD, aspirin use, and diabetes status. We did not adjust for stage at diagnosis, because we do not have stage at diagnosis for all the survivors, however, we adjusted for stage at diagnosis in the analysis of most common individual cancers as described below. We compared the associations in cancer survivors and in cancer-free participants (test set) by calculating the p-value for multiplicative interaction between PAC acceleration and cancer status in Model 3.

We also examined whether sex or race modified the association between PAC acceleration and all-cause mortality in cancer survivors and in cancer-free participants (test set), because the rate of biological aging in males may exceed that of females⁷⁵ and Black individuals have higher death rates than White individuals in the U.S.¹¹² For cancer survivors' analysis stratified by sex, we also created a category of non-sex-related cancers including all cancers that could have occurred in both males and females, and excluding breast, cervical, endometrial, ovarian, and prostate cancers. We did not examine whether sex or race modify the association with CVD mortality or cancer mortality because of the small number of deaths attributed to CVD or cancer during the follow-up period. Interactions with sex and race were examined using a multiplicative term.

In addition, we also examined the association between PAC acceleration and risk of all-cause mortality in survivors of most common cancer types such as breast, prostate, and colorectal cancers in Models 1-3. We did not include survivors of other cancers, e.g., lung cancer because of a small number of survivors (N=25). In the analysis of individual cancer survivors, we additionally adjusted Model 3 for stage at diagnosis when

calculating the HRs for all-cause mortality.

C.8 Sensitivity analysis

In a sensitivity analysis of CVD mortality and cancer mortality, deaths from other causes were treated as competing events using Fine and Gray method.^{72,73}

C.9 Exploratory analysis

Using ToppGene,⁶⁸⁻⁷⁰ we conducted an overrepresentation analysis (ORA) for the proteins included in CaPAC5. ORA is a type of pathway analysis that shows whether proteins from pre-defined pathways, for example those participating in the same molecular process or are regulated by the same transcription factor (i.e., to a specific GO term or KEGG pathways) are present more than would be expected (overrepresented) in our data.⁷¹ UniProt IDs were provided as the inputs, the background was set to all protein-coding genes, and the False Discovery Rate (FDR) significance level was set to 0.05.

D. Results

D.1 Pearson correlation between PAC and chronological age

Among cancer-free participants, the Pearson correlation coefficients (r) between CaPAC5 and chronological age were 0.89 (median absolute error (MAE)=1.76 years) in the training set and 0.75 (MAE=2.19 years) in the test set (**Figure 5.1a**), respectively. Pearson's r between Lehallier's PAC and chronological age was 0.80 (MAE=2.21 years) in the training set and 0.70 (MAE = 2.50 years), in the test set (**Figure 5.1c**). In the test

set, Pearson's r between CaPAC5 and Lehallier's PAC was 0.89.

Among cancer survivors, Pearson's r between chronological age and CaPAC5 was 0.75 (MAE=2.45 years) and Pearson's r between chronological age and Lehallier's PAC was 0.70 (MAE=2.48 years) (**Figures 5.1b and 5.1d**). Among cancer survivors, Pearson's r between CaPAC5 and Lehallier's PAC was 0.89.

D.2 Distribution of participant characteristics across tertiles of PAC acceleration

The distributions of age acceleration for CaPAC5 and Lehallier's PAC among 806 White and Black cancer survivors and 1,233 White and Black cancer-free participants (test set) are shown in **Supplemental Figure 5.1**. The means for CaPAC5 and Lehallier's PAC adjusted for chronological age were 76.1 and 76.2 years for cancer survivors and cancer-free participants (test set), respectively. The ranges of CaPAC5 acceleration were -6.68 to 15.07 years for cancer survivors and -9.86 to 15.28 years for cancer-free participants (test set), respectively (**Table 5.1**). Among cancer survivors, those with higher age acceleration for both PACs were more likely to be White, and have CVD, less education, a lower level of physical activity, and a lower eGFR (**Table 5.1**). Similar to cancer survivors, among cancer-free participants, those with higher acceleration for both PACs were more likely to be White, and have CVD, a lower level of physical activity, and a lower eGFR, but the education level was similar across PAC tertiles. In addition, cancer-free participants (test set) with a higher age acceleration for CaPAC5 but not for Lehallier's PAC tended to have diabetes (**Table 5.1**).

D.3 Associations between PAC acceleration and mortality

D.3.1 Associations with all-cause, CVD, and cancer mortality in all cancer survivors

Among 806 cancer survivors, 272 died by 2019 (median follow-up = 6.98 years). In cancer survivors, age acceleration for both PACs was similarly associated with all-cause mortality. HRs (95% CIs) per 1 SD for all-cause mortality in Model 3 were 1.42 (1.24, 1.62) for CaPAC5 and 1.40 (1.22, 1.61) for Lehallier's PAC (**Table 5.2 and Supplemental Table 5.1**). The associations with all-cause mortality in cancer survivors for both PACs were not statistically modified by sex or race (**Supplemental Table 5.2**). In addition, in cancer survivors, acceleration for both PACs was significantly associated with CVD mortality and cancer mortality in Model 3. For CaPAC5: HRs (95% CIs) per 1 SD were 1.35 (1.03, 1.76) for CVD mortality and 1.45 (1.17, 1.81) for cancer mortality in cancer survivors (**Table 5.2 and Supplemental Table 5.1**).

D.3.2 Associations with all-cause mortality in survivors of individual cancer types

Age acceleration for both PACs were significantly associated with all-cause mortality in postmenopausal breast cancer survivors. For CaPAC5: HRs (95% CIs) per 1 SD was 1.54 (1.05, 2.25) in Model 3 (**Table 5.3 and Supplemental Table 5.1**). In addition, CaPAC5 acceleration was significantly associated with all-cause mortality in colorectal cancer (CRC) survivors [HRs (95% CIs) per 1 SD = 1.96 (1.19, 3.22)], but not age acceleration for Lehallier's PAC (**Table 5.3 and Supplemental Table 5.1**). Neither of age accelerations was associated with all-cause mortality in prostate cancer survivors (**Table 5.3 and Supplemental Table 5.1**). In the same group of participants, additional adjustment for stage at diagnosis did not change the association between age acceleration

for both PACs and all-cause mortality among postmenopausal breast cancer survivors or prostate cancer survivors and the associations for CRC survivors for both PACs were in the same direction (**Supplemental Table 5.3**).

D.3.3 Association with all-cause and CVD mortality in cancer-free participants

Among 1,233 cancer-free participants (test set), 224 died by 2019 (median follow-up = 7.25 years). Age acceleration for both PACs were associated with all-cause mortality: HRs (95% CIs) per 1 SD were 1.50 (1.28, 1.76) for CaPAC5 and 1.61 (1.39, 1.87) for Lehallier's PAC in Model 3 (**Table 5.2 and Supplemental Table 5.1**). These associations were not statistically modified by sex or race (**Supplemental Table 5.2**). Age acceleration for both PACs was also associated with CVD mortality: HRs (95% CIs) per 1 SD were 1.48 (1.51, 1.91) for CaPAC5 and 1.68 (1.33, 2.12) for Lehallier's PAC (**Table 5.2 and Supplemental Table 5.1**).

D.3.4 Comparison of associations with all-cause and CVD mortality in cancer survivors and cancer-free participants

When cancer survivors and cancer-free participants (test set) were included into the same model, cancer status did not modify the associations of age acceleration for both PAC with all-cause mortality or CVD mortality (p-interactions = 0.11-0.62) (**Table 5.2**).

D.4 Sensitivity analysis

Among cancer survivors, after accounting for non-CVD deaths as competing events, the association with CVD mortality became nonsignificant for the age

acceleration for both PACs: HRs (95% CIs) per 1 SD were 1.16 (0.88, 1.52) for CaPAC5 and 1.15 (0.88, 1.49) for Lehallier's PAC (**Table 5.2**). Among cancer survivors, after accounting for non-cancer deaths as competing events using the Fine and Gray method, the association with cancer mortality remained significant but became weaker for CaPAC5 acceleration [HR (95% CI) per 1 SD = 1.34 (1.09, 1.64)] and became nonsignificant for Lehallier's PAC [HR (95% CI) per 1 SD = 1.19 (0.94, 1.51)] (**Table 5.2**).

Among cancer-free participants (test set), after accounting for non-CVD deaths as competing events, the association with CVD mortality remained the same (compared to Cox proportional hazards model) for the age acceleration for both PACs: HRs (95% CIs) per 1 SD were 1.42 (1.03, 1.96) for CaPAC5 and 1.61(1.22, 2.13) (**Table 5.2**).

D.5 Exploratory analysis

Using the Gene Ontology Biological Process database, we found that 1290 pathways were overrepresented (FDR p-value < 0.05) for the proteins included in CaPAC5. The Top 40 pathways overrepresented were listed in **Supplemental Table 5.4**. They included “inflammatory response”, “regulation of immune system process”, “positive regulation of cell differentiation”, and “regulation of nervous system development”.

E. Discussion

In a large prospective study of White and Black men and women aged 66-90 years that measured more than 5,000 proteins using the SomaScan assay, we constructed

and validated a cancer-specific PAC, CaPAC5, in older cancer-free participants. In the cancer-free participants (test set), the correlation between CaPAC5 and chronological age was 0.75. We found that CaPAC5 acceleration was associated with risk of all-cause mortality and cancer mortality (in both Cox proportional hazards model and the competing risk model) in cancer survivors. We also found that CaPAC5 acceleration was associated with risk of all-cause mortality and CVD mortality (in both Cox proportion hazards model and the competing risk model) in cancer-free participants. The associations of CaPAC5 acceleration with all-cause mortality were similar in cancer survivors and cancer-free participants. CaPAC5 was strongly correlated with Lehallier's PAC (published PAC) and these two PACs were similarly associated with all-cause mortality in cancer survivors.

In this study, we found that the Pearson correlation coefficients between CaPAC5 and chronological age were similar among cancer survivors and cancer-free participants. Although cancer survivors had a slightly higher median absolute error (MAE) compared to cancer-free participants, the difference in MAE between cancer survivors and cancer-free participants (in the test set) was very small. This may be because 80% of the cancer survivors in our study have survived for five or more years. A previous ARIC study found that the overall health of cancer survivors who survived five or more years was similar to their overall health before their cancer diagnosis.¹¹³

In this study, we compared our CaPAC5 with the published Lehallier's PAC. The acceleration for both PACs showed similar association with participant characteristics in cancer survivors and cancer-free participants. Those with higher acceleration were likely to be White, have a lower physical activity, a history of CVD, and a lower eGFR. The

age acceleration for these PACs showed similar associations with all-cause mortality in cancer survivors. After accounting for deaths from other causes as competing events, the associations of age acceleration for both PACs with CVD mortality and cancer mortality in cancer survivors became weaker or nonsignificant, indicating that non-CVD deaths and non-cancer deaths are competing events for CVD mortality and cancer mortality, respectively, in cancer survivors. In addition, the age acceleration for both PACs was significantly associated with all-cause mortality in survivors of postmenopausal breast cancer even after additional adjustment for stage. To our knowledge, our study is the first study that examined the association between PACs and risk of mortality in cancer survivors. A few studies examined the association with mortality in cancer survivors using epigenetic clocks. A pooled analysis of seven prospective studies of adult cancer survivors [mean chronological age (SD) = 59.6 (7.5), 99.3% participants aged 40-69] found that age acceleration for Hannum clock was associated with all-cause mortality [per 5 year: HR (95% CI)=1.05 (1.01, 1.10)],³⁶ while in our study that association was much stronger [HRs (95% CIs) per 5 year = 1.96 (1.52, 2.54) for CaPAC5 and 1.91 (1.46, 2.51) for Lehallier's PAC]. Of note, here we used the unit of PAC acceleration of 5 years rather than 1 SD (SDs=2.59 years for CaPAC5 and 2.58 years for Lehallier's PAC) to compare our findings to the findings in published studies. Another study, the Cancer Genome Atlas (TCGA) study [63.5% patients \geq 60 years] found that positive vs. negative acceleration for Horvath's clock was associated with a 97% increase in risk of all-cause mortality in CRC survivors [positive Horvath age acceleration vs. negative: HR (95% CI) = 1.97 (1.14, 1.39)] after adjusting for stage-and molecular subtype.³⁹ In our study, we found that CaPAC5 acceleration was significantly associated with all-cause

mortality in CRC survivors in a multivariable-adjusted model but without adjustment for stage at diagnosis [HR (95% CI) per 1 SD = 1.96 (1.19, 3.22)]. However, after additional adjustment for stage, the association with all-cause mortality in CRC survivors became nonsignificant, which may be due to the small number of CRC survivors with stage information (N=52) in our study. In summary, CaPAC5 associated with all-cause mortality in cancer survivors. However, we cannot conclude whether PAC acceleration was stronger associated with all-cause mortality than age acceleration for epigenetic clocks in cancer survivors, because we have an older study population than in the published studies.

There is a growing concern about accelerated aging among the rapidly increasing number of cancer survivors. Until recently, measures of biological age have been mostly limited to functional measures, e.g., walking speed, maximal oxygen consumption, Fried frailty phenotype,¹¹⁴ clinical geriatric assessment, and cognitive assessments.^{10,115} These measures are unlikely to account for all aspects of aging and it is time-consuming and difficult to collect them, especially in the elderly. However, PACs are objective measures and can be measured in a small drop of blood. In the future, PAC could help screen for the aging process in routine follow-ups. Although the association between PAC acceleration and mortality was similar in those with and without cancer in our study, PAC may be especially important for cancer survivors to inform on cancer treatment and potentially on the use of anti-aging interventions. Targeting biological age in cancer survivors may be advantageous, because it can simultaneously decrease the incidence of several age-related diseases together and prolong cancer survivors' healthy lifespan. PACs could also serve as surrogate endpoints in clinical trials to test anti-aging lifestyle interventions and drugs. Lastly,

the proteins and protein pathways comprising PACs can serve as targets for novel anti-aging therapies for cancer survivors.

A limitation of our study is that cancer survivors markedly differed in the amount of time between cancer diagnosis and the time of blood collection (range: 2-25 years). However, the adjustment for time since diagnosis did not noticeably change the association between age acceleration and mortality. Other limitations are that we had a small group of individual cancer survivors, and we did not have information about cancer treatment and information about stage of diagnosis was available for survivors of several cancer types only. However, additional adjustment for stage at diagnosis did not change the association between PAC acceleration and all-cause mortality in postmenopausal breast cancer survivors or prostate cancer survivors, or direction of the association for CRC survivors. Further, the possibility of protein degradation during long-term storage cannot be excluded. However, the blood samples were frozen right after their collection and have never been thawed limiting the possibility of degradation.⁷⁶ Lastly, SomaScan provide relative quantification instead of absolute quantification.¹⁰⁴ The strengths of this population-based study include the prospective design, a large population-based sample of White and Black individuals, and, validated information about cancer incidence and some causes of death. Another strength is that the version of SomaScan assay used in ARIC measured more than 5000 proteins, greater than the panels of proteins used in most previous studies. With a larger panel of proteins, we may be able to develop a PAC that more accurately predicts chronological age. In addition, we compared the associations in cancer survivors and in cancer-free participants.

In conclusion, we constructed CaPAC5 in White and Black individuals and showed that this PAC was able to predict risk of mortality in cancer survivors. The proteins and protein pathways comprising CaPAC5 hold promise as anti-aging drug targets for cancer survivors. Future studies are needed to validate our CaPAC5 in a larger population and examine associations with age-related diseases in cancer survivors.

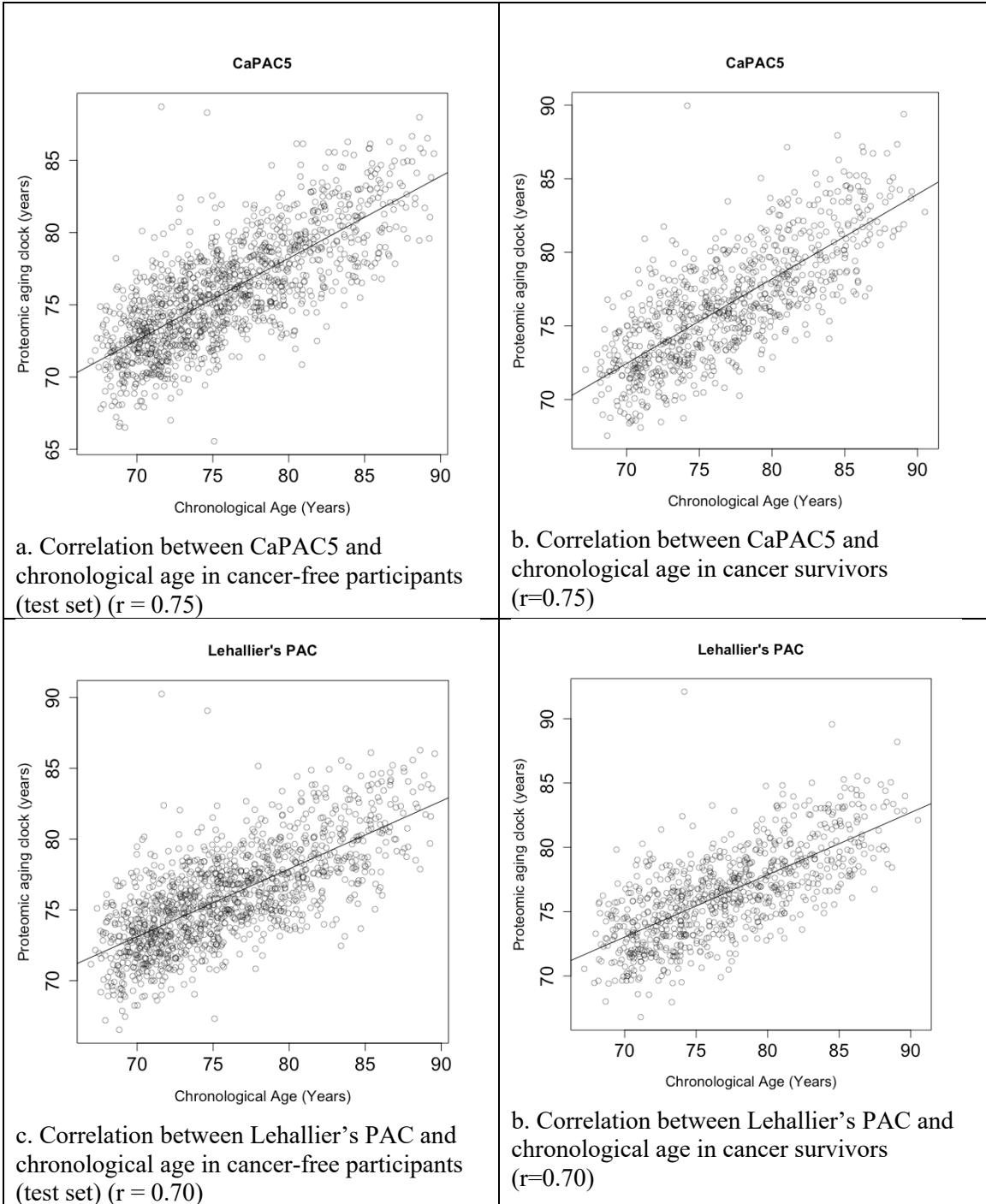


Figure 5.1 Pearson correlation (r) between PAC and chronological age in cancer-free participants and cancer survivors, Visit 5

Table 5.1 Participant characteristics across tertiles of age acceleration for PAC in cancer survivors and cancer-free participants; Visit 5

	Cancer survivors							
	CaPAC5				Lehallier's PAC			
	T1 (N = 268)	T2 (N = 269)	T3 (N = 269)	P-value	T1 (N = 268)	T2 (N = 269)	T3 (N = 269)	P-value
Range of age acceleration (years)	-6.7 to -1.3	-1.3 to 0.9	1.0 to 15.1		-6.8 to -1.3	-1.2 to 1.0	1.1 to 17.0	
Demographic								
Age, years (SD)	77.1 (4.9)	76.9 (5.2)	77.1 (5.4)	0.84	77.0 (5.2)	76.8 (4.8)	77.2 (5.5)	0.63
Male, %	52.2	58.4	49.8	0.12	58.2	50.2	52.0	0.15
White, %	76.5	80.7	86.3	0.02	76.9	80.7	85.9	0.03
Education, %								
Less than high school	12.7	12.3	19.0		11.2	15.2	17.5	
High school equivalent	41.4	39.4	41.3	0.10	38.8	42.7	40.5	0.15
Greater than high school	45.9	48.3	39.8		50.0	42.0	42.0	
Lifestyle/Medical factors								
BMI, kg/m ² (SD)	29.5 (5.3)	28.9 (5.7)	27.9 (6.0)	<0.01	29.6 (5.2)	29.0 (5.4)	27.8 (6.3)	<0.01
Smoking status, %								
Current smoker	5.0	6.0	7.3		3.7	5.8	8.7	
Former smoker	55.0	58.2	54.3	0.73	56.3	56.2	55.0	0.22
Never smoker	40.1	35.7	38.5		40.0	37.9	36.3	
Alcohol intake, %								
Current drinker	52.1	55.0	53.2		54.5	52.3	53.6	
Former drinker	30.0	29.1	28.8	0.95	31.0	28.7	28.2	0.70
Never drinker	17.9	15.9	18.0		14.5	19.0	18.3	
Physical activity, score (SD)	2.6 (0.8)	2.5 (0.8)	2.4 (0.8)	<0.01	2.7 (0.7)	2.5 (0.8)	2.4 (0.8)	<0.01
CVD, %	21.4	29.7	34.9	<0.01	16.8	29.4	39.8	<0.01
Diabetes, %	38.8	31.2	35.7	0.18	33.2	37.9	34.6	0.50
Aspirin use in the past two weeks, %	70.8	71.9	70.2	0.90	67.4	72.0	73.4	0.28
eGFR, mL/min/1.73 m ² (SD)	71.0 (18.1)	65.2 (18.4)	60.0 (20.3)	<0.01	70.9 (17.9)	65.6 (18.4)	59.6 (0.4)	<0.01
Cancer-free participants (test set)								
	CaPAC5				Lehallier's PAC			
	T1 (N = 411)	T2 (N = 411)	T3 (N = 411)	P-value	T1 (N = 411)	T2 (N = 411)	T3 (N = 411)	P-value
Range of age acceleration (years)	-9.8 to -1.2	-1.1 to 1.0	1.1 to 15.3		-8.2 to -1.2	-1.1 to 1.0	1.1 to 16.4	

Demographic									
Age, years (SD)	76.1 (5.2)	75.5 (5.0)	76.3 (5.3)	0.06	76.0 (5.1)	75.8 (5.1)	76.1 (5.3)	0.73	
Male, %	43.6	40.6	40.2	0.56	39.2	43.3	41.9	0.47	
White, %	75.2	80.5	82.7	0.04	75.4	79.8	83.2	0.02	
Education, %									
Less than high school	12.9	13.6	14.7		12.4	14.9	13.9		
High school equivalent	40.9	42.8	42.1	0.89	43.8	38.3	43.7	0.44	
Greater than high school	46.2	43.6	43.3		43.8	46.8	42.4		
Lifestyle/Medical factors									
BMI, kg/m ² (SD)	29.4 (5.3)	29.0 (5.8)	27.6 (5.7)	<0.01	29.1 (4.8)	29.0 (6.1)	28.0 (5.9)	0.01	
Smoking status, %									
Current smoker	5.8	5.5	7.2		5.2	6.5	6.8		
Former smoker	55.1	52.9	47.0	0.25	51.5	53.9	49.4	0.64	
Never smoker	39.1	41.6	45.8		43.3	39.5	43.8		
Alcohol intake, %									
Current drinker	48.5	48.1	54.1		46.3	51.4	53.0		
Former drinker	30.8	28.7	23.8	0.21	31.5	27.7	23.9	0.18	
Never drinker	20.8	23.3	22.2		22.3	20.9	23.1		
Physical activity, score (SD)	2.7 (0.8)	2.7 (0.8)	2.5 (0.8)	0.01	2.7 (0.8)	2.6 (0.8)	2.6 (0.8)	0.01	
CVD, %	19.9	21.7	29.8	<0.01	16.6	22.9	31.9	<0.01	
Diabetes, %	34.8	30.2	26.0	0.02	32.9	29.9	28.2	0.35	
Aspirin use in the past two weeks, %	69.9	66.9	70.2	0.56	68.4	67.5	71.2	0.49	
eGFR, mL/min/1.73 m ² (SD)	73.7 (15.8)	69.8 (19.3)	65.5 (18.8)	<0.01	72.7 (17.0)	71.4 (17.5)	64.9 (19.5)	<0.01	

Abbreviations: V5 – Visit 5; PAC – proteomic ageing clock; CaPAC – cancer-specific proteomic aging clock; SD – standard deviation; CVD – cardiovascular disease; BMI – body mass index; eGFR - estimated glomerular filtration rate.

Table 5.2 Associations of PAC acceleration with all-cause and CVD mortality among cancer survivors and cancer-free participants, and cancer mortality among cancer survivors

		No of deaths	Total person-years	CaPAC5 HR (95% CI) per 1 SD ^a Model 3 ^b	Lehallier's PAC HR (95% CI) per 1 SD ^a Model 3 ^b
All-cause mortality	cancer survivors	272	4,963	1.42 (1.24, 1.62)	1.40 (1.22, 1.61)
	cancer-free participants (test set)	224	8,247	1.50 (1.28, 1.76)	1.61 (1.39, 1.87)
	p-interaction ^c			0.62	0.20
CVD mortality	cancer survivors	75	4,963	1.35 (1.03, 1.76)	1.36 (1.02, 1.82)
	cancer-free participants (test set)	82	8,247	1.48 (1.15, 1.91)	1.68 (1.33, 2.12)
	p-interaction			0.46	0.11
Cancer mortality	cancer survivors	86	4963	1.45 (1.17, 1.81)	1.29 (1.02, 1.63)
Fine and gray competing risk model					
CVD mortality	cancer survivors			1.16 (0.88, 1.52)	1.15 (0.88, 1.49)
	cancer-free participants (test set)			1.42 (1.03, 1.96)	1.61 (1.22, 2.13)
Cancer mortality	cancer survivors			1.34 (1.09, 1.64)	1.19 (0.94, 1.51)

Abbreviations: PAC – proteomic ageing clock; CaPAC – cancer-specific proteomic aging clock; CVD – cardiovascular disease; SD – standard deviation; BMI – body mass index; eGFR – estimated glomerular filtration rate; HR – hazard ratio; CI – confidence interval.

^a SDs for PAC acceleration were: among cancer survivors = 2.59 and 2.58 years for CaPAC5 and Lehallier's PAC, respectively; and among cancer-free participants (test set) = 2.55 and 2.49 years for CaPAC5 and Lehallier's PAC, respectively.

^b Model adjusted for chronological age, sex, race, study center, education, BMI, smoking status, alcohol intake, physical activity, time since cancer diagnosis (for cancer survivors' analysis only), eGFR, history of CVD, aspirin use, and diabetes.

^c Interaction with cancer status was examined using a multiplicative term.

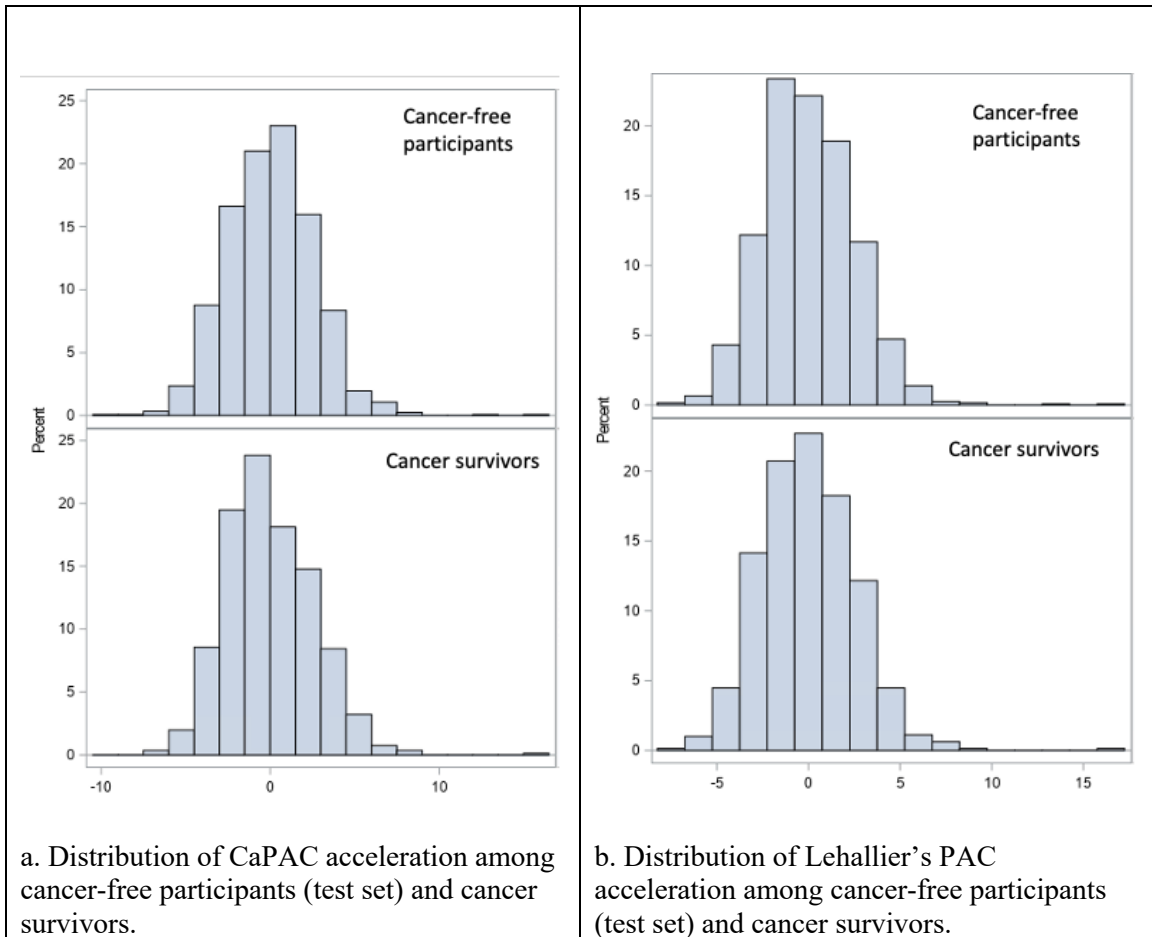
Table 5.3 Associations of PAC acceleration with all-cause mortality among survivors of postmenopausal breast, prostate, and colorectal cancers in Model 3

	No of deaths	Total person-years	CaPAC5 HR (95% CI) per 1 SD ^a Model 3 ^b	Lehallier's PAC HR (95% CI) per 1 SD ^a Model 3 ^b
Postmenopausal breast cancer survivors	46	1,114	1.54 (1.05, 2.25)	1.72 (1.13, 2.64)
Prostate cancer survivors	88	1,551	1.19 (0.92, 1.53)	1.28 (0.97, 1.69)
CRC survivors	35	412	1.96 (1.19, 3.22)	1.38 (0.87, 2.17)

Abbreviations: PAC – proteomic ageing clock; CaPAC – cancer-specific proteomic aging clock; CVD – cardiovascular disease; SD – standard deviation; CRC – colorectal cancer; BMI – body mass index; eGFR – estimated glomerular filtration rate; HR – hazard ratio; CI – confidence interval.

^a SDs for PAC acceleration were: CaPAC5 = 2.60, 2.45, and 2.52 years for breast, prostate, and colorectal cancer survivors, respectively; and Lehallier's PAC = 2.42, 2.57, and 2.78 years for breast, prostate, and colorectal cancer survivors, respectively.

^b Model 3 was adjusted for chronological age, sex (for CRC survivors' analysis only), race, study center, education, BMI, smoking status, alcohol intake, physical activity, time since cancer diagnosis, eGFR, history of CVD, aspirin use, and diabetes.



Supplemental Figure 5.1 Distribution of PAC acceleration among cancer-free participants (test set) and cancer survivors, Visit 5

Supplemental Table 5.1 Associations of PAC acceleration with all-cause and CVD mortality among cancer survivors and cancer-free participants, and cancer mortality among cancer survivors in Models 1-3

		CaPAC5			Lehallier's PAC		
		HR (95% CI) per 1 SD ^a					
		Model 1 ^b	Model 2 ^c	Model 3 ^d	Model 1 ^b	Model 2 ^c	Model 3 ^d
All-cause mortality	All cancer survivors	1.53 (1.36, 1.72)	1.43 (1.25, 1.63)	1.42 (1.24, 1.62)	1.51 (1.34, 1.70)	1.44 (1.25, 1.65)	1.40 (1.22, 1.61)
	Cancer-free participants (test set)	1.58 (1.41, 1.79)	1.50 (1.29, 1.75)	1.50 (1.28, 1.76)	1.66 (1.48, 1.85)	1.58 (1.37, 1.82)	1.61 (1.39, 1.87)
CVD mortality	All cancer survivors	1.51 (1.20, 1.90)	1.33 (1.02, 1.73)	1.35 (1.03, 1.76)	1.53 (1.22, 1.93)	1.42 (1.07, 1.89)	1.36 (1.02, 1.82)
	Cancer-free participants (test set)	1.59 (1.31, 1.92)	1.51 (1.19, 1.91)	1.48 (1.15, 1.91)	1.73 (1.46, 2.06)	1.68 (1.36, 2.06)	1.68 (1.33, 2.12)
Cancer mortality	All cancer survivors	1.47 (1.20, 1.81)	1.46 (1.17, 1.82)	1.45 (1.17, 1.81)	1.29 (1.04, 1.60)	1.30 (1.03, .64)	1.29 (1.02, 1.63)
All-cause mortality for survivors of postmenopausal breast, prostate, and colorectal cancers							
Postmenopausal breast cancer survivors		1.46 (1.10, 1.93)	1.54 (1.06, 2.23)	1.54 (1.05, 2.25)	1.50 (1.10, 2.04)	1.77 (1.18, 2.66)	1.72 (1.13, 2.64)
Prostate cancer survivors		1.25 (1.01, 1.55)	1.22 (0.94, 1.58)	1.19 (0.92, 1.53)	1.36 (1.09, 1.71)	1.36 (1.04, 1.77)	1.28 (0.97, 1.69)
CRC survivors		1.69 (1.19, 2.40)	1.87 (1.23, 2.84)	1.96 (1.19, 3.22)	1.37 (0.97, 1.93)	1.43 (0.96, 2.13)	1.38 (0.87, 2.17)

Abbreviations: PAC – proteomic ageing clock; CaPAC – cancer-specific proteomic aging clock; CVD – cardiovascular disease; SD – standard deviation; CRC – colorectal cancer; BMI – body mass index; eGFR – estimated glomerular filtration rate; HR – hazard ratio; CI – confidence interval.

^a Among cancer survivors, SDs for PAC acceleration were: CaPAC5 = 2.59, 2.60, 2.45, and 2.52 years for all, breast, prostate, and colorectal cancer survivors, respectively; and Lehallier's PAC = 2.58, 2.42, 2.57, and 2.78 years for all, breast, prostate, and colorectal cancer survivors, respectively. Among cancer-free participants (test set), SDs for PAC acceleration were 2.55 and 2.49 years for CaPAC5 and Lehallier's PAC, respectively.

^b Model 1 was adjusted for chronological age, sex, race, and study center

^c Model 2 was adjusted for chronological age, sex, race, study center, education, BMI, smoking status, alcohol intake, physical activity, time since cancer diagnosis (for cancer survivors' analysis only), and eGFR.

^d Model 3 was adjusted for chronological age, sex, race, study center, education, BMI, smoking status, alcohol intake, physical activity, time since cancer diagnosis (for cancer survivors' analysis only), eGFR, history of CVD diseases, aspirin use, and diabetes.

Supplemental Table 5.2 Associations of PAC acceleration with all-cause mortality among cancer survivors and cancer-free participants stratified by race and sex

		No. of death	Total person-years	CaPAC5 HR (95% CI) ^a per 1 SD	Lehallier's PAC HR (95% CI) ^a per 1 SD
Stratified by race ^b					
All cancer survivors ^c	White	219	4,204	1.41 (1.22, 1.64)	1.39 (1.19, 1.63)
	Black	53	758	1.30 (0.96, 1.76)	1.29 (0.95, 1.76)
	P-interaction			0.77	0.47
Cancer-free participants (test set) ^d	White	178	6,889	1.48 (1.22, 1.79)	1.52 (1.27, 1.82)
	Black	46	1,358	1.60 (1.12, 2.27)	2.05 (1.52, 2.76)
	P-interaction			0.82	0.32
Stratified by sex ^b					
All cancer survivors ^c	Women	115	2,351	1.38 (1.12, 1.70)	1.33 (1.06, 1.68)
	Men	157	2,612	1.40 (1.17, 1.68)	1.46 (1.20, 1.76)
	P-interaction			0.80	0.73
Non-sex-related cancer survivors ^f	Women	58	868	1.62 (1.16, 2.24)	1.26 (0.90, 1.76)
	Men	69	1,054	2.20 (1.52, 2.69)	1.89 (1.43, 2.49)
	P-interaction			0.47	0.22
Cancer-free participants (test set) ^g	Women	120	4,816	1.57 (1.26, 1.96)	1.72 (1.39, 2.12)
	Men	104	3,431	1.42 (1.11, 1.81)	1.49 (1.18, 1.87)
	P-interaction			0.32	0.23

Abbreviations: PAC – proteomic ageing clock; CaPAC – cancer-specific proteomic aging clock; CVD – cardiovascular disease; SD – standard deviation; BMI – body mass index; eGFR – estimated glomerular filtration rate; HR – hazard ratio; CI – confidence interval.

^a model was adjusted for chronological age, sex, race, study center, education, BMI, smoking status, alcohol intake, physical activity, time since cancer diagnosis (for cancer survivors' analysis only), eGFR, history of CVD diseases, aspirin use, and diabetes.

^b Interactions with sex and race were examined using a multiplicative term.

^c Among cancer survivors, SDs for PAC acceleration were: CaPAC5 = 2.61 and 2.55 years for White and Black participants, respectively; and Lehallier's PAC: 2.55 and 2.67 years for White and Black participants, respectively.

^d Among cancer-free participants (test set), SDs for PAC acceleration were: CaPAC5 = 2.43 and 2.96 years for White and Black participants, respectively; and Lehallier's PAC = 2.34 and 2.97 years for White and Black participants, respectively.

^e Among all cancer survivors, SDs for PAC acceleration were: CaPAC5 = 2.58 and 2.60 years for females and males, respectively; and Lehallier's PAC = 2.42 and 2.72 years for females and males, respectively.

^f Non-sex related cancer survivors are survivors diagnosed with any cancers other than breast, cervical, endometrial, ovarian, or prostate cancers. Among non-sex related cancer survivors, SDs for PAC acceleration were: CaPAC5 = 2.48 and 2.73 years for females and males, respectively; and Lehallier's PAC = 2.42 and 2.83 years for females and males, respectively.

^g Among cancer-free participants (test set), SDs for PAC acceleration were: CaPAC5 = 2.56 and 2.53 years for females and males, respectively; and Lehallier's PAC = 2.47 and 2.52 years for females and males, respectively.

Supplemental Table 5.3 Associations of PAC acceleration with all-cause mortality among a subset of survivors of postmenopausal breast, prostate, and colorectal cancers with the information on stage of diagnosis

	No. of deaths	Total person-years	CaPAC5		Lehallier's PAC	
			HR (95% CI) per 1 SD ^a		HR (95% CI) per 1 SD	
			Model 3 ^b	Model 3 + stage	Model 3 ^b	Model 3 + stage
Postmenopausal breast cancer survivors	31	743	2.60 (1.41, 4.79)	2.68 (1.44, 4.99)	2.63 (1.43, 4.83)	2.63 (1.43, 4.85)
Prostate cancer survivors	80	1,400	1.19 (0.91, 1.56)	1.25 (0.95, 1.65)	1.27 (0.95, 1.71)	1.32 (0.97, 1.80)
CRC survivors	23	335	1.84 (0.95, 3.58)	1.49 (0.72, 3.06)	5.29 (1.70, 16.46)	3.97 (1.14, 13.86)

Abbreviations: PAC – proteomic ageing clock; CaPAC – cancer-specific proteomic aging clock; CVD – cardiovascular disease; SD – standard deviation; CRC – colorectal cancer; BMI – body mass index; eGFR – estimated glomerular filtration rate; HR – hazard ratio; CI – confidence interval.

^a SDs for PAC acceleration were: CaPAC5 = 2.38, 2.44, and 2.60 years for breast, prostate, and colorectal cancer survivors; and Lehallier's PAC = 2.29, 2.59, and 2.75 years for breast, prostate, and colorectal cancer survivors, respectively.

^b Model 3 was adjusted for chronological age, sex, race, study center, education, BMI, smoking status, alcohol intake, physical activity, time since cancer diagnosis, eGFR, history of CVD, aspirin use, and diabetes.

Supplemental Table 5.4 Top 40 pathways overrepresented in CaPAC5

No.	Pathway	FDR p-value	No.	Pathway	FDR p-value
1	regulation of multicellular organismal development	3.453E-19	21	cellular response to endogenous stimulus	3.829E-11
2	negative regulation of multicellular organismal process	1.64E-16	22	positive regulation of intracellular signal transduction	4.136E-11
3	cell adhesion	1.64E-16	23	neuron development	5.239E-11
4	positive regulation of multicellular organismal process	9.406E-16	24	positive regulation of MAPK cascade	6.583E-11
5	positive regulation of signal transduction	6.975E-15	25	neuron projection development	1.128E-10
6	locomotion	1.222E-14	26	tube development	1.847E-10
7	external encapsulating structure organization	1.222E-14	27	regulation of peptidase activity	2.86E-10
8	positive regulation of cell population proliferation	5.151E-14	28	enzyme-linked receptor protein signaling pathway	3.225E-10
9	extracellular matrix organization	1.566E-13	29	response to hormone	3.663E-10
10	extracellular structure organization	1.566E-13	30	regulation of protein phosphorylation	3.663E-10
11	chemotaxis	3.522E-13	31	regulation of MAPK cascade	4.102E-10
12	taxis	3.644E-13	32	positive regulation of cell differentiation	5.958E-10
13	inflammatory response	3.644E-13	33	regulation of nervous system development	6.908E-10
14	regulation of response to external stimulus	7.609E-13	34	blood vessel morphogenesis	8.724E-10
15	positive regulation of gene expression	1.055E-12	35	vasculature development	1.109E-09
16	cell-cell adhesion	1.247E-12	36	regulation of endopeptidase activity	1.217E-09
17	positive regulation of developmental process	2.246E-12	37	growth	1.282E-09
18	response to wounding	1.071E-11	38	defense response to other organism	1.625E-09
19	regulation of immune system process	1.287E-11	39	cell activation	1.646E-09
20	developmental growth	3.261E-11	40	cellular component morphogenesis	1.955E-09

Chapter 6 Summary

Recent evidence suggests that individual's aging process can be slowed down and the slowing down of aging can improve an individual's health over their lifespan. Given that the older U.S. population is expanding, there is a need to develop accurate measures of biological age. In turn, the accurate measures of biological age will enable the development of anti-aging interventions. The aims of this dissertation were to construct PACs using a large population-based study – ARIC, and test their associations with mortality, cancer risk and survival.

In the first manuscript, we used the proteomics data measured in middle-aged healthy participants (at Visit 2, aged 46-70 years) to construct and compare five new PACs with the goal to identify the best PAC among them. We utilized different types of penalized regressions and different transformations of proteins. All these five PACs were similarly and strongly correlated with chronological age and with each other; thus, we selected the simplest PAC – V2 ARIC PAC. This V2 ARIC PAC acceleration was significantly associated with all-cause mortality and CVD mortality (in both Cox proportional hazards model and competing risk model). V2 ARIC PAC was significantly associated with cancer mortality in Cox proportional hazards model, but not in the competing risk model.

We also constructed a de novo PAC (V5 ARIC PAC) using healthy participants of older age (at Visit 5, aged 66-90 years). Similar to V2 ARIC PAC, V5 ARIC PAC was also strongly correlated with chronological age and V5 ARIC PAC acceleration was significantly associated with all-cause and CVD mortality (in both Cox proportional hazards model and competing risk model). In addition, V5 ARIC PAC acceleration was

significantly associated with cancer mortality (in both Cox proportional hazards model and competing risk model).

Finally, we compared V2 ARIC PAC to three published PACs constructed at Visit 2. V2 ARIC PAC was correlated with three published PACs and the age acceleration for these four PACs was similarly associated with mortality.

In summary, PACs trained using different statistical methods and with different transformations of proteins show very similar performance and PACs that were developed in different populations were similarly associated with mortality. These findings suggest the robustness of PACs.

In the second manuscript, using proteomics data at Visit 2, we constructed a cancer-specific PAC (CaPAC) using participants who remained cancer-free until 2015. Similar to PAC constructed in healthy participants in the first manuscript, CaPAC was strongly correlated with chronological age and CaPAC acceleration was significantly associated with risk of overall, colorectal, and lung cancers. In addition, CaPAC was significantly associated with kidney cancer risk in Black participants but not White participants.

For the third manuscript, using the proteomics data measured in cancer-free participants at Visit 5, we constructed another cancer-specific PAC (CaPAC5). CaPAC5 was strongly correlated with chronological age in both cancer survivors and in cancer-free participants. CaPAC5 acceleration was significantly associated with all-cause mortality and cancer mortality (in both Cox proportional hazards model and competing risk model) in cancer survivors and all-cause mortality and CVD mortality (in both Cox proportional hazards model and competing risk model) in cancer-free participants. The

associations with all-cause mortality were similar for cancer survivors and cancer-free participants.

In addition, for manuscripts 2 and 3, we compared the newly constructed CAPACs (PACs created in White and Black cancer-free people) to the published PAC developed by Lehallier (Lehallier's PAC, created in White healthy people). We chose Lehallier's PAC rather than other published PACs, because it was most strongly correlated with chronological age among published PACs. The associations for cancer risk and all-cause mortality in cancer survivors and cancer-free participants were similar for Lehallier's PAC and the newly constructed CAPACs. Taken together, the findings from manuscripts 2 and 3 supported the finding from manuscript 1 about the robustness of PACs.

Overall, this dissertation extended our knowledge of PACs as measures of biological age to predict mortality, cancer risk and cancer survival. To our knowledge, our study is the first to examine associations with cancer risk and survival. If the association is causal, by slowing aging, it would be possible to slow down cancer development and progression.

Chapter 7 References

1. Sebastiani P, Thyagarajan B, Sun F, et al. Biomarker signatures of aging. *Aging Cell*. Apr 2017;16(2):329-338. doi:10.1111/accel.12557
2. Johnson AA, Shokhirev MN, Wyss-Coray T, Lehallier B. 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*. 07 2020;60:101070. doi:10.1016/j.arr.2020.101070
3. de Magalhães JP, Stevens M, Thornton D. The Business of Anti-Aging Science. *Trends Biotechnol*. 11 2017;35(11):1062-1073. doi:10.1016/j.tibtech.2017.07.004
4. Justice JN, Nambiar AM, Tchkonja T, et al. Senolytics in idiopathic pulmonary fibrosis: Results from a first-in-human, open-label, pilot study. *EBioMedicine*. Feb 2019;40:554-563. doi:10.1016/j.ebiom.2018.12.052
5. Xu M, Pirtskhalava T, Farr JN, et al. Senolytics improve physical function and increase lifespan in old age. *Nat Med*. 08 2018;24(8):1246-1256. doi:10.1038/s41591-018-0092-9
6. Guida JL, Agurs-Collins T, Ahles TA, et al. Strategies to Prevent or Remediate Cancer and Treatment-Related Aging. *J Natl Cancer Inst*. Apr 2020;doi:10.1093/jnci/djaa060
7. Schafer MJ, White TA, Evans G, et al. Exercise Prevents Diet-Induced Cellular Senescence in Adipose Tissue. *Diabetes*. 06 2016;65(6):1606-15. doi:10.2337/db15-0291
8. Soto-Perez-de-Celis E, Li D, Yuan Y, Lau YM, Hurria A. Functional versus chronological age: geriatric assessments to guide decision making in older patients with cancer. *Lancet Oncol*. 06 2018;19(6):e305-e316. doi:10.1016/S1470-2045(18)30348-6
9. Guida JL, Ahles TA, Belsky D, et al. Measuring Aging and Identifying Aging Phenotypes in Cancer Survivors. *J Natl Cancer Inst*. 12 2019;111(12):1245-1254. doi:10.1093/jnci/djz136
10. Hurria A, Jones L, Muss HB. Cancer Treatment as an Accelerated Aging Process: Assessment, Biomarkers, and Interventions. *American Society of Clinical Oncology educational book American Society of Clinical Oncology Annual Meeting*. 2016;35:e516-22. doi:10.1200/edbk_156160
11. Hoffman MC, Mulrooney DA, Steinberger J, Lee J, Baker KS, Ness KK. Deficits in physical function among young childhood cancer survivors. *J Clin Oncol*. Aug 2013;31(22):2799-805. doi:10.1200/JCO.2012.47.8081
12. Jylhävä J, Pedersen NL, Hägg S. Biological Age Predictors. *EBioMedicine*. Jul 2017;21:29-36. doi:10.1016/j.ebiom.2017.03.046
13. Tanaka T, Biancotto A, Moaddel R, et al. Plasma proteomic signature of age in healthy humans. *Aging Cell*. Oct 2018;17(5):e12799. doi:10.1111/accel.12799
14. Lehallier B, Shokhirev MN, Wyss-Coray T, Johnson AA. Data mining of human plasma proteins generates a multitude of highly predictive aging clocks that reflect different aspects of aging. *Aging Cell*. 11 2020;19(11):e13256. doi:10.1111/accel.13256
15. Sierra F, Hadley E, Suzman R, Hodes R. Prospects for life span extension. *Annu Rev Med*. 2009;60:457-69. doi:10.1146/annurev.med.60.061607.220533
16. Ferrucci L, Gonzalez-Freire M, Fabbri E, et al. Measuring biological aging in humans: A quest. *Aging Cell*. Feb 2020;19(2):e13080. doi:10.1111/accel.13080
17. Horvath S. DNA methylation age of human tissues and cell types. *Genome Biol*. 2013;14(10):R115. doi:10.1186/gb-2013-14-10-r115

18. Hannum G, Guinney J, Zhao L, et al. Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol Cell*. Jan 24 2013;49(2):359-367. doi:10.1016/j.molcel.2012.10.016
19. Levine ME, Lu AT, Quach A, et al. An epigenetic biomarker of aging for lifespan and healthspan. *Aging (Albany NY)*. 04 2018;10(4):573-591. doi:10.18632/aging.101414
20. Lu AT, Quach A, Wilson JG, et al. DNA methylation GrimAge strongly predicts lifespan and healthspan. *Aging (Albany NY)*. 01 2019;11(2):303-327. doi:10.18632/aging.101684
21. Belsky DW, Caspi A, Corcoran DL, et al. DunedinPACE, a DNA methylation biomarker of the pace of aging. *Elife*. Jan 14 2022;11doi:10.7554/eLife.73420
22. Belsky DW, Caspi A, Arseneault L, et al. Quantification of the pace of biological aging in humans through a blood test, the DunedinPoAm DNA methylation algorithm. *Elife*. 05 2020;9doi:10.7554/eLife.54870
23. Weidner CI, Lin Q, Koch CM, et al. Aging of blood can be tracked by DNA methylation changes at just three CpG sites. *Genome Biol*. Feb 2014;15(2):R24. doi:10.1186/gb-2014-15-2-r24
24. Zhang Y, Wilson R, Heiss J, et al. DNA methylation signatures in peripheral blood strongly predict all-cause mortality. *Nat Commun*. Mar 17 2017;8:14617. doi:10.1038/ncomms14617
25. Levine ME, Hosgood HD, Chen B, Absher D, Assimes T, Horvath S. DNA methylation age of blood predicts future onset of lung cancer in the women's health initiative. *Aging (Albany NY)*. Sep 2015;7(9):690-700. doi:10.18632/aging.100809
26. Yang Z, Wong A, Kuh D, et al. Correlation of an epigenetic mitotic clock with cancer risk. *Genome Biol*. Oct 03 2016;17(1):205. doi:10.1186/s13059-016-1064-3
27. Bocklandt S, Lin W, Sehl ME, et al. Epigenetic predictor of age. *PLoS One*. 2011;6(6):e14821. doi:10.1371/journal.pone.0014821
28. Garagnani P, Bacalini MG, Pirazzini C, et al. Methylation of ELOVL2 gene as a new epigenetic marker of age. *Aging Cell*. Dec 2012;11(6):1132-4. doi:10.1111/acel.12005
29. Tanaka T, Basisty N, Fantoni G, et al. Plasma proteomic biomarker signature of age predicts health and life span. *Elife*. 11 2020;9doi:10.7554/eLife.61073
30. Roetker NS, Pankow JS, Bressler J, Morrison AC, Boerwinkle E. Prospective Study of Epigenetic Age Acceleration and Incidence of Cardiovascular Disease Outcomes in the ARIC Study (Atherosclerosis Risk in Communities). *Circ Genom Precis Med*. 03 2018;11(3):e001937. doi:10.1161/CIRCGEN.117.001937
31. Chen BH, Marioni RE, Colicino E, et al. DNA methylation-based measures of biological age: meta-analysis predicting time to death. *Aging (Albany NY)*. 09 2016;8(9):1844-1865. doi:10.18632/aging.101020
32. Morales Bernstein F, McCartney DL, Lu AT, et al. Assessing the causal role of epigenetic clocks in the development of multiple cancers: a Mendelian randomization study. *Elife*. 03 29 2022;11doi:10.7554/eLife.75374
33. Dugué PA, Bassett JK, Wong EM, et al. Biological Aging Measures Based on Blood DNA Methylation and Risk of Cancer: A Prospective Study. *JNCI Cancer Spectr*. 02 2021;5(1)doi:10.1093/jncics/pkaa109

34. Chung M, Ruan M, Zhao N, et al. DNA methylation aging clocks and pancreatic cancer risk: Pooled analysis of three prospective nested case-control studies. *medRxiv*. 2020;
35. Kresovich JK, Xu Z, O'Brien KM, Weinberg CR, Sandler DP, Taylor JA. Methylation-Based Biological Age and Breast Cancer Risk. *J Natl Cancer Inst*. 10 2019;111(10):1051-1058. doi:10.1093/jnci/djz020
36. Dugué PA, Bassett JK, Joo JE, et al. DNA methylation-based biological aging and cancer risk and survival: Pooled analysis of seven prospective studies. *Int J Cancer*. Apr 15 2018;142(8):1611-1619. doi:10.1002/ijc.31189
37. Levine ME, Crimmins EM. Is 60 the New 50? Examining Changes in Biological Age Over the Past Two Decades. *Demography*. 04 2018;55(2):387-402. doi:10.1007/s13524-017-0644-5
38. Ambatipudi S, Horvath S, Perrier F, et al. DNA methylome analysis identifies accelerated epigenetic ageing associated with postmenopausal breast cancer susceptibility. *Eur J Cancer*. 04 2017;75:299-307. doi:10.1016/j.ejca.2017.01.014
39. Zheng C, Li L, Xu R. Association of Epigenetic Clock with Consensus Molecular Subtypes and Overall Survival of Colorectal Cancer. *Cancer Epidemiol Biomarkers Prev*. Oct 2019;28(10):1720-1724. doi:10.1158/1055-9965.Epi-19-0208
40. Zaghlool SB, Kühnel B, Elhadad MA, et al. Epigenetics meets proteomics in an epigenome-wide association study with circulating blood plasma protein traits. *Nat Commun*. Jan 3 2020;11(1):15. doi:10.1038/s41467-019-13831-w
41. Santos R, Ursu O, Gaulton A, et al. A comprehensive map of molecular drug targets. *Nature reviews Drug discovery*. 2017;16(1):19-34.
42. Sun BB, Maranville JC, Peters JE, et al. Genomic atlas of the human plasma proteome. *Nature*. 06 2018;558(7708):73-79. doi:10.1038/s41586-018-0175-2
43. Sathyan S, Ayers E, Gao T, et al. Plasma proteomic profile of age, health span, and all-cause mortality in older adults. *Aging Cell*. Nov 2020;19(11):e13250. doi:10.1111/acel.13250
44. Lehallier B, Gate D, Schaum N, et al. Undulating changes in human plasma proteome profiles across the lifespan. *Nat Med*. 12 2019;25(12):1843-1850. doi:10.1038/s41591-019-0673-2
45. Organization WH. Cancer prevention. <https://www.who.int/cancer/prevention/en/#:~:text=Between%2030%2D50%25%20of%20all,for%20the%20control%20of%20cancer.>
46. Partridge L, Fuentealba M, Kennedy BK. The quest to slow ageing through drug discovery. *Nat Rev Drug Discov*. Aug 2020;19(8):513-532. doi:10.1038/s41573-020-0067-7
47. Campisi J. Aging, cellular senescence, and cancer. *Annual review of physiology*. 2013;75:685-705.
48. Il'yasova D, Colbert LH, Harris TB, et al. Circulating levels of inflammatory markers and cancer risk in the health aging and body composition cohort. *Cancer Epidemiology and Prevention Biomarkers*. 2005;14(10):2413-2418.
49. Zuo L, Prather ER, Stetskiv M, et al. Inflammaging and oxidative stress in human diseases: From molecular mechanisms to novel treatments. *International journal of molecular sciences*. 2019;20(18):4472.

50. Greten FR, Grivennikov SI. Inflammation and cancer: triggers, mechanisms, and consequences. *Immunity*. 2019;51(1):27-41.
51. Institute NC. statistics, Graphs, and Definitions. Accessed Jan 4, 2021. <https://cancercontrol.cancer.gov/ocs/statistics#stats>
52. Wright JD, Folsom AR, Coresh J, et al. The ARIC (Atherosclerosis Risk In Communities) Study: JACC Focus Seminar 3/8. *J Am Coll Cardiol*. Jun 2021;77(23):2939-2959. doi:10.1016/j.jacc.2021.04.035
53. Joshi CE, Barber JR, Coresh J, et al. Enhancing the Infrastructure of the Atherosclerosis Risk in Communities (ARIC) Study for Cancer Epidemiology Research: ARIC Cancer. *Cancer Epidemiol Biomarkers Prev*. 03 2018;27(3):295-305. doi:10.1158/1055-9965.EPI-17-0696
54. Gold L, Ayers D, Bertino J, et al. Aptamer-based multiplexed proteomic technology for biomarker discovery. *PLoS One*. Dec 7 2010;5(12):e15004. doi:10.1371/journal.pone.0015004
55. Gold L, Walker JJ, Wilcox SK, Williams S. Advances in human proteomics at high scale with the SOMAscan proteomics platform. Review. *New biotechnology*. Jun 15 2012;29(5):543-9. doi:10.1016/j.nbt.2011.11.016
56. Kim CH, Tworoger SS, Stampfer MJ, et al. Stability and reproducibility of proteomic profiles measured with an aptamer-based platform. *Sci Rep*. May 30 2018;8(1):8382. doi:10.1038/s41598-018-26640-w
57. Candia J, Cheung F, Kotliarov Y, et al. Assessment of Variability in the SOMAscan Assay. *Sci Rep*. Oct 27 2017;7(1):14248. doi:10.1038/s41598-017-14755-5
58. Tin A, Yu B, Ma J, et al. Reproducibility and Variability of Protein Analytes Measured Using a Multiplexed Modified Aptamer Assay. *The journal of applied laboratory medicine*. Jul 2019;4(1):30-39. doi:10.1373/jalm.2018.027086
59. SomaLogic. Somascan® data standardization. <https://somallogic.com/wp-content/uploads/2017/06/SSM-071-Rev-0-Technical-Note-SOMAscan-Data-Standardization.pdf>
60. Rooney MR, Tang O, Pankow JS, Selvin E. Glycaemic markers and all-cause mortality in older adults with and without diabetes: the Atherosclerosis Risk in Communities (ARIC) study. *Diabetologia*. 02 2021;64(2):339-348. doi:10.1007/s00125-020-05285-3
61. Horvath S, Raj K. DNA methylation-based biomarkers and the epigenetic clock theory of ageing. *Nat Rev Genet*. Jun 2018;19(6):371-384. doi:10.1038/s41576-018-0004-3
62. Baker GT, Sprott RL. Biomarkers of aging. *Exp Gerontol*. 1988;23(4-5):223-39. doi:10.1016/0531-5565(88)90025-3
63. Galkin F, Mamoshina P, Aliper A, de Magalhães JP, Gladyshev VN, Zhavoronkov A. Biohorology and biomarkers of aging: Current state-of-the-art, challenges and opportunities. *Ageing Res Rev*. 07 2020;60:101050. doi:10.1016/j.arr.2020.101050
64. Tworoger SS, Hankinson SE. Collection, processing, and storage of biological samples in epidemiologic studies: sex hormones, carotenoids, inflammatory markers, and proteomics as examples. *Cancer Epidemiol Biomarkers Prev*. Sep 2006;15(9):1578-81. doi:10.1158/1055-9965.epi-06-0629

65. Rai AJ, Gelfand CA, Haywood BC, et al. HUPO Plasma Proteome Project specimen collection and handling: towards the standardization of parameters for plasma proteome samples. *Proteomics*. Aug 2005;5(13):3262-77. doi:10.1002/pmic.200401245
66. Zou H, Hastie T. Regularization and Variable Selection via the Elastic Net. *Journal of the Royal Statistical Society Series B (Statistical Methodology)*. 2022/12/28/2005;67(2):301-320.
67. Mozaffarian D, Benjamin EJ, Go AS, et al. Heart disease and stroke statistics--2015 update: a report from the American Heart Association. *Circulation*. Jan 27 2015;131(4):e29-322. doi:10.1161/CIR.0000000000000152
68. Chen J, Xu H, Aronow BJ, Jegga AG. Improved human disease candidate gene prioritization using mouse phenotype. *BMC Bioinformatics*. Oct 16 2007;8:392. doi:10.1186/1471-2105-8-392
69. Chen J, Aronow BJ, Jegga AG. Disease candidate gene identification and prioritization using protein interaction networks. *BMC Bioinformatics*. Feb 27 2009;10:73. doi:10.1186/1471-2105-10-73
70. Chen J, Bardes EE, Aronow BJ, Jegga AG. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Res*. Jul 2009;37(Web Server issue):W305-11. doi:10.1093/nar/gkp427
71. Trent SC, Peterson P, Baker E, McGaw B. Overrepresentation of Culturally and Linguistically Diverse Students in Special Education. *International Encyclopedia of Education (Third Edition)*. Elsevier; 2010:774-779.
72. Fine JP, Gray RJ. A Proportional Hazards Model for the Subdistribution of a Competing Risk. *Journal of the American Statistical Association*. 1999;94(446):496-509. doi:10.1080/01621459.1999.10474144
73. Austin PC, Fine JP. Practical recommendations for reporting Fine-Gray model analyses for competing risk data. *Stat Med*. Nov 30 2017;36(27):4391-4400. doi:10.1002/sim.7501
74. Scott AJ, Ellison M, Sinclair DA. The economic value of targeting aging. *Nature Aging*. 2021;1(7):616-623. doi:10.1038/s43587-021-00080-0
75. Kankaanpää A, Tolvanen A, Saikkonen P, et al. Do Epigenetic Clocks Provide Explanations for Sex Differences in Life Span? A Cross-Sectional Twin Study. *J Gerontol A Biol Sci Med Sci*. 09 01 2022;77(9):1898-1906. doi:10.1093/gerona/glab337
76. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. Research Support, U.S. Gov't, P.H.S. *American journal of epidemiology*. Apr 1989;129(4):687-702.
77. SomaLogic. <https://somalogic.com/>
78. Lu J, Lu Y, Krumholz HM, Jiang L. Prevalence and control of hypertension - Authors' reply. *Lancet*. 10 13 2018;392(10155):1306. doi:10.1016/S0140-6736(18)31895-6
79. Inker LA, Eneanya ND, Coresh J, et al. New Creatinine- and Cystatin C-Based Equations to Estimate GFR without Race. *N Engl J Med*. 11 04 2021;385(19):1737-1749. doi:10.1056/NEJMoa2102953
80. ATS statement--Snowbird workshop on standardization of spirometry. *Am Rev Respir Dis*. May 1979;119(5):831-8. doi:10.1164/arrd.1979.119.5.831

81. Crapo RO, Morris AH, Gardner RM. Reference spirometric values using techniques and equipment that meet ATS recommendations. *Am Rev Respir Dis*. Jun 1981;123(6):659-64. doi:10.1164/arrd.1981.123.6.659
82. Harik-Khan RI, Fleg JL, Muller DC, Wise RA. The effect of anthropometric and socioeconomic factors on the racial difference in lung function. *Am J Respir Crit Care Med*. Nov 01 2001;164(9):1647-54. doi:10.1164/ajrccm.164.9.2106075
83. Lung function testing: selection of reference values and interpretative strategies. American Thoracic Society. *Am Rev Respir Dis*. Nov 1991;144(5):1202-18. doi:10.1164/ajrccm/144.5.1202
84. Glindmeyer HW, Lefante JJ, McColloster C, Jones RN, Weill H. Blue-collar normative spirometric values for Caucasian and African-American men and women aged 18 to 65. *Am J Respir Crit Care Med*. Feb 1995;151(2 Pt 1):412-22. doi:10.1164/ajrccm.151.2.7842200
85. Bell EJ, Lutsey PL, Windham BG, Folsom AR. Physical activity and cardiovascular disease in African Americans in Atherosclerosis Risk in Communities. *Med Sci Sports Exerc*. May 2013;45(5):901-7. doi:10.1249/MSS.0b013e31827d87ec
86. Folsom AR, Yatsuya H, Nettleton JA, et al. Community prevalence of ideal cardiovascular health, by the American Heart Association definition, and relationship with cardiovascular disease incidence. *J Am Coll Cardiol*. Apr 19 2011;57(16):1690-6. doi:10.1016/j.jacc.2010.11.041
87. Eriksson H, Caidahl K, Larsson B, et al. Cardiac and pulmonary causes of dyspnoea--validation of a scoring test for clinical-epidemiological use: the Study of Men Born in 1913. *Eur Heart J*. Sep 1987;8(9):1007-14. doi:10.1093/oxfordjournals.eurheartj.a062365
88. Loehr LR, Rosamond WD, Chang PP, Folsom AR, Chambless LE. Heart failure incidence and survival (from the Atherosclerosis Risk in Communities study). *Am J Cardiol*. Apr 01 2008;101(7):1016-22. doi:10.1016/j.amjcard.2007.11.061
89. Rosamond WD, Chang PP, Baggett C, et al. Classification of heart failure in the atherosclerosis risk in communities (ARIC) study: a comparison of diagnostic criteria. *Circ Heart Fail*. Mar 01 2012;5(2):152-9. doi:10.1161/CIRCHEARTFAILURE.111.963199
90. The National Survey of Stroke. National Institute of Neurological and Communicative Disorders and Stroke. *Stroke*. 1981 Mar-Apr 1981;12(2 Pt 2 Suppl 1):I1-91.
91. Rosamond WD, Folsom AR, Chambless LE, et al. Stroke incidence and survival among middle-aged adults: 9-year follow-up of the Atherosclerosis Risk in Communities (ARIC) cohort. *Stroke*. Apr 1999;30(4):736-43. doi:10.1161/01.str.30.4.736
92. Polter EJ, Onyeaghala G, Lutsey PL, et al. Prospective Association of Serum and Dietary Magnesium with Colorectal Cancer Incidence. *Cancer Epidemiol Biomarkers Prev*. Aug 2019;28(8):1292-1299. doi:10.1158/1055-9965.epi-18-1300
93. Chen LY, Norby FL, Gottesman RF, et al. Association of Atrial Fibrillation With Cognitive Decline and Dementia Over 20 Years: The ARIC-NCS (Atherosclerosis Risk in Communities Neurocognitive Study). *J Am Heart Assoc*. 03 07 2018;7(6)doi:10.1161/JAHA.117.007301

94. Knopman DS, Ryberg S. A verbal memory test with high predictive accuracy for dementia of the Alzheimer type. *Arch Neurol*. Feb 1989;46(2):141-5. doi:10.1001/archneur.1989.00520380041011
95. Wechsler D. *The Adult Intelligence Scale—Revised*. New York, NY: Psychological Corporation; 1981.
96. M. L. *Neuropsychological Assessment*. Oxford, UK: Oxford University Press; 1995.
97. White MC, Holman DM, Boehm JE, Peipins LA, Grossman M, Henley SJ. Age and cancer risk: a potentially modifiable relationship. *Am J Prev Med*. Mar 2014;46(3 Suppl 1):S7-15. doi:10.1016/j.amepre.2013.10.029
98. Institute NC. Age and Cancer Risk. <https://www.cancer.gov/about-cancer/causes-prevention/risk/age>
99. Cupit-Link MC, Kirkland JL, Ness KK, et al. Biology of premature ageing in survivors of cancer. *ESMO open*. 2017;2(5):e000250. doi:10.1136/esmoopen-2017-000250
100. Zheng Y, Joyce BT, Colicino E, et al. Blood Epigenetic Age may Predict Cancer Incidence and Mortality. *EBioMedicine*. Mar 2016;5:68-73. doi:10.1016/j.ebiom.2016.02.008
101. Durso DF, Bacalini MG, Sala C, et al. Acceleration of leukocytes' epigenetic age as an early tumor and sex-specific marker of breast and colorectal cancer. *Oncotarget*. Apr 2017;8(14):23237-23245. doi:10.18632/oncotarget.15573
102. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2022. *CA Cancer J Clin*. 01 2022;72(1):7-33. doi:10.3322/caac.21708
103. Wang T, Maden SK, Luebeck GE, et al. Dysfunctional epigenetic aging of the normal colon and colorectal cancer risk. *Clin Epigenetics*. Jan 3 2020;12(1):5. doi:10.1186/s13148-019-0801-3
104. SomaScan. <https://somalogic.com/somascan-platform/>
105. Suh E, Stratton KL, Leisenring WM, et al. Late mortality and chronic health conditions in long-term survivors of early-adolescent and young adult cancers: a retrospective cohort analysis from the Childhood Cancer Survivor Study. *Lancet Oncol*. 03 2020;21(3):421-435. doi:10.1016/S1470-2045(19)30800-9
106. Wang S, Prizment A, Thyagarajan B, Blaes A. Cancer Treatment-Induced Accelerated Aging in Cancer Survivors: Biology and Assessment. *Cancers (Basel)*. Jan 23 2021;13(3)doi:10.3390/cancers13030427
107. López-Otín C, Blasco MA, Partridge L, Serrano M, Kroemer G. The hallmarks of aging. *Cell*. Jun 2013;153(6):1194-217. doi:10.1016/j.cell.2013.05.039
108. Tanaka T, Biancotto A, Moaddel R, et al. Plasma proteomic signature of age in healthy humans. *Aging Cell*. Oct 2018;17(5):e12799. doi:10.1111/accel.12799
109. Nussinov R, Tsai CJ, Jang H. Protein ensembles link genotype to phenotype. *PLoS Comput Biol*. 06 2019;15(6):e1006648. doi:10.1371/journal.pcbi.1006648
110. Rutledge J, Oh H, Wyss-Coray T. Measuring biological age using omics data. *Nat Rev Genet*. Dec 2022;23(12):715-727. doi:10.1038/s41576-022-00511-7
111. Williams SA, Kivimaki M, Langenberg C, et al. Plasma protein patterns as comprehensive indicators of health. *Nat Med*. Dec 2019;25(12):1851-1857. doi:10.1038/s41591-019-0665-2

112. Benjamins MR, Silva A, Saiyed NS, De Maio FG. Comparison of All-Cause Mortality Rates and Inequities Between Black and White Populations Across the 30 Most Populous US Cities. *JAMA Netw Open*. 01 04 2021;4(1):e2032086. doi:10.1001/jamanetworkopen.2020.32086
113. Petrick JL, Foraker RE, Kucharska-Newton AM, et al. Trajectory of overall health from self-report and factors contributing to health declines among cancer survivors. *Cancer Causes Control*. Sep 2014;25(9):1179-86. doi:10.1007/s10552-014-0421-3
114. Fried LP, Tangen CM, Walston J, et al. Frailty in older adults: evidence for a phenotype. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*. 2001;56(3):M146-M157.
115. Guida JL, Ahles TA, Belsky D, et al. Measuring aging and identifying aging phenotypes in cancer survivors. *JNCI: Journal of the National Cancer Institute*. 2019;111(12):1245-1254.