From Department of Medical Epidemiology and Biostatistics
Karolinska Institutet, Stockholm, Sweden

BIOLOGICAL AGES: CORRELATIONS, GENETIC DETERMINANTS, AND HEALTH OUTCOMES

Xia Li

李夏

Stockholm 2021

Published by Karolinska Institutet.
Printed by Universitetsservice US-AB, 2021

ISBN 978-91-8016-373-6

Biological ages: Correlations, genetic determinants, and health outcomes

THESIS FOR DOCTORAL DEGREE (Ph.D.)

By

Xia Li

November 11th, 2021, 16:00
Lecture hall Petrén, Nobels väg 12B, Karolinska Institutet, Solna

Principal Supervisor:
Dr. Sara Hägg
Karolinska Institutet
Department of Medical Epidemiology and Biostatistics

Co-supervisor(s):
Dr. Juulia Jylhävä
Karolinska Institutet
Department of Medical Epidemiology and Biostatistics

Dr. Patrik Magnusson
Karolinska Institutet
Department of Medical Epidemiology and Biostatistics

Dr. Alexander Ploner
Karolinska Institutet
Department of Medical Epidemiology and Biostatistics

Opponent:
Dr. Brandon Pierce
University of Chicago
Department of Public Health Sciences

Examination Board:
Dr. Amaia Calderón-Larrañaga
Karolinska Institutet
Department of Neurobiology, Care Sciences and Society, Aging Research Center

Dr. Serhiy Dekhtyar
Karolinska Institutet
Department of Neurobiology, Care Sciences and Society, Aging Research Center

Dr. Anna Zettergren
University of Gothenburg
Department of Psychiatry and Neurochemistry

To everyone
— who shall experience aging in one way or another

ABSTRACT

Population aging is a global trend and requires better evidence-based guidance. This thesis studied population aging from a molecular epidemiological angle. The overall work is centered on the "biological age (BA)", which, in a broad sense, is a quantification of any aging-related changes in the cellular, organ-, system-, and/or organismal features. Ideally, BA provides additional information in the assessment and prediction of aging risks independent of chronological age.

The first and second studies explored the correlations between BAs and mortality associations.

Study 1 examined a frailty measure, the frailty index (FI), with all-cause and cause-specific mortality in 42,953 twins. Increased FI was associated with higher risks of death due to all-cause, cardiovascular diseases, and respiratory-related causes. Particularly, the effect was independent of familiar factors and declined with growing age.

Study 2 focused on a list of established BAs, including telomere length (TL), DNA methylation-based age estimators (DNAmAges), a multiple biomarker-derived BA score, and functional BA measures in 846 adults. Correlations were generally stronger between BAs of the same type, with TL showing the weakest correlations to other BAs, and the remaining demonstrating moderate to high correlations across BAs. Individually, all BAs except for TL were associated with mortality risk; jointly, two DNAmAges and the FI were predictive of mortality risk independent of the other BAs.

The third and fourth studies incorporated genetic information to disentangle relationships between genetic factors, clinical biomarkers, and aging phenotypes.

Study 3 investigated a set of clinical biomarkers in relation to healthspan, i.e., disease-free lifespan, and used genetically predicted biomarkers as instrumental variables in 12,098 participants. Glycemic, lipid-, and inflammatory biomarkers were associated with altered risks of healthspan. In addition, genetic predisposition to elevated fasting blood glucose was associated with a higher risk of encountering an end of healthspan during the follow-up.

Study 4 interrogated rare and functional genetic determinants of C-reactive protein (CRP) and the clinical relevance of the discovered genetic mutations in 161,430 adults. Carrying a protein-altering or loss-of-function mutation in the CRP gene was significantly associated with decreased serum CRP concentration. Mutation carriers were less affected by obese status in terms of the magnitude of increased CRP level.

In conclusion, BAs can capture distinct aspects of aging-related information. Making use of a set of multi-dimensional BAs could provide complementary evidence for risk assessment and intervention/treatment effect evaluation in research as well as in clinical practices. Measurements of BA together with genetic assessments would facilitate the delivery of precision medicine.

LIST OF SCIENTIFIC PAPERS

I. **Li X**, Ploner A, Karlsson IK, Liu X, Magnusson PK, Pedersen NL, Hägg S, Jylhävä J. The frailty index is a predictor of cause-specific mortality independent of familial effects from midlife onwards: a large cohort study. *BMC Medicine*. 2019 May 15;17(1):94.

II. **Li X**, Ploner A, Wang Y, Magnusson PK, Reynolds C, Finkel D, Pedersen NL, Jylhävä J, Hägg S. Longitudinal trajectories, correlations and mortality associations of nine biological ages across 20-years follow-up. *eLife*. 2020 Feb 11;9:e51507.

III. **Li X**, Ploner A, Wang Y, Zhan Y, Pedersen NL, Magnusson PK, Jylhävä J, Hägg S. Clinical biomarkers and associations with healthspan and lifespan: evidence from observational and genetic data. *EBioMedicine*. 2021 Apr 1;66:103318.

IV. **Li X**, Ploner A, Wang Y, Mak J, Lu Y, Magnusson PK, Jylhävä J, Hägg S. Associations of exome-wide rare genetic variants with serum C-reactive protein. (*Manuscript*)

CONTENTS

1 INTRODUCTION....1
2 BACKGROUND....3
2.1 Biological aging....3
2.2 Biological age....4
2.3 Correlations of BA....10
2.4 Mortality associations of BA....11
2.5 Clinical biomarkers and healthspan....12
2.6 Genetic variants of CRP....13
3 RESEARCH AIMS....15
4 MATERIALS AND METHODS....17
4.1 Study population....17
4.1.1 Screening Across the Lifespan Twin Study (SALT)....17
4.1.2 Swedish Adoption/Twin Study of Aging (SATSA)....17
4.1.3 TwinGene....18
4.1.4 UKB....18
4.2 Measurements....19
4.2.1 Biological ages....19
4.2.2 Clinical biomarkers....22
4.2.3 Genetic variants....22
4.2.4 Health outcomes....23
4.3 Statistical analysis....24
4.3.1 Generalized survival models....24
4.3.2 Correlations and mortality associations of multiple BAs....25
4.3.3 Cox regression models of clinical biomarkers and PRSs....25
4.3.4 Gene-based burden test....26
5 RESULTS....27
5.1 Study 1....27
5.1.1 Population characteristics....27
5.1.2 Time-constant FI-mortality associations....28
5.1.3 Time-dependent FI-mortality associations....28
5.2 Study 2....29
5.2.1 Population characteristics....29
5.2.2 BA correlations and clustering....30
5.2.3 BA-mortality associations....32
5.3 Study 3....32
5.3.1 Population characteristics....32
5.3.2 Serum biomarkers and healthspan....33
5.3.3 Biomarker PRSs and healthspan....34
5.4 Study 4....35
5.4.1 Population characteristics....35
5.4.2 Rare functional genetic variants and CRP....35

5.4.3 Gene-BMI interaction effect on CRP 37
5.4.4 CRP-associated rare mutations in relation to diseases and aging phenotypes 38
6 DISCUSSION 39
6.1 BA correlations and BA-mortality associations 39
6.2 Clinical biomarkers and healthspan 41
6.3 Rare functional genetic variants and serum CRP 41
6.4 Methodological considerations 42
6.5 Ethical considerations 44
7 CONCLUSIONS 45
8 FUTURE PERSPECTIVE 47
9 ACKNOWLEDGEMENTS 49
10 REFERENCES 51

LIST OF ABBREVIATIONS

ApoA1	Apolipoprotein A1
ApoB	Apolipoprotein B
BA	Biological age
BMI	Body mass index
CA	Chronological age
CAD	Coronary artery disease
CDS	Coding sequence
CHF	Coronary heart failure
COPD	Chronic obstructive pulmonary disease
CRP	C-reactive protein
CVD	Cardiovascular disease
DNAmAges	DNA methylation-based age estimators
DZ	Dizygotic
FAI	Functional aging index
FBG	Fasting blood glucose
FDR	False discovery rate
FI	Frailty index
FP	Frailty phenotype
G*E	Gene-environment
GSM	Generalized survival model
GWAS	Genome-wide association study
HALE	Health-adjusted life expectancy
Hb	Hemoglobin
HbA1c	Glycated hemoglobin
HDL-C	High-density lipoprotein cholesterol
HR	Hazard ratio
IBD	Inflammatory bowel disease
IPT	In-person testing
IQR	Interquartile range
KDM	The BA method proposed by Klemera and Doubal
LD	Linkage disequilibrium
LDL-C	Low-density lipoprotein cholesterol

MAF	Minor allele frequency
MI	Myocardial infarction
Multi-biomarker BA	Multiple biomarker-derived BA score
MZ	Monozygotic
OR	Odds ratio
PC	Principal component
PCA	Principal component analysis
PolyPhen	Polymorphism phenotyping
PRS	Polygenetic risk score
QC	Quality control
qPCR	Quantitative polymerase chain reaction
SALT	Screening across the lifespan twin study
SATSA	Swedish Adoption/Twin Study of Aging
SD	Standard deviation
SIFT	Sorting intolerant from tolerant
SNV	Single nucleotide variant
STR	Swedish Twin Registry
TC	Total cholesterol
TG	Triglyceride
TIA	Transient ischemic attack
TL	Telomere length
UKB	UK Biobank
WES	Whole-exome sequencing
WHO	World Health Organization

1 INTRODUCTION

Population aging is a global trend. According to the World Population Prospects 2019 reported by the World Health Organization (WHO), the average life expectancy at birth has been growing worldwide, rising from 64.2 years in 1990 up to 72.6 years in 2019 and a projected number of 77.1 years in 2050. (1)

This longevity achievement is reached unevenly across genders and geographical regions, with women (75.0 versus 70.2 years in men) and Australia/New Zealand region (83.2 versus 65.2 years in the least developed countries) showing a longer life expectancy in 2019. Moreover, population age structures accelerate the burden of population aging at varying speeds. On a global scale, individuals aged 65 years or over are estimated to make up 9.1% and 15.9% of the total population in 2019 and 2050, respectively. Region-wise, the Sub-Saharan Africa region shows the lowest 65+ proportion (3.0% in 2019 and 4.5% in 2050) while the Europe and Northern America region presents the highest (18.0% in 2019 and 26.1% in 2050). (1)

A growing number of the older population would lead to a remarkable change in social, economic, and medical systems that requires better evidence-based guidance. In the meantime, more and more people reaching advanced age and the advent of the big-data era make large-scale data of human aging available and accessible. Together, the situation provides researchers valuable opportunities to study the heterogeneity of aging-related phenotypes and to elucidate the reasons underneath the phenomenon that some people could experience aging "successfully" while others could not.

Towards these ends, this thesis studied population aging from a molecular epidemiological angle. The overall work is centered on the concept of "biological age (BA)", which, in a broad sense, is a quantification of any aging-related changes in the cellular, organ-, system-, and/or organismal features. Ideally, BA provides additional information in the assessment and prediction of aging risks independent of chronological age (CA).

Of all BA candidates, the present thesis focused on several established ones, namely telomere length (TL), DNA methylation-based age estimators (DNAmAges), several clinical biomarkers, a multiple biomarker-derived BA score (multi-biomarker BA), physical and cognitive function, as well as the frailty index (FI). Specifically, research interests went to 1) the correlations between different types of BA, 2) the associations of one and a set of BAs with mortality, 3) a set of clinical biomarkers in association with the end of healthspan, and 4) the genetic determinants of one inflammatory biomarker, C-reactive protein (CRP), and their clinical relevance in relation to health outcomes.

2 BACKGROUND

2.1 BIOLOGICAL AGING

Aging, the age-related senescence, is ubiquitous in nature, although some organisms, such as hydra, (2) do manifest biological immortality. Evolutionary theories, including antagonistic pleiotropy hypothesis and mutation accumulation theory, proposed that aging results from the accumulation of detrimental mutation because the force of natural selection against late-acting alleles is not as efficient as against early-onset mutation, especially during the post-reproductive period. (3-7) In empirical observations, human biological aging is often viewed as multifaceted changes occurring with increasing age, manifested as physiological deteriorations, functional impairments, and increased susceptibilities to adverse outcomes. (8-12) Depending on the biological scales and domains being inspected, researchers depicted aging differently (Figure 2.1.1).

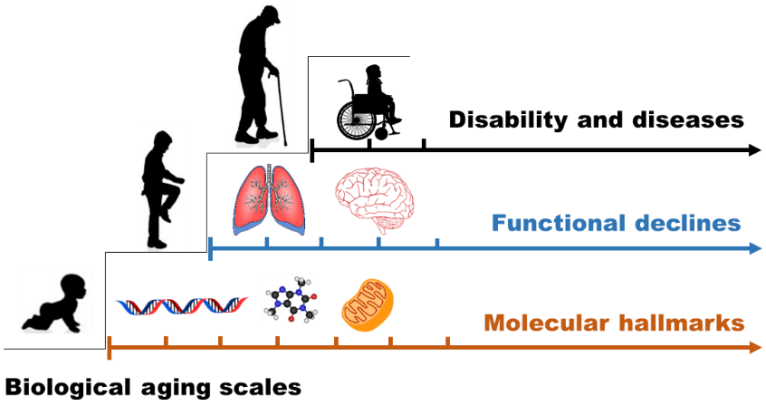

Figure 2.1.1 Biological aging scales. Biological aging changes can be projected to different scales. Common quantifications include molecular hallmarks, functional declines, and vulnerability to disability and diseases, of which molecular changes could occur throughout lifespan while organ- and organismal dysfunction are often observable from the middle age onwards.

Molecular and cellular hallmarks

A network of biological mechanisms underlies the aging process. In 2013, Lopez-Otın et al. put forward an integrative scheme consisting of nine interrelated molecular and cellular hallmarks of aging. Broadly, the hallmarks fall into three categories, 1) primary causes of cellular damage (genomic instability, telomere attrition, epigenetic alterations, and loss of proteostasis), 2) responses to damage (deregulated nutrient sensing, mitochondrial dysfunction, and cellular senescence), and 3) integrative phenotypes (stem cell exhaustion, and altered intercellular communication). (13) Later on, other reviews continued to propose similar molecular and cellular frameworks, such as seven pillars of aging and cellular aging defects, to summarize the mechanisms shared by aging and age-related diseases. (14-16) Unlike functional declines of organ systems which are often observable in adulthood, molecular and cellular changes occur and accumulate throughout the lifespan. (13)

Functional declines

Aging-related physiological changes accumulate with age and affect organ-, system-, and organismal functions. Khan et al. summarized the age-related characteristics of a range of organ systems, including neurological, pulmonary, cardiovascular, renal, gastrointestinal, endocrine, reproductive, and musculoskeletal systems, and observed that relative rates of functional decline vary greatly. (8) The most rapid decline is seen for female reproductive function, with female fertility peaking in the early twenties and exhausted at the age of menopause. (17, 18) On the contrary, gastrointestinal function only undergoes mild deterioration from middle-age and onwards and liver chemistries were mostly age-independent. (8, 19) At the organismal level, function declines, like physical dysfunction and cognitive aging, are often observable across multiple organs and systems at advanced ages. (20-22)

Increased susceptibility to diseases and death

As age is a major risk factor for most chronic diseases and some mechanisms are shared between aging and disease pathologies, a spectrum of diseases are commonly referred to as age-related diseases or conditions. (12, 14, 23) Among those, Alzheimer's disease, Parkinson's disease, some cancers, cardiovascular disease (CVD), chronic obstructive pulmonary disease (COPD), maculopathy, periodontitis, sarcopenia, osteoarthritis, and osteoporosis are extensively studied. (23) Previous evidence found the risk of developing a list of chronic diseases in the population grows exponentially from 30 to 70 years. (24, 25) However, the speed of disease burden seems to slow down among the oldest old, as evidence suggested around a quarter of centenarians could escape clinical symptoms of diseases successfully and the prevalent dementia risks leveled off at around the age of 95. (26, 27) Another obvious feature of human aging is the increased risk of death, the ultimate endpoint of life. Population statistics from the United States showed that the mortality rate almost doubled across every 10-year age group from 15–24 years to 75–84 years. (28, 29)

2.2 BIOLOGICAL AGE

Biological aging encompasses complex features and thus measuring the aging process in populations has long been a challenge. A well-known metric of aging is CA, the calendar periods that have passed by since the time of birth. With an easy-to-measure and easy-to-access definition, CA serves as a convenient metric of aging in most medical research. However, people of the same CA manifest diverse aging-related phenotypes and the magnitude of this heterogeneity often enlarges with growing age. Gerontologists and medical researchers need new indexes to keep track of the physiological and functional changes beyond measuring CA alone. Therefore, the concept of BA has been discussed.

BA concept

The American Federation for Aging Research has suggested that a 'biomarker of aging', often used synonymously as 'BA', should meet the following requirement (30, 31):

1. It must predict the rate of aging. In other words, it would tell exactly where a person is in their total lifespan. It must be a better predictor of life span than chronological age.
2. It must monitor a basic process that underlies the aging process, not the effects of disease.
3. It must be able to be tested repeatedly without harming the person. For example, a blood test or an imaging technique.
4. It must be something that works in humans and in laboratory animals, such as mice. This is so that it can be tested in lab animals before being validated in humans.

Despite years of efforts in the pursuit of qualified BAs, no such measure fulfilling all of the above criteria has been found. Some even believe that a single BA index is impossible to find for measuring a process as complex as biological aging. (32)

In this thesis, the definition of BA is relaxed from the above criteria. In a broad sense, a BA can be an indicator that changes in parallel with one or more aging aspects, including any cellular, organ-, system-, and/or organismal features. That is, identifying BA could be achieved by either pinpointing the causes of aging, such as the cellular hallmarks of aging, or measuring proxy indicators, which are correlated with aging-related phenotypes but do not necessarily play a causal role in the process of aging, for instance, aging-induced functional declines.

Utility of BA

Ideally, BA provides additional information in the assessment and prediction of aging risks independent of CA. Therefore, the CA-independent part of BA is of particular interest to researchers and often termed as BA residual, age acceleration, or delta age in the literature. Despite a lack of consensus, delta age is often calculated from a naïve subtraction between BA and CA; whereas BA residual and age acceleration refer to the difference between the observed BA and the expected BA among the people of the same CA, derived from a regression model (Figure 2.2.1).

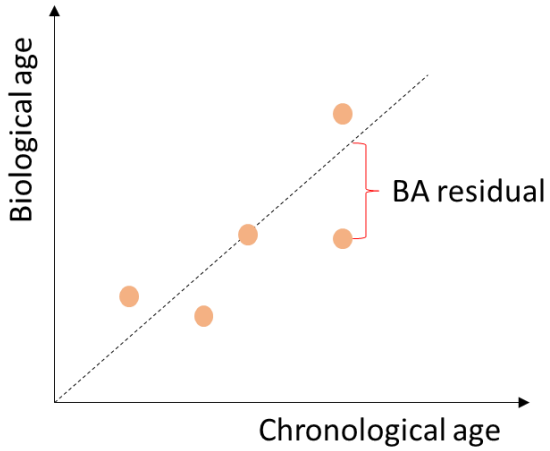

Figure 2.2.1 A simplified illustration of CA, BA, and BA residual. The expected BA-CA relationship in a population (dashed line) is fitted using the individual-level observations (dots). A BA residual

refers to the difference between the observed BA and the expected BA among the people of the same CA, of which the red curly bracket denotes an example.

Successful development of BA comes with opportunities to improve clinical practice and to facilitate academic research. First, by assessing aging-related health risks in the population, people at higher risks can be identified and delivered with personalized medical service. Second, by monitoring the aging changes repeatedly through BA measures, medical research could evaluate the efficacy of interventions and treatments to delay population aging.

BA candidates

A variety of measures depicting one or a combination of measures of physiological and/or functional deterioration can be seen as potential BAs. It could be a metric that is initially used to measure molecular and cellular changes, for instance, genetic instability, (33) telomere attrition, (34) and epigenetic alteration, (35) as well as comprehensive organismal-level functions like cognition, (36) physical functioning, (37) and frailty. (38) Indexes measuring the risk of death or developing aging-related conditions could also be treated as a type of BA, because increased vulnerability to morbidities and mortality occurs with aging. (39) Furthermore, BA can be a mixed measure of several aging mechanisms and aging manifestations to integrate their complex nature. (40, 41) So far, several review papers (42-53) have been devoted to discussing the multiple choices of BAs and some commonly-used BA measures are summarized in Table 2.2.1.

Table 2.2.1 Aging mechanisms, aging manifestations and potential BA measures

Underlying aging mechanisms ←----------→ Organismal aging manifestations					
Molecular and cellular mechanisms		Physiological and functional declines measured at organ-, system-, and organism-level		Disease and death risks	
Hallmarks	BA measures (50)	Declines	BA measures (8)	Conditions	BA measures (presence of the diseases) (23)
Genomic instability	Aneuploidy	Cardiovascular	Blood pressure	Premature aging disorders	Progeria, Werner syndrome
Telomere attrition	Leukocyte telomere length	Pulmonary	Forced vital capacity	Neurodegenerative disorders	Alzheimer's Disease, Parkinson's Disease
Epigenetic alterations	Methylation patterns	Renal	Serum creatinine	Cancer	Prostate cancer
Loss of proteostasis	Glycans (54)	Immune system	Bone marrow mass	Cardiovascular diseases	Hypertension, atherosclerosis
Mitochondrial dysfunction	Reactive oxygen species (55)	Neurologic function	Cerebral tissue atrophy	Pulmonary diseases	Chronic obstructive pulmonary disease
Cellular senescence	Senescence-associated beta-galactosidase (56)	Musculoskeletal	Muscle mass	Musculoskeletal disorders	Sarcopenia, osteoporosis
Deregulated nutrient sensing	Insulin-like growth factor 1 (57)	Sensory	Sensory cell loss	Inflammatory diseases	Rheumatoid arthritis
Comprehensive measures					
Epigenetics	Epigenetic clocks (58)	Cardiorespiratory fitness	Cardiorespiratory fitness (59)	Morbidity	Comorbidity (60), multi-morbidity (61), healthy lifespan (62)
Transcriptomics	Transcriptomic age (63)	Cognitive function	General cognitive ability (36)	Mortality	All-cause mortality risk
Proteomics	Proteomic signature of age (64)	Physical function	Global physical functioning scale (37)		
Metabolomics	Metabolic Age Score (65)	Frailty	Frailty phenotype (66), frailty index (38)		
Transcripts, proteins, metabolites, cytokines, microbes and clinical laboratory values	Ageotype (41)	Metabolic, cardiac, lung, kidney, liver, immune function	PhenoAge (40, 67)		

BAs studied in this thesis

Of all the potential BAs aforementioned, details will be given for four types of BA, namely TL, DNAmAges, multi-biomarker BA, and the FI, in the following section and Table 2.2.2, as they were extensively studied in this thesis.

Telomere length

The telomere is a sequence of repetitive nucleotide sequences located at the end of each chromosome. The telomere protects the chromosome's end from fusion and its length shortens during cell division. When TL shortening reaches a critical length, cells will no longer get replicated. Therefore, TL is seen as a cell mitotic clock and a measure of cellular aging. (68, 69)

Epidemiological evidence showed leukocyte TL attrition occurs with increasing age in both cross-sectional and longitudinal studies, of which Marioni et al reported an average loss of 48–67 base pairs/year in an older population. (70-72) Most studies modeled the TL declines in a linear trend, (70) i.e., a constantly changing rate, while some suggested the rate of TL-decline slightly accelerates after the age of 69. (73)

DNAmAges

Epigenetic changes can alter gene activities without changing the DNA sequence. (74) Methylation is a common type of epigenetic modification, by which a methyl group is added to the DNA molecule and often the consequence is the repression of gene transcription. The methylation levels at many cytosine-phosphate-guanine (CpG) dinucleotide sites change with age, showing a pattern of global hypomethylation and regional hypermethylation. (75, 76) Therefore, by aggregating methylation information of aging-related CpG sites through (penalized) regression models, a combined BA estimator, often referred to as DNAmAge or epigenetic clock, could be generated.

According to the nature of training targets, there are two groups of DNAmAges broadly. The first type is the CA estimator. Both Horvath and Hannum clocks are CA predictors, of which Horvath trained multiple tissues while Hannum used blood samples only. (58, 77) The second type of DNAmAges took a step closer to the BA estimators as they used aging-related phenotypes as training targets. For instance, DNAmPhenoAge and DNAmGrimAge utilized a biomarker-derived PhenoAge and biomarkers plus mortality risk in the training process, respectively. (67, 78) By construction, epigenetic clocks are CA-calibrated and, thus, increase with CA in a linear fashion in training populations. This positive BA-CA correlation has been supported repeatedly in different studies. (79, 80)

Multi-biomarker BA

Some biomarkers that are routinely assessed at clinics as well as in population surveys, including serum assays, urine biomarkers, and anthropometric markers, demonstrate age-

dependent patterns. Given a single biomarker only reflects organ or system function to a narrow extent, aggregating many biomarkers as a composite score could potentially maximize the aging-related biological information. Both unsupervised methods, such as principal component analysis (PCA), and supervised models, like a simple linear regression and a multi-step regression method proposed by Klemera and Doubal (KDM), can perform the aggregation task. (40, 81)

The biological relevance of a multi-biomarker BA largely depends on the component biomarkers and the training target in the BA development. Several biomarker-based BAs were developed in different populations using KDM and showed positive relationships between multi-biomarker BA and CA. (81-84)

FI

The frailty describes the gradual decline in "physiological reserve", leading to impaired robustness and resilience and characterized by the increased vulnerability to adverse outcomes. (22, 85-87) Two instruments are commonly used to assess frailty in population studies. The first one is proposed by Fried, the frailty phenotype (FP), and measures physical frailty through five clinical syndromes: unintentional weight loss, exhaustion, muscle weakness, walking speed in the lowest 20%, and low level of activity. (66) Another popular instrument is the FI, which considers aging as a process of deficit accumulation. The FI is calculated as the ratio of the number of deficits presented in a given individual to the total number of deficits, usually varying between 30 and 60, that have been taken into consideration. (38) Health deficits could be abnormal symptoms, signs, laboratory tests, disabilities, and diseases. Previous studies observed the frailty among different population settings and found that the FI gradually develops with increased CA within different age groups. (88)

Table 2.2.2 Four types of BA included in this thesis

	Data	Biological relevance	Measurement unit	BA-CA relationship
TL	Leukocyte TL	Cell mitotic aging	Base pair or a relative length	Negative
DNAmAges	Leukocyte DNA Methylation levels	Aging-related DNA methylation	Year	Positive
Multi-biomarker BA	Biomarkers such as blood and urine assays	Aging-related biomarker change	Year	Positive
FI	Health deficits like symptoms/signs, comorbidities, functions, lab values	Physiological reserve at organismal level	Ratio of the number of deficit presented to the total number of deficit considered	Positive

This thesis attempted to address BA-related research questions mainly from four aspects. In the following section, epidemiological evidence will be summarized.

2.3 CORRELATIONS OF BA

Almost every few months, a novel BA is described in the literature. With an overwhelming number of BA measures becoming available in the field, correlations could help us understand their connections.

Population studies showed leukocyte TL was only weakly correlated (absolute correlation coefficient <=0.2) with DNAmAge and DNAmAge acceleration, with inconsistent directions being reported. (72, 82, 89-91) Analysis of 6731 individuals in the United States found leukocyte TL was not correlated with a multi-biomarker BA, (92) which was further supported by another study conducted among 964 middle-aged participants in the Dunedin Study. (82) As for the frailty syndrome, Carvalho et al. meta-analyzed nine separate studies and found a weak but significant negative association of TL with frailty, assessed by both the FP and the FI. (93)

DNAmAges of different types generally presented moderate to high correlations. (78, 82, 94, 95) Grodstein et al estimated the correlations between four DNAmAges (Horvath, Hannum, PhenoAge, and Cortical) and found in blood samples pair-wise correlation coefficients ranged from 0.58 to 0.80. (95) DNAmAge accelerations were moderately correlated, with correlation coefficients between 0.17 and 0.45 reported. (78) With regard to associations with multi-biomarker BAs, DNAmAges, especially Hannum clock, showed a weak and positive correlation. (82) Further, the development of frailty, measured by both the FP and the FI, was associated with increased epigenetic age and age acceleration in old populations. (96, 97)

In the Singapore Longitudinal Aging Study, researchers built up a multi-biomarker BA from eight measures of kidney, lung, cognitive, and physical functions, and found that higher BA levels were predictive of the FP. (83) Chan et al. developed a 72 biomarkers-based BA in the UK Biobank (UKB) and observed an increased risk of hospital admissions (a subset of hospital frailty risk score) among biologically older individuals. (84) However, correlations between different multi-biomarker BAs are less studied in BA comparison studies.

Despite the rationale and implementation behind the FP and the FI not being quite the same, moderate correlations were found in different populations. (98-101)

In summary, previous studies found low to moderate correlations between different types of BA, including TL, DNAmAges, multi-biomarker BAs, and frailty measures. Between different BA types, TL was in a weak or null correlation with the others, while the remaining BAs presented positive correlations with varying magnitudes. Among DNAmAges, positive and moderate correlations were observed. A majority of the previous studies only took into account two BA types. Studies analyzing a range of BAs in the same population could gain an integrated view of BA correlations yet are largely lacking.

2.4 MORTALITY ASSOCIATIONS OF BA

Mortality is the ultimate endpoint of the aging process and estimating mortality associations with BAs could help interpret the utility of BAs.

The relationship between leukocyte TL at baseline and the risk of death during the follow-up were investigated among different populations and inconsistent results were found. (102-104) A meta-analysis pooled the data from 21 studies together and observed shorter LTL predicted a higher risk of all-cause mortality in the general population, and the effect size was weaker among people aged 80 years and over. (105) In addition, testing methods affected the precision of telomere measurements and seem to explain in part the inconsistent findings.

The ability of mortality prediction has been tested extensively during the development of the DNAmAges as a means of validation. (35, 79, 80, 106, 107) Despite being trained to be CA estimators, Horvath and Hannum clocks presented the ability to predict mortality risk independently of CA. Some recent clocks are explicitly trained to capture mortality-related methylation information. (78, 108) Indeed, mortality-oriented epigenetic clocks, as expected, outperformed other types of epigenetic clocks in the mortality prediction with respect to the magnitude of the effect size.

Using different biomarker panels, significant associations between higher multi-biomarker BAs and increased death risks were found in Asian, American, and European populations. (40, 83, 84, 109-111) Since the definitions of multi-biomarker BA are study-specific, a comparison across BAs would be informative but remains lacking.

The FI was a robust predictor of mortality risk across a number of populations. Despite the number and nature of health deficits considered by studies were slightly different, increased FI was robustly related to the increased risk of death, and the effect was independent of age, sex, and other lifestyle factors. (112) However, previous studies mostly focused on the general population and the old population. Thus, it remains understudied whether the FI effect on mortality would be modified by age and familial background.

Besides exploring a single BA, some studies aimed to gain a more global picture by investigating multiple BAs in the same population. Belsky et al. evaluated eleven aging quantifications using data from the Dunedin Study and found Hannum DNAmAge was consistently associated with aging-related outcomes, including physical functioning, cognitive decline, and subjective signs of aging. (82) However, mortality risk was not studied due to a lack of data availability. Recently, a few studies compared the BA-mortality association across several BA types. (91, 97, 113-115) Consistent findings are that different types of BAs tended to show low correlations, and the FI and mortality-oriented DNAmAges outperformed others in mortality prediction.

In summary, short TL showed inconsistent mortality association in individual studies and was related to slightly higher death risk in a meta-analysis, while the remaining BAs were robust predictors of mortality risk and showed negative associations. Given the well-recognized

correlations between distinct BAs, it is of interest and importance to know how BAs could work together to inform health risks in the population. However, evidence in this regard is limited.

2.5 CLINICAL BIOMARKERS AND HEALTHSPAN

Mortality risk is extensively studied in population studies. Improving longevity does not always go in parallel with the extension of good health. (62) The WHO reported an increase of 6.6 and 5.4 years from 2000 to 2019 in life expectancy and health-adjusted life expectancy (HALE) at birth, respectively, suggesting a healthy lifespan and lifespan might feature differently. (116) However, conceptualizing healthspan in the population is no easy task. At a population level, WHO uses HALE as a summary measure to quantify population health by accounting for disability rate and disability weights in each region. (117) However, at an individual level, the controversy of measuring healthspan lies in defining the threshold between healthy and unhealthy states. (118-120) Recently, a morbidity-defined healthspan measure came into use in a large-scale population study, in which healthspan is the age at the first occurrence of any diagnosis across seven chronic diseases. (25, 121) However, this morbidity-based measure could overlook other aging aspects such as functional changes, and could be heavily driven by early-onset diseases rather than late-onset conditions. Besides, this healthspan is, in essence, dichotomizing the health level; a two-level health framework is, at best, an over-simplification of the real health scale. Despite its limitation, this is a practical way to quantify healthspan in population studies and this healthspan definition will be the focus of this thesis.

At clinics, some biomarkers, such as fasting glucose and lipids, are routinely monitored by medical specialists. These disease-specific biomarkers have the potential to inform healthspan for two reasons. First, the aging process and disease pathology have shared mechanisms. Second, morbidity-defined healthspan is ended by disease onset. Therefore, associating clinical biomarkers with healthspan could potentially identify aging risk factors that are shared by a range of chronic diseases. Two previous studies have tested this hypothesis by estimating associations of the clinical biomarkers analyzed in this thesis. Using data from the Framingham Heart Study, Terry et al. assessed total cholesterol and glucose measured at age 40-50 in association with morbidity-free survival at age 85 years among 2,531 participants. (122) They found people with a lower level of total cholesterol and an absence of glucose intolerance had higher odds of survival. Similarly, among 2,008 individuals from the Rotterdam Study, Newson et al. observed a higher probability to be free from major diseases, including cardiovascular diseases, stroke, cancer, and dementia, among participants with a lower concentration of serum CRP at baseline. (123)

In summary, healthspan refers to the lifetime spent in good health condition and, practically, researchers treated the onset of chronic diseases as the endpoint of the healthspan. Total cholesterol, glucose tolerance, and CRP were found to be associated with the probability of morbidity-free survival at the age of 85 years. However, evidence for the other serum

biomarkers and the genetically proxied clinical biomarkers in relation to healthspan remains unknown.

2.6 GENETIC VARIANTS OF CRP

Serum CRP is a sensitive inflammatory marker. In the inflection/injury-induced acute inflammatory phase, serum CRP concentration could go up drastically by >1000 folds; while chronic inflammation is often denoted by a constant and low-grade increase in CRP level. (124, 125) CRP is also a non-specific biomarker. In other words, an increase in CRP indicates inflammation, but often gives no information on the affected organ or system. In clinical practice, CRP change needs to be accompanied by more specific biomarkers and symptoms/signs in order to make a meaningful inference. (126)

The serum concentration of CRP is found to be moderately heritable, with a twin-based heritability ranging from 0.1 to 0.65. (127-132) Genome-wide association studies (GWAS) and fine-mapping studies further scanned the common and low-frequency genetic variants in the relation to serum CRP. (133-145) The latest meta-GWAS identified 58 genetic loci that reached GWAS significance and GWAS signals were seen across almost all autosomal chromosomes. (143) The distinct variants at all identified loci and at the CRP locus explained 11.0% and 4.3% of the serum CRP variance, respectively. In addition, body mass index (BMI)-adjusted analyses suggested BMI-related pathways only explain the gene-CRP associations to a minimal extent. Recently, another research group performed a comprehensive GWAS analysis for common biomarkers using data from the UKB. (146) The fine-mapping analysis identified 99 distinct variants, which together explained 6.1% of the residual variance. Particularly, a low to moderate level of polygenicity was observed as the top 1% and 5% most heritable linkage disequilibrium (LD) blocks explained 27.8% and 39.5% of all heritability.

Limited by statistical power, standard GWAS could only interrogate genetic variants that are common or of low-frequency (minor allele frequency [MAF] >0.1%) even in the UKB, where genotypes were available for nearly half a million participants. (146) Rare variants are characterized by a lower LD with flanking variants, from which inferring causal loci is more straightforward than from a complex LD structure. (147) Additionally, the maximal magnitude of rare variant effect is, in theory, greater than that of common variants due to the force of natural selection, making rare variants ideal instruments to identify biological mechanisms and prioritize therapeutic targets. (148) Therefore, to complement our current understanding of the genetic influence on CRP, examining rare genetic variants is warranted.

Because of the low frequency, rare genetic variants are often grouped together by a functional unit and are subsequently associated with the trait of interest. (149, 150) A common collapsing approach of this type is the gene-based burden test. Schick et al. did an exome-wide burden test among 6,050 European Americans and 3,109 African Americans and did not identify any significant signal, likely due to the small sample size. (151) Recently, Cirulli and colleagues associated gene-based mutation burden with CRP among 46,765 UKB participants

and identified only one significant signal in the CRP gene region. (152) Of the 37,867 individuals with European ancestry, 66 (0.17%) were carriers of rare functional mutation and showed a significantly lower serum CRP level. Overall, the current search on CRP and rare genetic variants is limited in terms of sample sizes.

In summary, CRP is a systemic inflammatory biomarker, with the potential to inform aging due to “inflammaging”, i.e., the chronic and low-grade inflammation during aging. (153) GWASs have made major achievements in understanding the CRP-related mechanisms that are regulated by common and low-frequency genetic variants. Rare genetic variants in relation to serum CRP could further reveal mechanistic insights and drugable targets, yet are so far understudied. In addition, evidence to interpret the potential clinical relevance of rare variants, including gene-environment (G*E) interaction as well as association with diseases and aging-related phenotypes, is lacking.

3 RESEARCH AIMS

The overarching aim of this thesis is to understand correlation patterns, genetic determinants, and associations with health outcomes of multiple BAs (including composite scores and biomarkers that reflect aging-related changes). Towards this end, four individual studies were conducted with the following aims (Figure 3.1.1):

Aim 1 To estimate the associations of the FI with all-cause and cause-specific mortality, taking into account familial factors, and to test whether the associations are time-dependent.

Aim 2 To analyze correlations of nine BAs (TL, DNAmAges, multi-biomarker BA, functional BAs) and their associations with all-cause mortality.

Aim 3 To explore the associations of ten clinical biomarkers (glycemic, lipid-, inflammatory, and hematological markers) with healthspan, i.e., disease-free lifespan, and lifespan, and to identify putative causal relationships by leveraging genetic instruments.

Aim 4 To examine rare functional genetic variants in association with serum CRP at the whole exome-wide scale, to test a potential interaction effect between rare mutation and BMI, and to elucidate the clinical relevance of the identified gene by estimating the relative risks for a set of health outcomes, including diseases, frailty, and death.

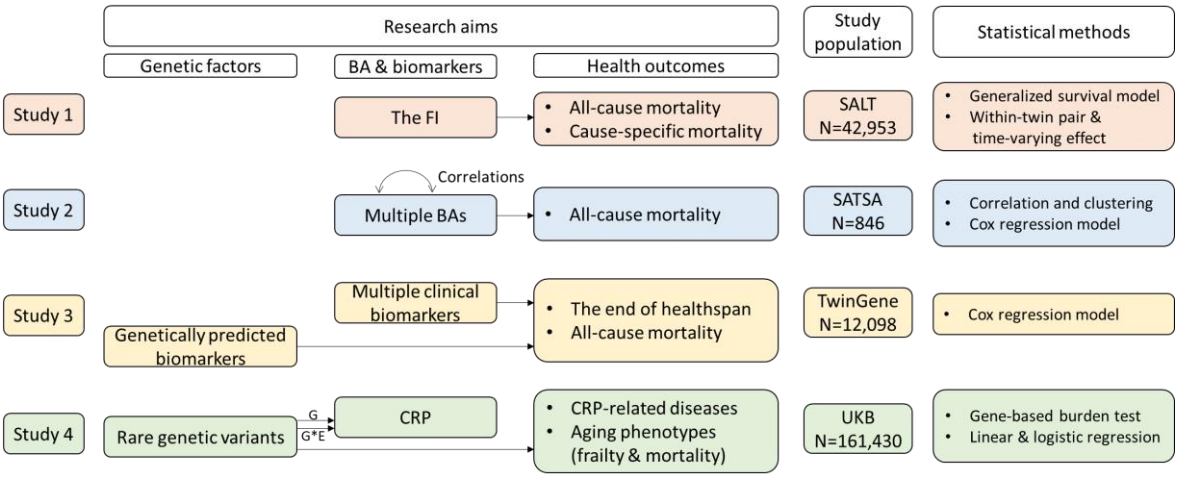

Figure 3.1.1 Overview of the research aims and the study designs

4 MATERIALS AND METHODS

4.1 STUDY POPULATION

All study populations came from two large-scale population resources, the Swedish Twin Registry (STR) and the UKB.

The STR was established in the late 1950s. Initially, the purpose was to investigate the effects of smoking and alcohol consumption on the risk of cancer and cardiovascular diseases while controlling for genetic background. (154, 155) To date, the STR has developed into a comprehensive infrastructure that facilitates research with broad interests through the data collection of genetic information, molecular biomarkers, lifestyle behaviors, and disease diagnoses. (156, 157) The latest updates in 2019 reported that 216,258 twins born between 1886 and 2015 were enrolled in the STR. (158) The STR comprises a number of sub-cohorts, and each of them comes with distinct measurements and study designs. Three sub-cohorts were analyzed in the first three studies of this thesis, respectively.

4.1.1 Screening Across the Lifespan Twin Study (SALT)

SALT is a telephone-based screening among all STR twins who were born in 1958 and earlier. A full-scale telephone survey started in 1998 and was completed in 2002. The telephone survey was structured to collect information about illnesses, health, prescription and nonprescription medication use, occupation, education, and lifestyle behaviors. A total of 44,919 twin individuals completed the interview. (154-157)

SALT participants were included in Study 1 if they 1) had less than 20% missing data across the 44 frailty items of interest, and 2) had valid follow-up information on all-cause and cause-specific mortality. Eventually, 42,953 twin individuals aged from 41 to 95 years were included in Study 1. The study population was further categorized into three overlapping subgroups within each sex: 1) 32,794 single responders, i.e., unrelated individuals, including twins whose partner did not respond, twins from opposite-sex twin pairs and one randomly selected member of each same-sex twin pair, 2) 11,812 DZ twin individuals from complete same-sex dizygotic (DZ) pairs, 3) 8,506 monozygotic (MZ) twin individuals from complete MZ pairs.

4.1.2 Swedish Adoption/Twin Study of Aging (SATSA)

SATSA is a longitudinal study consisting of twin pairs who were reared apart in early childhood and matched twin pairs who were reared together. (154, 155, 159) A subset of study individuals who were older than 50 years of age were invited to in-person testings (IPTs), which include questionnaire survey, cognitive assessment, and blood sampling. Nine complete IPT waves have been conducted from 1986 through 2014 (IPT1 to IPT10, except for IPT 4, where only telephone interviews were performed). The number of individuals participating in each IPT wave ranged from 645 in IPT1 to 269 in IPT10. In total, 859 individuals participated in at least one IPT wave in SATSA.

SATSA participants were included in Study 2 if they 1) had any BA measure, including TL, four DNAmAges (Horvath, Hannum, PhenoAge, and GrimAge), multi-biomarker BA, and three functional BAs (cognitive function, functional aging index [FAI], and the FI), assessed once across IPT waves, and 2) had known all-cause mortality information. Eventually, 846 individuals were included in Study 2.

4.1.3 TwinGene

Between 2004 and 2008, SALT participants were invited to join the TwinGene project. (157, 158) Individuals who responded to the invitation underwent a series of examinations, including self-reported questionnaire queries focused on chronic diseases and medicine use, a simple health check-up, and blood sample collection. Blood samples were later used for biochemistry assessments and array genotyping. A total of 12,646 individuals participated in TwinGene.

TwinGene participants were included in Study 3 if they 1) had at least one biomarker value across ten clinical biomarkers (fasting blood glucose [FBG], glycated hemoglobin [HbA1c], total cholesterol [TC], high-density lipoprotein cholesterol [HDL-C], low-density lipoprotein cholesterol [LDL-C], Apolipoprotein A1 [ApoA1], Apolipoprotein B [ApoB], triglyceride [TG], CRP, and hemoglobin [Hb]), 2) had non-missing values across educational attainment, BMI, and smoking status, and 3) had known information on disease diagnosis and vital status. Eventually, 12,098 individuals were included in Study 3.

4.1.4 UKB

The UKB is a large-scale prospective cohort and recruited >500,000 participants aged 40–69 years from across the United Kingdom in 2006–2010. (160-162) Participants are extensively phenotyped and genotyped through multiple assessments, including questionnaires, physical measures, bio-sample assays, genome-wide array genotyping, and whole-exome sequencing (WES). In addition, diagnosis and death for all participants are followed up through hospital inpatient data and mortality data.

In Study 4, we included UKB participants who had 1) non-missing CRP measurement at baseline, i.e., the initial assessment visit, 2) European ancestry, 3) contributed a sample that passed quality control (QC) for WES genotypes and arrayed genotypes. This resulted in an eligible sample set with 177,242 individuals. Next, we retained one of the members randomly among sets of people with third-degree or higher relatedness, so that all the study participants in the analysis are approximately unrelated. Furthermore, we separated participants randomly into two groups, containing 75% and 25% of the independent participants, which was used as discovery and replication set, respectively. The combination of discovery and replication yielded a total of 161,430 participants and will be referred to as “meta-analysis set” hereafter.

An overview of assessment timeline is displayed in Figure 4.1.1.

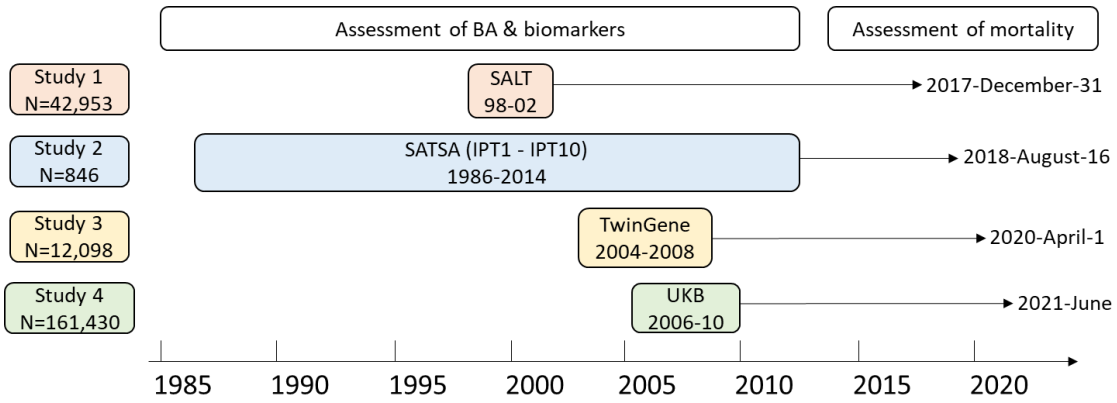

Figure 4.1.1 Timeline of assessments

4.2 MEASUREMENTS

4.2.1 Biological ages

FI in SALT

The FI was constructed from self-reported data. We considered forty-four symptoms, signs, disabilities, and diseases covering a wide range of biological systems and associated with health status. The deficit items and scoring are presented in Table 4.2.1. An FI value for each individual was calculated as the number of deficits present divided by the total number of deficits (n=44). For example, an individual reporting 5 deficits has an FI value of 5/44=0.11. The FI value was scaled to a unit of 10% in the associational analysis.

Table 4.2.1 44 frailty items and the coding rules

No.	Questions	Coding
1	How do you estimate your general health?	Excellent=0, Good=0.25, Average=0.5, Not so good=0.75, Bad=1
2	Do you think your health status prevents you from doing things you want to do?	Not at all=0, To some extent=0.5, A great deal=0
3	How many times a year do you get serious infections (other than respiratory)?	0-1 times=0, 2-4 times=0.5, 5 times or more =1
4	Do you have buzzing in the ears?	Both ears or one ear=1, No=0
5	Do you have or have you had angina pectoris	No=0, Yes=1
6	Do you have or have you had heart attack	No=0, Yes=1
7	Do you have or have you had heart failure	No=0, Yes=1
8	Do you have or have you had high blood pressure	No=0, Yes=1
9	Do you have or have you had lipid disorder, for example high cholesterol or high triglycerides	No=0, Yes=1
10	Do you have or have you had vascular spasm in the legs (intermittent claudication)	No=0, Yes=1
11	Do you have or have you had clot in the leg (venous thrombosis)	No=0, Yes=1
12	Do you have or have you had cerebral hemorrhage or clot in the brain (stroke)	No=0, Yes=1

No.	Questions	Coding
13	Do you have or have you had TIA attacks (temporary weakness or paralysis or reduction of sensibility)	No=0, Yes=1
14	Do you have or have you had irregular cardiac rhythm/atrial fibrillation	No=0, Yes=1
15	Do you have or have you had chronic lung disease (including chronic bronchitis and emphysema)	No=0, Yes=1
16	Do you have or have you had dizziness	No=0, Yes=1
17	Do you have or have you had rheumatoid arthritis	No=0, Yes=1
18	Do you have or have you had knee joint problem	No=0, Yes=1
19	Do you have or have you had sciatica	No=0, Yes=1
20	Do you have or have you had osteoporosis	No=0, Yes=1
21	Do you have or have you had hip joint problem	No=0, Yes=1
22	Do you have or have you had back pain	No=0, Yes=1
23	Do you have or have you had neck pain	No=0, Yes=1
24	Do you have or have you had diabetes (including old age diabetes, and excluding pregnancy diabetes)	No=0, Yes=1
25	Do you have or have you had goiter	No=0, Yes=1
26	Do you have or have you had glandular diseases (excluding goiter)	No=0, Yes=1
27	Do you have or have you had gall bladder problem	No=0, Yes=1
28	Do you have or have you had liver disease (for example, cirrhosis)	No=0, Yes=1
29	Do you have or have you had gout	No=0, Yes=1
30	Do you have or have you had kidney disease	No=0, Yes=1
31	Do you have or have you had stomach or intestine problems	No=0, Yes=1
32	Do you have or have you had recurring urinary tract problems	No=0, Yes=1
33	Do you have or have you had cancer, tumor disease or leukemia	No=0, Yes=1
34	Do you have or have you had migraine	No=0, Yes=1
35	Do you have or have you had asthma	No=0, Yes=1
36	Do you have or have you had allergy	No=0, Yes=1
37	Do you have recurrent periods of coughing?	No=0, Yes=1
38	You felt depressed during the past week?	Never or almost never=0, Seldom=0.5, Often, always or almost always=1
39	You were happy during the past week?	
40	You felt lonely during the past week?	
41	Do you have or have you had any physical handicap	No=0, Yes=1
42	Do you have or have you had Crohn's disease or Ulcerative colitis	No=0, Yes=1
43	How is your vision?	Good=0, Reduced=0.5, Highly reduced or blind=1
44	How is your hearing?	Good=0, Reduced=0.5, Highly reduced=1

Multiple BAs in SATSA

Telomere length

Leukocyte TL was measured by quantitative polymerase chain reaction (qPCR)-based technique. (73) The measurement is the relative length (T/S ratio) calculated as the ratio of the individual telomeric DNA (T) to a piece of reference DNA containing a single copy gene (S). The TL values were corrected for batch effect, and TL outliers (exceeding Mean±4*standard deviation [SD]) were omitted in the analyses.

DNAmAges

Genome-wide methylation levels in leukocytes were measured by Illumina's Infinium HumanMethylation 450K BeadChip and quantified as beta-values. (63, 163) DNAmAges are weighted sums of methylation levels across specific CpG sites, of which the feature selection and aggregation algorithms were developed elsewhere. (58, 67, 77, 78) In summary, we utilized four types of DNAmAges (Horvath, Hannum, PhenoAge, and GrimAge), which incorporated methylation levels of 353 age-related, 71 age-related, 513 phenotypic age-related, and 1,030 mortality risk-related CpG sites, respectively. DNAmAges is calibrated against the CA scale and presented in the year unit. That means a DNAmAge value of 50 years corresponds to the expected BA level among the people with a CA of 50 years in the training population. DNAmAges were calculated with the help of an online DNA Methylation Age Calculator. (164)

Physiological age (multi-biomarker BA)

Physiological age is a type of multi-biomarker BA. A set of serum biomarkers and clinical markers were taken into account in the BA development, including serum hemoglobin, glucose, cholesterol, Apolipoprotein B, triglyceride, BMI, waist-hip ratio, weight, waist circumference, hip circumference, systolic BP, and diastolic BP. We used PCA and KDM to combine CA and biomarkers into a single physiological age value in men and women separately. Similar to the DNAmAges, physiological age is calibrated against the CA scale and presented in the year unit.

Cognitive function

We used a general cognitive ability score to measure cognitive function. Four cognitive domains were taken into account, including crystallized, fluid, memory, and perceptual speed abilities assessed through in-person cognitive testing. (36) PCA and a T-score scaling were applied in the score development. The values of resulting cognitive function were distributed with a mean value of around 50 and a SD of around 10.

Functional aging index (FAI)

We used FAI to measure functional ability, with a focus on the physical aspects. (165) Four functional measurements were taken into consideration: 1) vision and hearing were combined to create a measure of self-reported sensory ability, 2) muscle strength, 3) walking speed time, and 4) lung function. The four indicators were standardized and then summed to create a composite score. The values of FAI were distributed with a mean value of around 50 and a SD of around 10.

FI

Similar to the FI in SALT, the FI in SATSA was constructed from 42 self-reported health deficits. Details of FI items are described elsewhere. (166)

BA residuals

BA residuals were constructed to represent the deviation between the observed BA and the expected BA among those with the same CA. This is done by regressing out the CA-related part from the respective BA using a generalized linear model. Natural splines of CA were used in the model to accommodate both linear and non-linear relationships between BA and CA.

All BAs and BA residuals were scaled to the SD unit in the associational analyses.

4.2.2 Clinical biomarkers

Multiple clinical biomarkers in TwinGene

We investigated circulating concentrations of clinical biomarkers that reflect glycemic control (FBG, HbA1c), lipid metabolism (TC, HDL-C, LDL-C, ApoA1, ApoB, TG), inflammation (CRP), and hematological function (Hb). HbA1c was assessed by ion exchange chromatography and the other clinical biomarkers were measured by a semi-automated biochemistry analyzer (Beckman Coulter, CA). Biomarkers that appeared strongly right-skewed (FBG, HbA1c, TG, and CRP) were firstly log-transformed; all biomarkers were then standardized to SD units.

CRP in UKB

High sensitivity CRP was measured by Beckman Coulter AU5800 at baseline from 2006 to 2010. The original value of CRP (unit: mg/L) presented a skewed distribution and the natural logarithm of CRP was computed and used in the associational analyses.

4.2.3 Genetic variants

Polygenetic risk score (PRS) in TwinGene

Genotypes were assessed on Illumina OmniExpress BeadChips. Arrayed genetic data were then imputed against the reference panels in the 1000 Genomes Project phase 1 version 3.

To calculate polygenic risk scores for each clinical biomarker, we used the summary statistics from previous GWASs (167) and the genotypes in TwinGene. A detailed procedure can be found elsewhere. (168) Briefly, we calculated the PRSs as a weighted sum of the biomarker-elevating alleles that were both GWAS significant (P<5E-8) and LD independent (r^2 <0.1). In summary, PRSs were derived from a different number of SNPs (from 63 for FBG to 590 for HDL-C), and were standardized to SD units.

Genotypes in UKB

Of the genetic variants released by the WES interim 200k release, we selected PA variants that are 1) quality controlled, 2) in the coding sequence (CDS) regions, 3) rare, namely MAF<0.1%, 4) single nucleotide variants (SNVs) or indels that lead to any of the following functional consequences: splice acceptor variant, splice donor variant, stop gained,

frameshift, stop lost, start lost, inframe insertion, inframe deletion, missense variant, and protein altering variant based on the annotation results from the Ensembl Variant Effect Predictor (VEP 103.1), (169) and 5) not benign missense mutation (Sorting Intolerant From Tolerant [SIFT]>0.05 and Polymorphism Phenotyping [PolyPhen]<0.15). (170, 171) Of all PA variants, a subset with high-impact consequences, including splice acceptor variant, splice donor variant, stop gained, frameshift, stop lost, and start lost, are classified as LOF variants. As a result, 1,776,249 PA and 266,226 LOF variants were included in the present analysis.

In the exome-wide burden test, we subsequently collapsed the PA variants according to mapped genes. That is, for each gene, participants were assigned to value 1 if carrying any PA mutation in the corresponding gene region and to value 0 otherwise (Figure 4.2.1). Similarly, LOF variants were collapsed using mapped genes. Eventually, 21,270 and 20,047 genes were analyzed in the PA and LOF burden test, respectively.

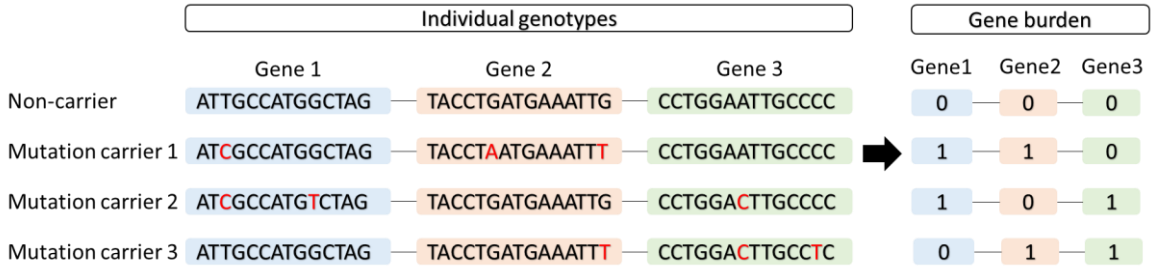

Figure 4.2.1 An illustration of gene burden calculation. The existence of any PA/LOF allele (in red letters) in one gene region leads to a value of 1, otherwise to a value of 0.

4.2.4 Health outcomes

Health outcomes in STR

All-cause and cause–specific mortality

All-cause mortality, including vital status and dates of death, were obtained from the Swedish Population Register. The all-cause mortality information was updated through 2017-December-31, 2018-August-16, and 2020-April-1 in Study 1-3, respectively.

Cause-specific mortality data were obtained from the Cause of Death Register. In study 1, three specific causes of death, including CVDs (including stroke), respiratory-related causes, and cancer, were analyzed, with the latest update on December 31, 2014.

The end of healthspan

In Study 3, we defined the end of the healthspan as the age at the first occurrence of any following conditions (referred to as “any chronic disease” hereafter): cancer, diabetes, CVDs (coronary heart failure [CHF], myocardial infarction [MI], stroke), COPD, dementia, and death. Disease diagnosis was ascertained through linkages to the Swedish National Patient Register. For each disease, we treated the earliest admission date (inpatient record) or record

date (outpatient care) as the onset date. Healthspan information was followed up through December 31st 2016.

ICD codes of diseases were described elsewhere. (168, 172)

Health outcomes in UKB

Frailty

Frailty at baseline was measured through two instruments, the FP and the FI, at baseline. (173, 174) The FP measures five domains, including weight loss, exhaustion, slowness, low physical activity, and weakness, of which weakness was assessed through grip strength and the other measures were self-reported. People who meet 1-2 and 3-5 criteria are defined as pre-frail and frail, respectively. The FI takes into account 49 self-reported frailty items. As the FI shows a skewed distribution in the population, we first log-transformed and then rescaled the FI to the SD unit. The transformed FI was used in the associational analysis.

Disease diagnosis

A list of CRP-associated diseases were examined in the present analysis, (175) including autoimmune and inflammatory (Celiac disease, inflammatory bowel disease [IBD, all types], Crohn's disease, Ulcerative colitis, Psoriatic arthritis, Rheumatoid arthritis, Type 1 diabetes, Knee osteoarthritis), cardiovascular (coronary artery disease [CAD], Ischemic stroke), metabolic (Type 2 diabetes, Chronic kidney disease), neurodegenerative (Alzheimer disease, Parkinson disease), and psychiatric diseases (Bipolar disorder, Major depressive disorder). The first occurrence of a diagnosis was ascertained through self-reported health conditions at baseline, hospital inpatient data as well as causes of death from registers. Inpatient data and death information were updated in June 2021.

All-cause mortality

All UKB participants were linked to the death register and were followed up until June 2021.

4.3 STATISTICAL ANALYSIS

4.3.1 Generalized survival models

In the first study, we used a generalized survival model (GSM) to estimate the association between baseline FI and mortality risk during the follow-up. (176) The event of interest was defined as the occurrence of death due to all causes or the corresponding causes of death in all-cause and cause-specific mortality models, respectively. All models controlled the effect of age implicitly through underlying time scale, sex, education years, smoking, and BMI.

First, we assumed that hazard ratios (HRs) were time-constant, i.e., proportional hazard. The survival models were fitted separately for single responders (i.e., unrelated participants), same-sex DZ pairs, and MZ pairs. Specifically, for twin-pair models, a between-within decomposition along with a random effect were introduced to the GSM. (177, 178) In this

manner, shared familial effects both due to shared patterns of exposure and measured confounders included in the model, as well as general unmeasured similarity in survival patterns within twin pairs were adjusted for. In other words, the HRs of interest represent the within-population, within-DZ pairs, and within-MZ pairs effects, respectively.

Second, we fitted analogous models for all three groups where the HRs were allowed to vary with time, i.e., time-varying HRs.

4.3.2 Correlations and mortality associations of multiple BAs

In the second study, as BAs were assessed in multiple SATSA waves, repeated measurements were used in the estimation of the correlation coefficients across nine BAs. We selected complete measurements where all BAs were assessed for the same individuals and performed a repeated measures correlation analysis to control the relatedness between repeated measurements. (179) The same correlation analyses were replicated for nine BA residuals to quantify the BA correlations that cannot be explained by CA. To further cluster BAs, we transformed correlation coefficients between BAs to distances and performed hierarchical cluster analysis on BAs and BA residuals using Ward's method.

In the survival analysis, we only used baseline measurements, i.e., the first available measurement when repeated measurements were available for the same person, and fitted the Cox regression model to estimate the association between baseline BAs or BA residuals and the risk of all-cause mortality during the follow-up. All models adjusted for age implicitly through underlying time scale, sex, education, BMI, and smoking status and accounted for the left truncation and right censoring. To adjust for relatedness within twin pairs and subjects, robust standard errors were introduced. First, all models took only one BA/BA residual into account (one-BA models). Second, all nine BAs entered into the same survival model altogether (multi-BA models) to estimate the BA-independent effect. To minimize collinearity between BAs due to their CA-related feature, only BA residuals were analyzed in the multi-BA model.

4.3.3 Cox regression models of clinical biomarkers and PRSs

In the third study, we applied Cox regression models to estimate the association between serum biomarkers at baseline and the hazard of outcomes (any chronic disease and death) during follow-up among all participants, men, and women, respectively. Each participant was followed up until the age of any chronic disease, death, or the end of follow-up. All models were adjusted for age implicitly through underlying time scale, sex, birth year category in decades, educational attainment, BMI, smoking status, and statin usage. We also used robust standard errors to account for relatedness within twin pairs and Benjamini-Hochberg false discovery rate (FDR) to correct multiple testing. Serum biomarkers were analyzed individually.

To estimate the associations between the genetic propensity of clinical biomarkers and the outcomes of interest, we adopted a similar Cox regression approach as above. All survival

models were adjusted for age implicitly through underlying time scale, sex, birth year category, and the first ten genomic principal components (PCs) to account for population stratification. Once significant PRS results were found, further analyses were conducted to explore the potential pathways that underlie the association (Figure 4.3.1). Taking FBG PRS and healthspan for instance, two additional models were estimated with further adjustment for 1) BMI and serum TG, LDL-C, HDL-C, CRP, Hb, and 2) serum FBG.

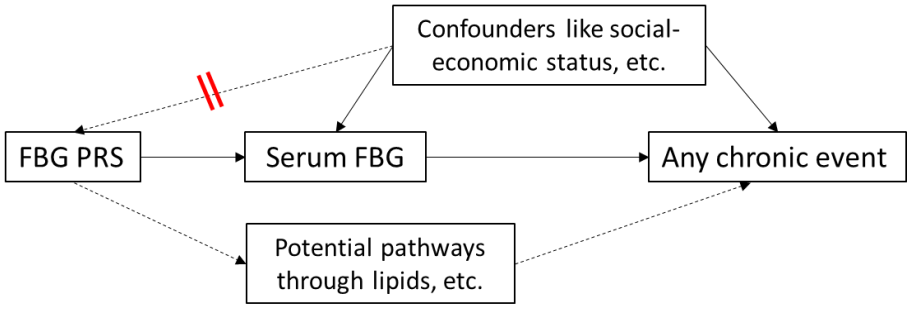

Figure 4.3.1 An illustration of the relationship between PRSs, clinical biomarkers, confounders, potential mediators, and outcomes. The relationship of interest is between the FBG PRS and the risk of any chronic event, i.e., healthspan, mediated through serum FBG. Common confounders influencing serum FBG-healthspan association, such as social-economic status, will no longer distort the FBG PRS-healthspan relationship because genes regulating FBG are determined at conception and are likely independent of factors exposed in childhood and adulthood. FBG PRS may affect healthspan through mediators other than serum FBG, such as by regulating serum lipid levels.

4.3.4 Gene-based burden test

We performed a whole exome-wide burden test by associating gene burden, i.e., the existence of any PA/LOF mutation in the gene region, with serum CRP concentration. The beta coefficients measure the effect of the gene burden on log-transformed CRP level, with a >0 value indicating CRP-increasing effect and <0 value denoting CRP-decreasing effect. We performed the CMC burden test (180) twice in two independent samples, both the discovery set and the replication set, meta-analyzed the effect estimates, and eventually report the results for the meta-analysis set. All models controlled the effects of age at baseline, sex, WES release, and 20 genomic PCs and were fitted in Rvtests (version: 20190205). (181) Once candidate genes from the above burden tests were identified, the single variant-CRP association at the candidate loci would be examined.

Further, to interpret the clinical relevance of the identified rare mutation, we did two additional analyses. First, we tested a possible interaction effect of rare gene*BMI on the level of serum CRP. Second, we estimated the associations between PA mutation burden in candidate genes and the risks of a series of diagnoses, frailty (pre-frail and frail defined by the FP and the FI), and all-cause mortality. Linear regression and logistic regression models were fitted to estimate beta coefficients and odds ratios (OR) when continuous values (serum CRP and the FI) and dichotomized values (diagnoses, pre-frail and frail, and mortality) were the outcomes of interest, respectively. All models were adjusted for age at baseline, sex, birth year in 10-year categories, WES release, and the first 20 genomic PCs.

5 RESULTS

5.1 STUDY 1

5.1.1 Population characteristics

Of 42,953 participants, 19,924 (46.4%) were men and 31,866 (74.2%) were comprised of complete twin pairs (19.8% MZ pairs, 27.5% same-sex DZ pairs, and 26.9% opposite-sex pairs). At baseline, the median levels of FI were 0.108, 0.097, and 0.119 in all, men, and women, respectively (Table 5.1.1). Specifically, the FI has a right-skewed distribution (Figure 5.1.1).

Table 5.1.1 Baseline characteristics of the study population in SALT

	All	Men	Women
Number of participants	42953	19924	23029
Age at baseline, mean (SD)	58.8 (10.7)	58.4 (10.4)	59.2 (11.0)
BMI (kg/m^2), mean (SD)	25.0 (3.5)	25.5 (3.1)	24.5 (3.7)
Education (year), mean (SD)	10.4 (3.2)	10.5 (3.2)	10.4 (3.2)
Tobacco products use, N (%)	25048 (58.3%)	12677 (63.6%)	12371 (53.7%)
FI, median (IQR)	0.108 (0.114)	0.097 (0.097)	0.119 (0.125)
History of diseases at baseline, N (%)			
CVD	15487 (36.1%)	6981 (35.0%)	8506 (36.9%)
Respiratory disease	4580 (10.7%)	1891 (9.5%)	2689 (11.7%)
Cancer	3022 (7.0%)	932 (4.7%)	2090 (9.1%)

IQR, interquartile range.

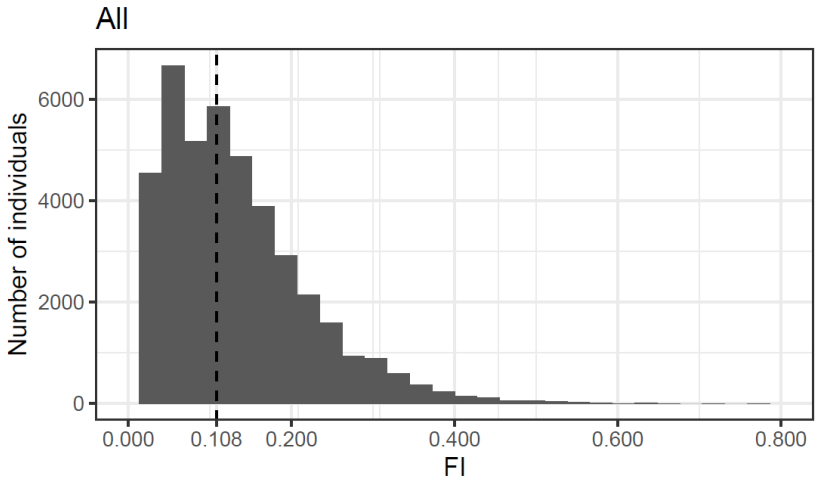

Figure 5.1.1 The distribution of the FI in all participants. The plot illustrated the number of individuals (y scale) against the level of FI (x scale) among all participants.

5.1.2 Time-constant FI-mortality associations

During the up to 20 years of follow-up, 12,222 (28.5%) deaths were documented, of which 3,270 (7.6%), 1,051 (2.4%), and 3,302 (7.7%) deaths were caused by CVDs, respiratory-related causes, and cancer, respectively.

An increase in FI was significantly associated with higher risks of deaths due to all causes, CVD, and respiratory-related causes (Table 5.1.2). No significant associations were observed for cancer mortality. Specifically, the within-pair associations were significant among both DZ and MZ twin pairs, suggesting the effects were independent of shared familial factors.

Table 5.1.2 Time-constant associations between 10% increase in the FI and mortality

	Single responders	DZ twins	MZ twins
Men(N=19924)			
All causes	1.28(1.24,1.32)	1.40(1.27,1.55)	1.34(1.13,1.58)
CVD	1.31(1.23,1.40)	1.35(1.11,1.66)	1.37(0.97,1.92)
Respiratory-related	1.23(1.11,1.38)	1.44(1.01,2.05)	2.03(1.14,3.60)
Cancer	1.06(1.00,1.14)	1.15(0.95,1.40)	0.99(0.73,1.34)
Women(N=23029)			
All causes	1.21(1.18,1.25)	1.25(1.15,1.35)	1.30(1.14,1.49)
CVD	1.27(1.15,1.34)	1.45(1.21,1.73)	1.83(1.35,2.49)
Respiratory-related	1.26(1.15,1.39)	1.28(0.97,1.69)	1.62(1.02,2.58)
Cancer	1.05(0.99,1.11)	0.96(0.81,1.13)	1.19(0.92,1.55)

The table is reproduced from Li et al. BMC Medicine 2019. (172)

5.1.3 Time-dependent FI-mortality associations

We next estimated the FI-mortality associations in an age-dependent manner and found evidence supporting a time-dependent effect (P for time effect <0.003 for all models). A relatively greater HR was associated with the FI at middle age, and effect sizes decreased gently with increasing age at FI assessment (Figure 5.1.2).

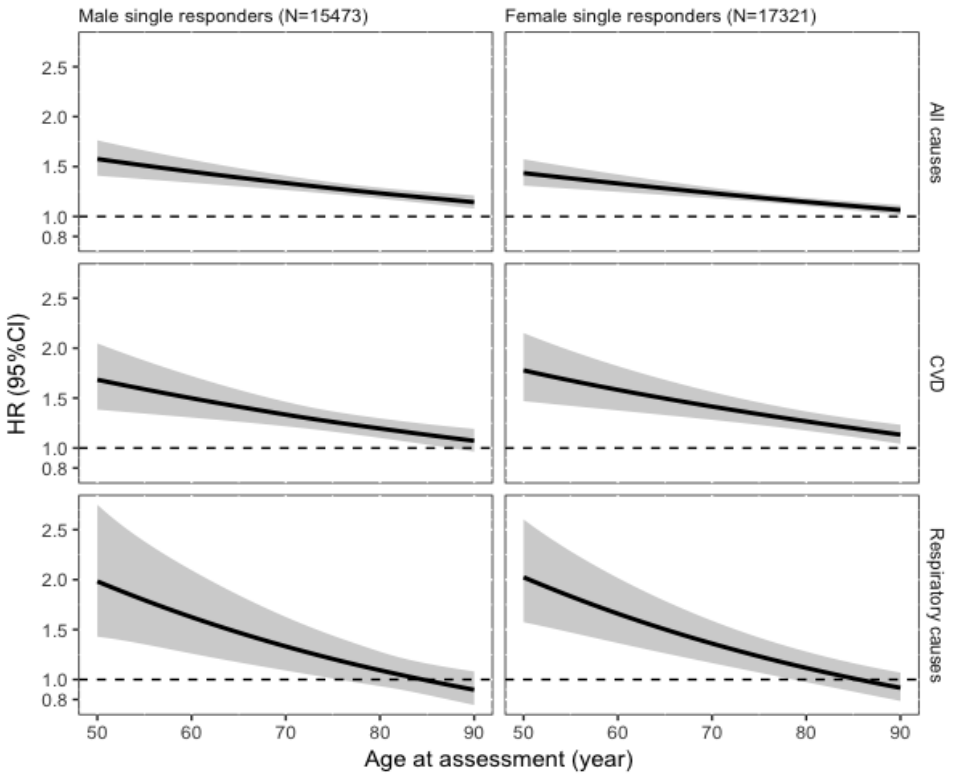

Figure 5.1.2 Time-dependent effects of 10% increase in the FI on mortality risk in single responders. The plot illustrated the estimated HRs (y scale) related to a 10% increase in the FI in association with three outcomes according to the age at FI assessment (x scale). The figure is reproduced from Li et al. BMC Medicine 2019. (172)

5.2 STUDY 2

5.2.1 Population characteristics

Of 845 participants, 342 (40.5%) were men and 800 (94.7%) were comprised of complete twin pairs (37.9% MZ pairs and 56.6% DZ pairs; Table 5.2.1). On average, each individual was assessed 4.7 times longitudinally. BA measures were available in a different number of individuals, ranging from 387 (for DNAmAges) to 829 (for cognitive function). The TL and cognitive function showed a negative relationship with CA, while the remaining BAs increased with growing CA (Figure 5.2.1).

Table 5.2.1 Baseline characteristics of the study population in SATSA

	All	Men	Women
N	845	342	503
Age at baseline, mean (SD)	63.6 (8.6)	62.7 (8.0)	64.3 (8.9)
BMI (kg/m^2), mean (SD)	25.6 (3.9)	25.8 (3.4)	25.5 (4.2)
Above primary education, N (%)	329 (40.3%)	143 (42.6%)	186 (38.7%)
Current and ex-smokers, N (%)	211 (25.0%)	117 (34.2%)	94 (18.7%)

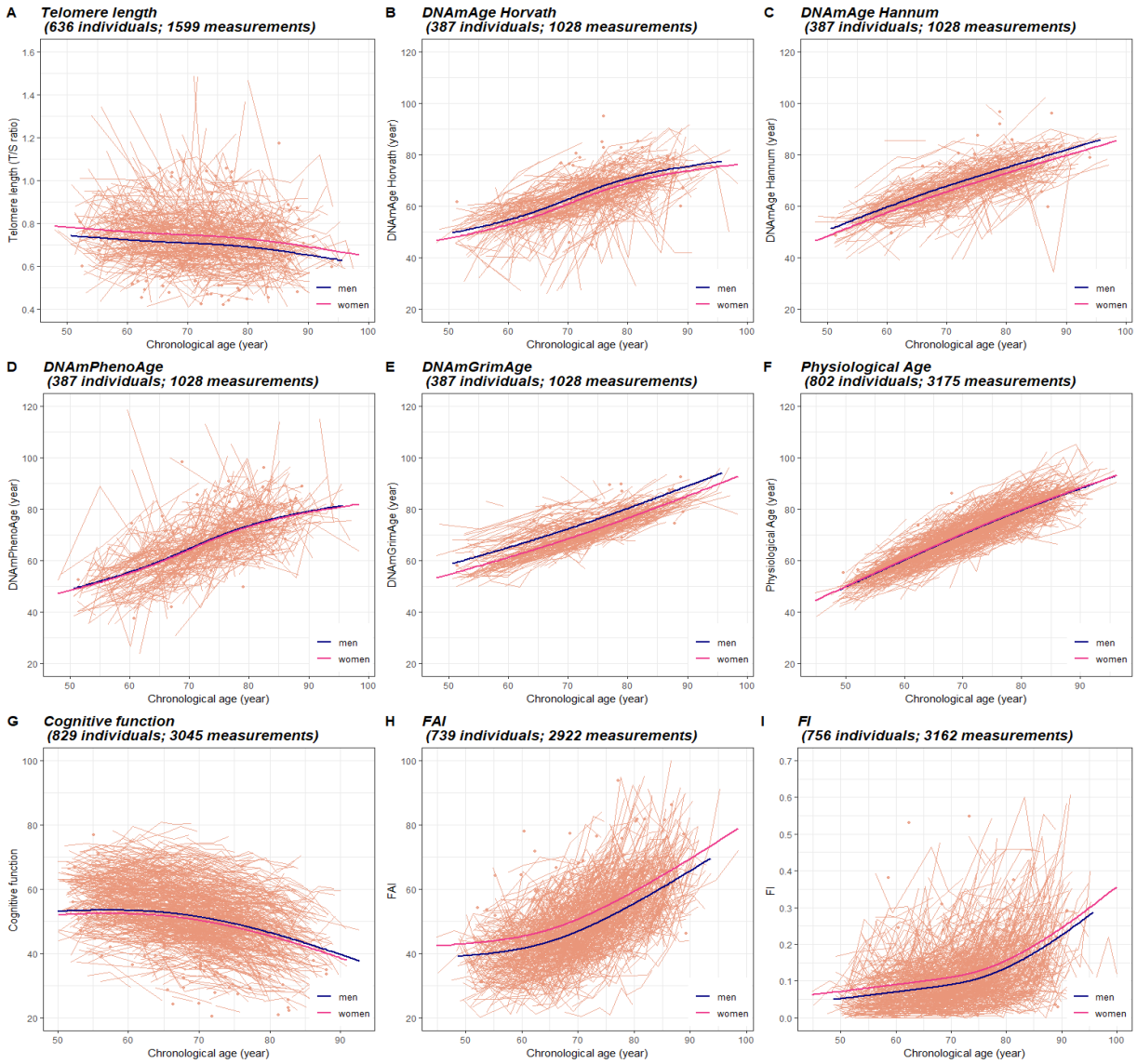

Figure 5.2.1 Longitudinal trajectories of BAs in 845 individuals (3973 measurements) with information on at least one BA. The plot illustrated the relationships between BAs (y scale) and CA (x scale). In each panel, individual-level data were denoted in orange and repeated measurements assessed among the same individuals were connected by broken lines. Blue and pink smooth lines represent the population average level among men and women, respectively. This figure is reproduced from Li et al eLife 2020. (182)

5.2.2 BA correlations and clustering

Telomere length showed low correlations with both CA and the other BAs ($r \leq 0.16$), while the remaining BAs were correlated to moderate and high degrees ($0.24 \leq |r| \leq 0.87$). Correlations between BA residuals, which represent the CA-independent information, were attenuated compared to BA correlations. Moderate correlations remained between DNAmAges (Horvath and Hannum) and between functional BAs (cognitive function and FAI, and FAI and FI; Table 5.2.2). In the hierarchical cluster analysis, the same types of BAs, i.e., DNAmAges and functional BAs, tended to be more closely related. Within DNAmAges residuals, GrimAge and PhenoAge were separated from the other two DNAmAges (Figure 5.2.2).

Table 5.2.2 Correlation coefficients of BAs in 288 individuals

	CA	TL	DNAmAge				Physiol	Functional BA		
			Hor	Han	Pheno	Grim		Cog	FAI	FI
Correlations of BAs										
CA	1.00									
TL	-0.11	1.00								
Hor	0.53	-0.09	1.00							
Han	0.64	-0.16	0.56	1.00						
Pheno	0.53	-0.07	0.27	0.41	1.00					
Grim	0.85	-0.09	0.44	0.59	0.49	1.00				
Physiol	0.87	-0.07	0.48	0.58	0.47	0.74	1.00			
Cog	-0.45	-0.10	-0.25	-0.24	-0.30	-0.42	-0.42	1.00		
FAI	0.54	-0.06	0.32	0.31	0.28	0.43	0.49	-0.50	1.00	
FI	0.45	-0.06	0.25	0.26	0.24	0.36	0.39	-0.29	0.48	1
Correlations of BA residuals										
CA	1.00									
TL	<0.01	1.00								
Hor	-0.02	-0.04	1.00							
Han	-0.04	-0.12	0.35	1.00						
Pheno	0.00	-0.01	-0.02	0.11	1.00					
Grim	0.07	-0.01	0.00	0.13	0.09	1.00				
Physiol	-0.10	0.04	0.06	0.05	0.02	0.03	1.00			
Cog	0.03	-0.17	-0.04	0.02	-0.06	-0.07	-0.12	1.00		
FAI	0.01	0.00	0.07	-0.04	-0.02	-0.08	0.07	-0.32	1.00	
FI	-0.07	-0.03	0.05	-0.01	-0.02	-0.04	0.02	-0.10	0.31	1

Hor Horvath DNAmAges, Han Hannum DNAmAges, Pheno DNAmPhenoAge, Grim DNAmGrimAge, Physiol Physiological age, Cog Cognitive function. This table is reproduced from Li et al eLife 2020. (182)

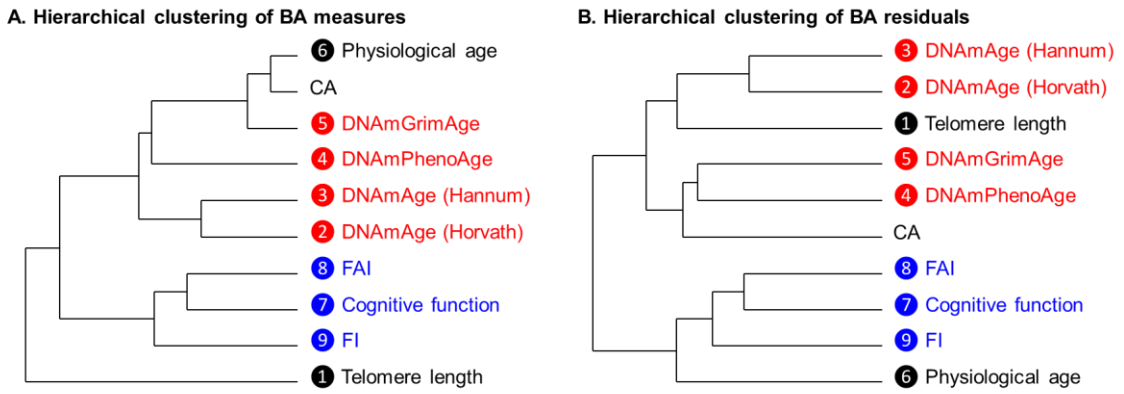

Figure 5.2.2 Hierarchical clustering of BAs in 288 individuals using 612 complete measurements. The dendrograms illustrated the hierarchical clustering according to the Euclidean distances derived from correlation coefficients. DNAmAges and functional BAs were marked in red and blue, respectively. This figure is reproduced from Li et al eLife 2020. (182)

5.2.3 BA-mortality associations

During a median follow-up time of more than 15 years, we documented 583 (69.2%) death cases among the total study population. We first analyzed the BA-mortality associations for each BA separately and found all BA, except for TL, were associated with mortality risk, although Hannum DNAmAge and physiological age showed insignificant results. An increase in cognitive function showed a protective effect, while an increase in the remaining BAs was associated with a higher death risk. Next, we took all BA residuals into consideration in the same model and observed the effects of three BAs (Horvath DNAmAge, DNAmGrimAge, and the FI) were independent of the other BAs (Table 5.2.3).

Table 5.2.3 HRs (95% CI) of a SD increase in baseline BAs with the risk of all-cause mortality

BAs	One-BA model	Nine-BA residual model
Telomere length	1.01 (0.92, 1.11)	1.03 (0.89, 1.19)
DNAmAge (Horvath)	1.17 (1.01, 1.36)	1.31 (1.08, 1.58)
DNAmAge (Hannum)	1.17 (0.98, 1.40)	1.03 (0.83, 1.28)
DNAmPhenoAge	1.26 (1.08, 1.47)	1.13 (0.91, 1.40)
DNAmGrimAge	1.39 (1.11, 1.75)	1.43 (1.11, 1.84)
Physiological age	1.13 (0.97, 1.31)	1.01 (0.87, 1.18)
Cognitive function	0.85 (0.76, 0.94)	1.01 (0.85, 1.20)
FAI	1.27 (1.10, 1.47)	1.04 (0.86, 1.27)
FI	1.32 (1.18, 1.48)	1.58 (1.32, 1.89)

This table is reproduced from Li et al eLife 2020. (182)

5.3 STUDY 3

5.3.1 Population characteristics

Of 12,098 participants, 5,469 (45.2%) were men and 9,300 (76.9%) were comprised of complete twin pairs (21.6% MZ pairs, 29.2% same-sex DZ pairs, and 26.0% opposite-sex DZ pairs). A total of 2,560 individuals had any prevalent chronic diseases of interest at baseline and therefore the remaining 9,538 participants were included in the healthspan analysis (Table 5.3.1). The most common causes leading to an end of healthspan were cancer (43.9%), diabetes (12.5%), and MI (11.5%; Table 5.3.2).

Table 5.3.1 Baseline characteristics in SATSA

	All	Men	Women
N	12098	5469	6629
Age (year)	64.9 (8.1)	65.3 (8.0)	64.6 (8.2)
Educational attainment (year)	10.8 (3.2)	10.7 (3.3)	10.8 (3.2)
BMI (kg/m2)	25.9 (3.9)	26.2 (3.4)	25.6 (4.2)
Ever-smokers (N [%])	6776 (56.0%)	3311 (60.5%)	3465 (52.3%)
Prevalent diseases at baseline (N [%])			
Cancer	1221 (10.1%)	536 (9.8%)	685 (10.3%)
Diabetes	482 (4.0%)	293 (5.4%)	189 (2.9%)
MI	686 (5.7%)	502 (9.2%)	184 (2.8%)
CHF	219 (1.8%)	133 (2.4%)	86 (1.3%)

	All	Men	Women
Stroke	348 (2.9%)	192 (3.5%)	156 (2.4%)
COPD	192 (1.6%)	86 (1.6%)	106 (1.6%)
Dementia	21 (0.2%)	15 (0.3%)	6 (0.1%)
Any prevalent chronic disease	2560 (21.2%)	1359 (24.8%)	1201 (18.1%)
Serum biomarkers at baseline (Mean [SD] or Median [IQR])			
FBG (mmol/L; median)	5.6 (1.2)	5.7 (1.3)	5.4 (1.0)
HbA1c (%; median)	4.80 (0.65)	4.84 (0.72)	4.78 (0.59)
TG (mmol/L; median)	1.2 (0.8)	1.2 (0.8)	1.1 (0.7)
TC (mmol/L)	5.8 (1.1)	5.5 (1.1)	6.0 (1.1)
HDL-C (mmol/L)	1.4 (0.4)	1.2 (0.3)	1.6 (0.4)
LDL-C (mmol/L)	3.8 (1.0)	3.7 (1.0)	3.9 (1.0)
ApoA1 (g/L)	1.6 (0.3)	1.5 (0.3)	1.7 (0.3)
ApoB (g/L)	1.08 (0.24)	1.06 (0.24)	1.09 (0.24)
CRP (mg/L; median)	1.7 (2.7)	1.6 (2.7)	1.7 (2.8)
Hb (g/L)	142.3 (11.9)	148.6 (11.0)	137.0 (9.9)

This table is reproduced from Li et al EBiomedicine 2021. (168)

Table 5.3.2 Number of events leading to an end of healthspan

	All	Men	Women
Cancer	1684 (43.9%)	836 (43.0%)	848 (44.9%)
Diabetes	479 (12.5%)	245 (12.6%)	234 (12.4%)
MI	442 (11.5%)	284 (14.6%)	158 (8.4%)
Stroke	364 (9.5%)	197 (10.1%)	167 (8.8%)
CHF	249 (6.5%)	125 (6.4%)	147 (7.8%)
COPD	231 (6.0%)	107 (5.5%)	124 (6.6%)
Dementia	230 (6.0%)	83 (4.3%)	124 (6.6%)
Death up to 2016-12-31	153 (4.0%)	65 (3.3%)	88 (4.7%)

This table is reproduced from Li et al EBiomedicine 2021. (168)

5.3.2 Serum biomarkers and healthspan

During a median follow-up time of 9.5 and 13.0 years, 3,681 any chronic event and 2,671 deaths were documented (Table 5.3.3). We found that an elevated level of serum glycemic markers (HbA1c and FBG), inflammatory marker (CRP), and TG were indicative of higher risks of both any chronic disease and all-cause death; in contrast, some lipid markers (HDL-C, ApoA1, TC, and LDL-C) were associated with lower risks for both outcomes. Specifically, increased Hb and ApoB were predictive of lower death risks, but were not associated with the end of the healthspan (Table 5.3.4).

Table 5.3.3 Outcomes documented during follow-up

	Participants in healthspan analysis	Participants in lifespan analysis
Number of individuals	9538	12098
Follow-up information (N [%] or Median [IQR])		
Follow-up time (year)	9.5 (3.4)	13.0 (1.5)
Number of incident cases	3681 (38.6%)	2674 (22.1%)
Onset age of incident cases / age at death	72.3 (11.8)	81.7 (12.1)

Table 5.3.4 Associations of serum biomarkers with healthspan and lifespan in all participants

Biomarker (SD units)	Any chronic event (end of healthspan)		Death (end of lifespan)	
	HR (95%CI)	FDR-corrected P	HR (95%CI)	FDR-corrected P
FBG	1.28 (1.23, 1.33)	5.15e-33	1.18 (1.13, 1.22)	5.45e-17
HbA1c	1.29 (1.24, 1.34)	9.07e-37	1.22 (1.17, 1.26)	6.42e-28
CRP	1.11 (1.08, 1.15)	2.21e-09	1.15 (1.10, 1.20)	8.17e-11
TG	1.07 (1.03, 1.11)	0.001	1.12 (1.07, 1.17)	1.76e-06
Hb	0.99 (0.95, 1.03)	0.791	0.89 (0.85, 0.93)	6.40e-07
ApoB	1.00 (0.96, 1.03)	0.833	0.95 (0.91, 0.99)	0.016
LDL-C	0.96 (0.93, 1.00)	0.060	0.89 (0.85, 0.93)	2.14e-07
TC	0.96 (0.92, 0.99)	0.022	0.90 (0.86, 0.94)	1.76e-06
ApoA1	0.93 (0.89, 0.96)	2.12e-04	0.91 (0.87, 0.96)	1.72e-04
HDL-C	0.92 (0.89, 0.96)	1.51e-04	0.91 (0.86, 0.96)	2.74e-04

This table is reproduced from Li et al EBiomedicine 2021. (168)

5.3.3 Biomarker PRSs and healthspan

Using PRSs, we found genetic propensity to higher serum FBG and CRP were associated with a higher risk of any chronic disease and a lower risk of death, respectively (Table 5.3.5). Particularly, the FBG PRS association was largely mediated by serum FBG level; however, the serum CRP could not explain the PRS effect, suggesting genetically determined CRP and inflammation-induced serum CRP likely represent different health conditions (Table 5.3.6).

Table 5.3.5 Associations of biomarker PRSs with healthspan and lifespan in all participants

PRS (SD units)	Any chronic event (end of healthspan)		Death (end of lifespan)	
	HR (95%CI)	FDR-corrected P	HR (95%CI)	FDR-corrected P
FBG	1.05 (1.02, 1.09)	0.050	1.00 (0.96, 1.05)	0.867
HbA1c	1.04 (1.00, 1.08)	0.255	1.01 (0.96, 1.05)	0.867
CRP	0.99 (0.96, 1.03)	0.922	0.91 (0.87, 0.95)	1.54e-04
TG	1.01 (0.98, 1.05)	0.706	1.03 (0.99, 1.08)	0.251
Hb	0.99 (0.96, 1.03)	0.922	0.98 (0.94, 1.02)	0.643
ApoB	1.02 (0.98, 1.05)	0.706	1.05 (1.01, 1.09)	0.066
LDL-C	1.02 (0.99, 1.06)	0.630	1.05 (1.01, 1.10)	0.064
TC	1.02 (0.99, 1.06)	0.630	1.05 (1.01, 1.10)	0.064

PRS (SD units)	Any chronic event (end of healthspan)		Death (end of lifespan)	
	HR (95%CI)	FDR-corrected P	HR (95%CI)	FDR-corrected P
ApoA1	1.00 (0.96, 1.03)	0.922	0.98 (0.94, 1.03)	0.643
HDL-C	1.00 (0.96, 1.03)	0.948	0.99 (0.94, 1.03)	0.643

This table is reproduced from Li et al EBiomedicine 2021. (168)

Table 5.3.6 Associations of a SD-increase in FBG PRS and CRP PRS with additional adjustment

	Additional Model 1 [1]		Additional Model 2 [2]	
	HR (95%CI)	P	HR (95%CI)	P
FBG PRS and healthspan	1.05 (1.01, 1.09)	0.008	1.01 (0.98, 1.05)	0.526
CRP PRS and death	0.90 (0.86, 0.94)	2.5e-6	0.86 (0.82, 0.90)	3.7e-10

[1] FBG PRS: Original model + adjustment for BMI and serum TG, LDL-C, HDL-C, CRP, Hb; CRP PRS: Original model + adjustment for BMI and serum TG, LDL-C, HDL-C, FBG, Hb. [2] FBG PRS: Original model + adjustment for serum FBG; CRP PRS: Original model + adjustment for serum CRP. This table is reproduced from Li et al EBiomedicine 2021. (168)

5.4 STUDY 4

5.4.1 Population characteristics

At baseline, 161,430 participants, of whom 45.1% were men, had an average age of 56.7 years, an average BMI of 27.3 kg/m^2, and a median serum CRP of 1.30 mg/L. The baseline characteristics of the discovery set were not appreciably different from the replication set (Table 5.4.1).

Table 5.4.1 Baseline characteristics of study population in UKB

	Meta-analysis set	Discovery set	Replication set	P[1]
Number of participants	161430	121072	40358	
Men (N, proportion)	72765 (45.1%)	54583 (45.1%)	18182 (45.1%)	0.92
Age (year, mean, SD)	56.7 (8.0)	56.7 (8.0)	56.6 (8.0)	0.22
High education (N, proportion)	54041 (33.8%)	40512 (33.7%)	13529 (33.8%)	0.86
High physical activity (N, %)	78072 (48.5%)	58672 (48.6%)	19400 (48.3%)	0.18
Ever-smoker (N, proportion)	99188 (61.7%)	74363 (61.6%)	24825 (61.7%)	0.72
BMI (kg/m^2, mean, SD)	27.3 (4.7)	27.3 (4.7)	27.3 (4.7)	0.88
In first 50k release (N, proportion)	39249 (24.3%)	29398 (24.3%)	9851 (24.4%)	0.61
Serum CRP				
CRP (mg/L, median, IQR)	1.30 (2.03)	1.30 (2.03)	1.30 (2.05)	0.86
Natural logarithm (mean, SD)	0.31 (1.06)	0.31 (1.05)	0.31 (1.06)	0.72

[1] P for group difference between discovery set and replication set, estimated by Fisher's exact test, t-test, and non-parametric median test whenever appropriate.

5.4.2 Rare functional genetic variants and CRP

Using the CMC burden test and a whole exome-wide significant level (P< 2.35E-06[=0.05/21,270]), we found that PA burden in the CRP gene (chromosome 1: 159712289-

159714589 on GRCh38 assembly) was significantly associated with the serum CRP concentration. Genes, other than the CRP gene, did not show significant effects (Figure 5.4.1). Carrying any PA allele in the CRP gene decreased natural log-transformed CRP by 0.676 (equivalent to a change of 0.64 SD), with a P value of 2.63E-30 estimated in the meta-analysis set. The burden test for the LOF mutation demonstrated a similar result as with PA burden; the CRP locus was again the only significant signal across the autosomal exome (beta=-0.723, P value=6.58E-11).

Of the 52 PA variants found in the CRP gene region, a majority (n=43) exhibited CRP-decreasing effects. Seven PA variants were associated with serum CRP at an FDR-corrected significance level, and all of them showed decreasing effect (Figure 5.4.2).

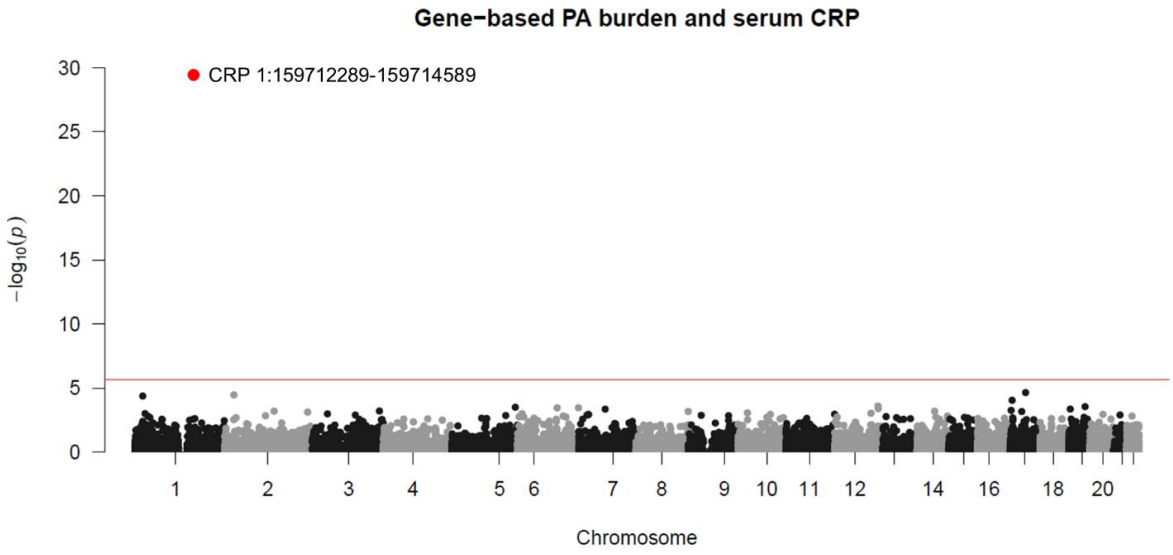

Figure 5.4.1 Associations between gene-based PA burden and serum CRP in the meta-analysis set. The Manhattan plot showed the associations between gene-based PA burden and serum CRP (natural-log transformed), estimated by the burden test in the meta-analysis set. Each dot represents a gene, of which the location is determined by the chromosomal position (x scale) and the statistic value of –log10(P) (y scale). The red horizontal line denotes the whole exome-wide significant level, i.e., -log10(2.35E-06). Annotations, including name, chromosome, and location, of the significant gene were displayed in text.

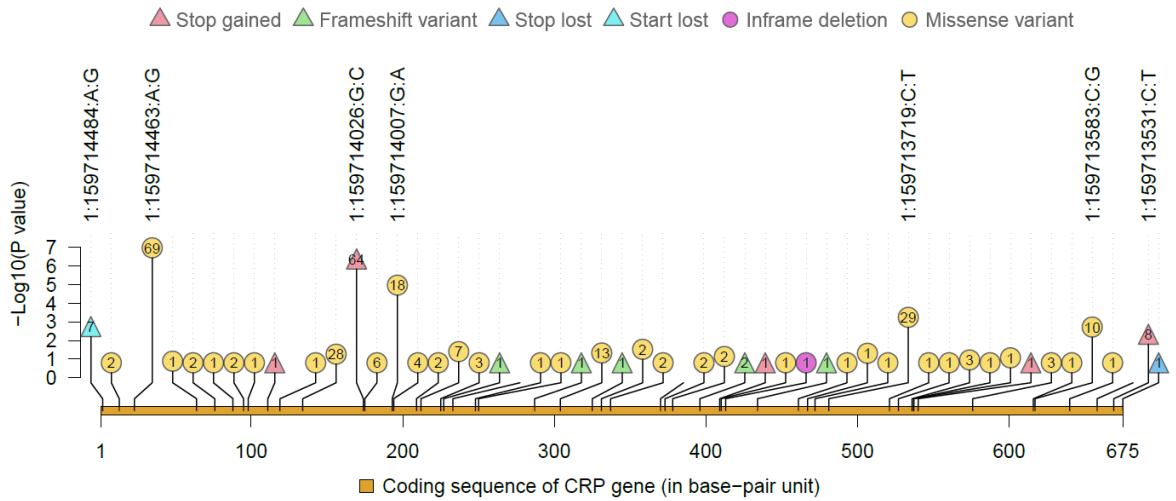

Figure 5.4.2 PA variants in the coding sequence of the CRP gene and the associations with serum CRP in the meta-analysis set. The Lolliplot shows the location of PA variants in the CDS region of the CRP gene (x scale) and the –log10(P) of the single variant-serum CRP association (y scale). LOF variants were displayed in triangles, while the remaining PA variants were displayed in circles, on top of which are the number of mutation carriers. Consequence types were denoted by different colors. Variant IDs of FDR-significant variants were displayed in text.

5.4.3 Gene-BMI interaction effect on CRP

Of the 161,430 participants, 335 individuals (134 men and 201 women) are heterozygous carriers of any PA allele in CRP locus, showing a decreased serum CRP level (median [IQR]=0.7[1.1]mg/L) compared to the non-carriers (median [IQR]=1.3[2.0]mg/L).

The serum CRP concentrations at baseline were significantly different across strata of the PA burden in the CRP gene, a previously reported common genetic variant in the CRP gene (rs1205), and a well-established environmental factor, BMI (Table 5.4.2). In particular, we observed a significant interaction effect between PA burden and obesity, with the CRP-raising effect of obesity being smaller in the PA mutation carriers (beta=0.717) than in non-carriers (beta=1.119; P for PA*obese interaction term =0.004).

Table 5.4.2 Effect of PA burden and BMI category on serum CRP (natural log transformed)

	Beta (SE)	P value [1]
Individual effect model		
Non-carriers of PA allele in CRP (reference)		
PA carriers	-0.668(0.057)	1.47e-31
rs1205 - TT (reference)		
TC	0.187(0.009)	<1e-100
CC	0.355(0.009)	<1e-100
Normal BMI (reference)		
Overweight	0.519(0.006)	<1e-100
Obese	1.118(0.006)	<1e-100
Interaction models (PA burden*BMI category)		
Non-PA carriers (normal BMI as reference)		
Overweight	0.519(0.006)	<1e-100
Obese	1.119(0.006)	<1e-100

	Beta (SE)	P value [1]
PA carriers (normal BMI as reference)		
Overweight	0.477(0.146)	1.17e-03
Obese	0.717(0.163)	1.49e-05

[1] P-value for testing the hypothesis that the corresponding Beta parameter is zero/null.

5.4.4 CRP-associated rare mutations in relation to diseases and aging phenotypes

Among the 334 PA mutation carriers, we documented a varying number of disease cases, ranging from one (Celiac disease) to 36 (Major depressive disorder) across all conditions. At a nominal significance level, carriers of PA mutation in the CRP gene had a higher risk of Crohn's disease (OR[95%CI]=3.15[1.40,7.08]) and Bipolar disorder (OR[95%CI]=3.36[1.38,8.16]) compared to non-carriers (Table 5.4.3). However, only six and five cases were observed for the two diseases among the carriers and the results were no longer considered significant after FDR corrections. Next, we tested the PA burden in association with aging-related phenotypes, including the FI, pre-frail and frail defined by the FP, and the mortality risk, and did not observe robust evidence to support a relationship.

Table 5.4.3 Associations between PA burden in CRP gene and health outcomes

Outcome	Cases among mutation carriers	Odds ratio (95%CI)	P value	FDR-corrected P
Autoimmune/inflammatory				
Celiac disease	1(0.3%)	0.45(0.06,3.22)	0.428	0.904
IBD (all types)	20(6%)	0.96(0.61,1.52)	0.877	0.994
Crohn's disease	6(1.8%)	3.15(1.40,7.08)	0.006	0.066
Ulcerative colitis	6(1.8%)	1.56(0.69,3.49)	0.285	0.890
Psoriatic arthritis	1(0.3%)	0.99(0.14,7.08)	0.994	0.994
Rheumatoid arthritis	2(0.6%)	0.25(0.06,1.01)	0.051	0.323
Type 1 diabetes	2(0.6%)	0.70(0.18,2.83)	0.621	0.983
Knee osteoarthritis	26(7.8%)	1.05(0.70,1.57)	0.819	0.994
Cardiovascular				
Coronary artery disease	29(8.7%)	1.00(0.67,1.48)	0.988	0.994
Ischemic stroke (all type)	3(0.9%)	0.57(0.18,1.77)	0.328	0.890
Metabolic				
Type 2 diabetes	23(6.9%)	1.02(0.66,1.57)	0.928	0.994
Chronic kidney disease	16(4.8%)	1.41(0.85,2.35)	0.187	0.888
Neurodegenerative				
Alzheimer disease	2(0.6%)	1.06(0.26,4.30)	0.932	0.994
Parkinson disease	3(0.9%)	1.39(0.44,4.36)	0.573	0.983
Psychiatric				
Bipolar disorder	5(1.5%)	3.36(1.38,8.16)	0.007	0.066
Major depressive disorder	36(10.7%)	1.16(0.82,1.65)	0.391	0.904
Aging-related phenotype				
FI (unit: SD) [1]	2.25(1.00)	-0.05(-0.16,0.05)	0.313	0.890
Pre-frail and frail	140(42.9%)	1.08(0.87,1.35)	0.496	0.942
All-cause death	20(6%)	0.94(0.59,1.49)	0.784	0.994

[1] Estimates for FI (unit: SD): mean (SD) and beta (95%CI)

6 DISCUSSION

6.1 BA CORRELATIONS AND BA-MORTALITY ASSOCIATIONS

Aging-related changes are intricate as they take place at multiple layers, from molecular- and cellular- levels scaling up to organ-, system-, and organism- levels. Making BA measuring even more challenging is the multi-domain feature of aging-related changes. However, in practice, a BA is often claimed when changes in some age-related layers and/or domains are quantified. As a consequence, the field is somewhat overwhelmed by the number of proposed BA measures, leading to an urgent quest for understanding their connection and utility. The first and second studies in the present thesis provide evidence in this regard from two aspects, 1) BA-BA correlations and 2) BA-mortality association.

BA correlations

First, we observed weak correlations of TL with the other types of BAs, which is in line with previous findings. (72, 82-84, 89, 90, 92, 183, 184) This weak correlation suggests that the biological process captured by TL, mitotic aging, is largely independent of the other included BAs. Second, a clear attenuation of the correlations between different types of BA was observed after we controlled CA-related change among BA residuals. In other words, among individuals of the same CA, BAs of different types are only weakly associated and have the potential to provide complementary aging-related information. This finding indicates using different types of BA has the potential to maximize aging-related information compared to using different BAs of the same type.

Therefore, given that the aging process is multi-faceted and one BA is only capable of providing limited information, to gain a global picture of the aging level, several BAs assessed at different layers and domains are suggested, rather than only relying on a single BA measure. In research, multiple lines of BA evidence are needed to evaluate the efficacy of interventions on aging; in clinical settings, multi-dimensional BA assessments would facilitate the delivery of precision medicine.

BA-mortality associations

Our first observation is related to the BA individually. As with previous findings, (40, 67, 84, 108, 112, 185, 186) we observed that all BAs, except for TL, were predictive of mortality risk individually. We did not observe evidence to support a TL-related mortality risk, while a previous meta-analysis found higher TL indicated lower mortality risk after pooling inconsistent evidence together. (105) Taken together, these results suggest DNAmAges, multi-biomarker BA, and the functional BAs could robustly identify people with higher death risks across population settings, while leukocyte TL is less robust.

Specifically, DNAmGrimAge exhibited the largest effect on mortality risk related to one SD change in BA. DNAmGrimAge is a second-generation DNAmAge and its calculation algorithm was trained explicitly on mortality risk in the US population. On the one hand, a

strong mortality association of DNAmGrimAge is expected by construction. On the other hand, the result that DNAmGrimAge showed a stronger association with death than FI, an extremely robust mortality predictor, is encouraging. That means using DNA methylation information and a proper training method could create valid predictors of complex phenotype, demonstrating a great potential of using the epigenome to assess other complex aging-related phenotypes. Therefore, researchers could continue to optimize the predictive utility of the genome-wide methylation data in relation to a wide range of aging-related phenotypes to help translate DNAmAges to more practical uses.

The second observation came from the joint effects of different BAs on mortality risk, where significant and independent associations of Horvath DNAmAge, DNAmGrimAge, and FI were found. The results suggest these BAs captured complementary information in terms of mortality prediction. It is interesting to find, among the functional BA cluster, FI is the only measure remaining significant and the effects of FAI and cognitive function attenuated almost entirely. This is likely explained by the multi-domain nature of the FI in that the FI development has to some extent taken the physical and cognitive function quantified by the FAI and cognitive ability into account already.

Therefore, given a strong individual effect of DNAmGrimAge and the FI, we suggest prioritizing DNAmGrimAge and the FI as the individual BA measure to identify subgroups with higher mortality risks among the population. Furthermore, the independent joint effect of Horvath DNAmAge, DNAmGrimAge, and the FI indicated future clinical practices and medical research could narrow down the number of BA in the risk assessment while maximizing the relevant information.

FI-mortality association

Particularly, a whole study in this thesis is devoted to the investigation of FI-mortality associations. Similar to previous evidence, (112, 187) we found a higher FI at mid-life is significantly associated with elevated risks of death due to all causes, cardiovascular diseases, and respiratory-related causes.

Notably, we found the FI effect was independent of shared familial factors. Factors shared by twins, such as genetic and childhood environmental factors, could influence both the FI level and the mortality risk. Unlike previous studies which were conducted among unrelated individuals, our study used a within-twin pair design and found that the FI effects remained after controlling for twin-constant factors. It suggests that the FI captures aspects of aging that are related to the individual's lifestyle and environmental exposures on top of the shared genetics and familial factors. Therefore, the FI could serve as a valid morality risk indicator even among people with the same family background.

Second, we observed the relative risk due to FI increase was declining gradually with age. Frailty is hypothesized to be a result of the reduced physiological reserve. (22, 85-87, 188) If younger adults, who are expected to have better resilience, show the same amount of FI

increase as older adults, then these younger adults are likely exposed to stronger stress events and could therefore face a higher chance of death. Consequently, not only the old population, but also middle-aged adults could potentially gain benefits from enrollment to frailty screening.

6.2 CLINICAL BIOMARKERS AND HEALTHSPAN

We found that cancer, diabetes, and MI are the most frequent causes leading to the end of healthspan, while COPD, dementia, and death are the least common reasons, which was largely reflective of the order of the average onset ages across different diseases. Therefore, late-onset diseases are likely to be under-represented among relatively young adults and over-represented among old adults. Future analysis could further inspect the age-dependent feature of the healthspan phenotype among large-scale populations.

We observed that seven out of ten serum clinical biomarkers are significantly associated with healthspan. Overall, the directions of the associations mostly agree with the biomarker-disease understanding, where elevated levels of glycemic markers, CRP, and triglyceride indicated higher risks of any chronic diseases, while high HDL-related markers conveyed beneficial effects. (189-195) In contrast, the protective effect of increased TC and LDL-C on any chronic diseases disagrees with some, but not all previous evidence. (195-201) The inconsistent results might be explained by the difference in population characteristics as the relationships between lipids and mortality were suggested to be age-dependent and in a dose-response manner. (202) However, the present study is limited by the statistical power to detect those patterns. Nevertheless, this suggests that beneficial effects could be indicated through higher levels of TC and LDL-C when it comes to the overall healthspan in some populations like TwinGene. Therefore, we suggest taking a range of health outcomes into consideration when making public health recommendations as the biomarker effect on a single disease might be conflicted with the evidence for the overall health outcomes.

Interestingly, we found a significant positive relationship between genetic predisposition to higher serum glucose and an increased risk of any chronic diseases, largely mediated through serum glucose level. This is putative causal evidence to link higher glucose to shorter healthspan. From an intervention perspective, glucose-lowing behaviors, including caloric restriction and exercising, as well as pharmaceutical treatment, such as glucose-lowering drugs, could be effective to maintain healthspan. It is worth noting that this relationship largely attenuated after we exclude diabetes cases in the sensitivity analysis. On the one hand, the attenuation suggested diabetes is the primary driver underlying the glucose-healthspan association. On the other hand, diabetic and pre-diabetic conditions are relatively early-onset among general populations, observed in both TwinGene and the UKB, (25) and, targeting glucose would nevertheless lead to health benefits from a population's perspective.

6.3 RARE FUNCTIONAL GENETIC VARIANTS AND SERUM CRP

We observed that all signals, except for the CRP gene, identified in the GWAS flattened to a statistically insignificant level across the exome-wide PA mutations, suggesting rare genetic

determinants were less polygenic relative to common variants. Another possible explanation is that the effects of genes other than the CRP were too small to be identified in the current analysis. Previously, a study used the UKB WES data from the first 50k release (informative sample size: 40,468) and observed that the CRP gene was the only significant signal in the burden test. (152) Even though the present analysis increased the sample size by four-fold (informative sample size: 161,430), we did not find additional signals on top of the previous test. Taken together, this observation supports a strong effect of PA burden in the CRP gene on serum CRP levels, and provides no evidence for a polygenetic architecture across the rare functional variants.

Further, we found a G*E interaction between rare variants and a well-established CRP-associated factor, BMI. The effect of obesity in elevating serum CRP was smaller in the rare mutation carriers than among the non-carriers. A similar G*E interaction feature was demonstrated for common genetic variants of CRP in a previous study, (137) albeit not in ours. Together, these observations suggest that both genetic and environmental factors need to be considered when interpreting serum CRP levels in clinical settings, especially among rare mutation carriers. Given the per-allele effect of PA variants were 4-7 times larger than the common variants in the CRP gene and the existence of G*E interaction, translating the knowledge to clinical uses would facilitate the delivery of precision medicine.

We saw suggestive evidence for increased risks of Crohn's disease and Bipolar disorder among PA mutation carriers. Crohn's disease is an inflammatory bowel disease and, in line with our finding, a nominal association has been reported before using polygenic risk scores of CRP as genetic instruments. (175, 193) Bipolar disorder is a psychiatric disease and previous MR studies have found an association in the opposite direction to our results. (143, 175) It is worth noting that we used self-reported information, hospital inpatient data, and death record to capture diagnoses. Therefore, the out-patient diagnosis could be missed and the psychiatric diagnosis captured in hospitals are more likely to be of the severe type. In addition, the number of cases among the mutation carriers is relatively small, and therefore sampling variability is a major concern. Examining the clinical relevance of rare coding mutations could reveal therapeutic insights for disease treatment and provide evidence for personalized medicine. Future investigations with a larger number of mutation carriers and longer follow-up times are warranted.

6.4 METHODOLOGICAL CONSIDERATIONS

This thesis is composed of four epidemiological studies. From a methodological point of view, three types of considerations are discussed below.

Selection bias

All analysis data in this thesis came from established population cohorts. As with many cohorts, the participants in SALT, SATSA, TwinGene, and UKB are volunteer-based, who were often more health-conscious and, therefore, not as representative of the general population. Previous studies have shown that the UKB participants were better off in terms of

socioeconomic deprivation, lifestyle behavior, and most disease conditions, compared to general UK individuals. (203, 204) Therefore, the generalizability of the absolute risks in this thesis is limited, and measures such as the prevalence of frailty, the level of BAs, the incidence of the end of healthspan, and the frequency of the mutation carriers, may not be interpreted as a representative estimate among the general population. However, the relative risks, namely the "exposure-outcome associations", would be influenced by the volunteer effects to only a limited degree, if we assume the true effects are similar among volunteers and non-volunteers. Recently, the generalizability of risk factor-mortality associations in UKB has been supported. (205)

Noteworthy is another type of selection bias introduced by survival conditioning in the exploration of genetic factors associated with serum CRP in the fourth study. As all participants joined the UKB in their adulthoods, the genetic variants that are associated with serum CRP yet are not compatible with middle-age survival would be neglected.

Misclassification

Measurements of exposure and outcomes in this thesis were ascertained through multiple lines of sources, including self-reported information, physical check-ups, sampling and arraying of bio-samples, statistical computations, and data acquisition from the health registers. Any type of measurement error would lead to misclassification of either "exposure" or "outcome".

Misclassification of the exposure can occur non-differentially or differentially. The former type means misclassification occurs randomly across groups of interest and will bias the relative risk toward the null; while the latter one denotes the probability of being misclassified depends on the factors of interest and will induce bias with varying directions and magnitudes. (206, 207) In this thesis, non-differential misclassification is likely to exist and should not change the main conclusion.

The outcome assessment influences the result interpretation. The health outcome ascertainment in this thesis heavily relies on register-based data. Particularly in the fourth study, out-patient data are not available. Diagnoses captured by the in-patient data would largely reflect the severe types. Minor conditions that only require out-patient medical services in the UKB are missed. Therefore, the interpretation of our results should be restricted to the severer types of diagnosis.

Confounding

Confounding is central to epidemiological research when causal inference is the goal. The first and second studies in this thesis emphasize the predictive ability of the BAs instead of causal relationships, because BAs by definition are indicators/proxies of aging-related changes and are not necessarily the causes of aging. Therefore, the research focus is given to the predictive value of BA independent of the traditional aging factors, such as age, sex, and BMI, instead of controlling confounders. Further, the third and fourth studies touched upon

the aim of causal inference. The third study used genetic predisposition as an instrumental variable to minimize confounding bias based on the assumption that genetic factors are determined at birth and are not associated with common confounders. (208, 209) Even though we did not observe associations between PRSs and common confounders, residual confounding might exist through a factor that is not measured in the study. The fourth study aimed to discover rare functional genetic variants that causally regulate serum CRP concentration. The nature of the rare functional genetic variants could reduce the possibility of confounding effects, since PA mutations have direct functions on amino acids based on biological knowledge. Population stratification could be a potential confounder in the genetic analysis and is adjusted for via genomic PCs. Albeit unlikely, we cannot preclude the possibility of assortative mating, meaning the serum biomarker concentrations have an influence on mate-selection, being a possible residual confounder. (210)

6.5 ETHICAL CONSIDERATIONS

The primary goal of scientific research is to generate and test hypotheses. This thesis aims to advance the knowledge of biological aging in populations. To do so, observational data collected from large-scale populations are analyzed. Multiple research steps, including data collection, data analyzing, and data publishing, could potentially give rise to risks and therefore the benefit-risk balances are worth reflecting.

First, individual data in both STR and UKB were obtained from self-reported, physical, cognitive, and bio-sample tests as well as health registers. Researchers need to inform participants of the purposes, procedures, and potential consequences of data collection and obtain participants' consents or equivalent permits before conducting the study. The analyses in this thesis were approved by the Regional Ethics Board in Stockholm (Dnr 2016/1888-31/1) and informed consent were obtained from all participants during the cohort establishment.

Second, sensitive personal data should be treated with caution in the analysis. Using sufficient data to answer the research question and protecting the participants' privacy are equally important. This thesis analyzed different types of sensitive data, such as genetic information, biomarkers, and diseases. To protect participants' privacy, the data which I could get access to is pseudonymized and unidentifiable.

Third, epidemiological researchers often put a lot of effort into data analysis and should not manipulate data in any way. With various statistical methods at our disposal, the analysis plan should be chosen based on solid scientific justifications instead of chasing small P values.

Fourth, publishing scientific articles is an essential way for researchers to share new knowledge with the field and for peer researchers to judge and/or criticize the study quality. Since epidemiologists' work usually ends up with a huge amount of analysis results, we should try to publish our findings in an unbiased manner instead of selecting results subjectively.

7 CONCLUSIONS

I. The FI was predictive of mortality risk in related individuals and middle-aged adults. Public health screenings could consider the FI as a valid morality risk indicator among family members and the middle-aged.

II. BAs of different types captured distinct aspects of mortality-related information. Making use of a set of multi-dimensional BAs could provide complementary evidence for risk assessment and intervention/treatment effect evaluation in research as well as in clinical practices.

III. Levels in circulating glycemic, lipid-, and inflammatory biomarkers were associated with the morbidity-defined healthspan. Glucose control was a putative causal mechanism and a potential intervention target for healthspan maintenance.

IV. Rare functional genetic mutations were strongly associated with serum CRP concentrations and showed no evidence for a polygenetic architecture. The observation of gene-environment interactions underscores the need to consider both genetic and environmental factors when making inferences and offering personalized medical service in clinics.

8 FUTURE PERSPECTIVE

Research on BA has accumulated at an unprecedented speed in the era of omics and big data. So far, a majority of the efforts were given to proving the validity of BAs by examining the BA-health outcome associations. Besides risk prediction, another essential utility of BAs lies in the evaluation of intervention/treatment effects, which however remains less well studied. BAs could proxy the aging process from distinct perspectives and provide continuous scales of health, compared to traditional dichotomized endpoints like disease diagnosis and death. Therefore, interrogating BAs in relation to aging-related intervention/treatment/exposure are essential topics for future studies. Two aspects are especially worth noting. First, as shown in this thesis, BAs of different types are complementary in capturing aging changes. Therefore a set of more or less independent BA rather than a single BA is suggested to assess the effect. Second, contrary to CA, BAs hold the potential to be reversible markers. To assess the reactions to intervention, large-scale longitudinal studies are needed.

In addition, the clinical use of BA could go beyond geriatric care, as recently demonstrated in COVID-related research. (211) The BA measurements in combination with domain-specific markers could allow medical specialists to provide individuals with more personalized services such as tailored suggestions on intervention and treatment. Therefore, the application of BA measures in a clinical setting is, without doubt, an important translational task for researchers to accomplish in the future.

Third, from a public health point of view, describing the changes of BAs in a representative population could keep track of the overall aging level among the population and evaluate the effectiveness of health policies. Currently, prevalence estimation and population-wide screening are mostly conducted in a disease-oriented manner, BA measures could make good candidates as a part of routinely assessed population statistics in the future.

Fourth, an assessment of genetic background could guide personalized medicine service for rare mutation carriers due to several reasons. First, biomarker variation patterns observed among the general population are no longer applicable to rare mutation carriers. Second, the existence of gene-environment interaction could interfere with the treatment/intervention effect. Third, elevated risks of health conditions could be monitored. Therefore, future investigations on rare genetic variants with larger sample sizes and in association with other clinical biomarkers are warranted to improve the delivery of precision medicine.

9 ACKNOWLEDGEMENTS

The past four years have turned out to be a journey beyond my wildest imagination, even not to mention COVID-19. I have gained so much, both research-wise and non-research-wise, and the credit goes to many brilliant people and organizations.

Primary thank goes to my main supervisor, **Sara Hägg**, for being an excellent researcher, educator, and a truly good person. This PhD program was built initially upon Sara's research interest around the "biological age" and her ability to acquire essential resources to support the work. I learned great lessons from her about not only conducting research but also communicating research as Sara is an extraordinary communicator on both academic and non-academic occasions. It is truly an honor for me to work with Sara during this PhD journey and I sincerely appreciate the time and the effort she has devoted to my education.

Equally important thanks are given to my co-supervisors. **Juulia Jylhävä** led the work of my first study and taught me a lot of "research 101" knowledge. I can't thank her enough for walking me through a solid and pleasant start. **Patrik Magnusson**, who always adds a cheerful flavor to my serious supervisory meetings, has given essential support to my STR studies and vividly demonstrated how to enjoy science in daily work. **Alexander Ploner** provided vital education and support to almost every part of my work, from computation coding, statistical inferencing, result interpreting, manuscript drafting, to revision rebutting. I will say no more about his intelligence; I do however want to thank him for always holding massive curiosity during every meeting we had, which is an incredible quality and has inspired me profoundly.

All co-authors played important roles in my studies. **Nancy Pedersen** has generously shared her invaluable insights gained through decades of research experience with almost all of my studies and I thank her sincerely. I had countless talks with **Yunzhang Wang**, **Yiqiang Zhan**, and **Xingrong Liu** to discuss the very details of my research and I truly appreciate their time and inputs. **Ida Karlsson, Chandra Reynolds, Deborah Finkel, Jonathan Mak**, and **Lu Yi** contributed to the formation and publication of my research; I could not thank their time and comments enough. I pay grateful respect to **all participants and staff in the Swedish Twin Registry and the UK Biobank**. My entire work is built on their efforts.

Special thanks are given to **Fang Fang** for being an outstanding and warm-hearted researcher, whose behaviors encourage me a lot. A special thank also goes to **Kristina Johnell** for agreeing to be the chairperson for my public defence and for providing solid leadership for MEB during the pandemic. **Ci Song** has been a fiendlike mentor to me and I have enjoyed all the talks and meals we had together.

Every colleague in the "Aging epidemiology @MEB" group is vital to keeping a vigorous scientific environment. I appreciated all the group meetings hosted by **Sara Hägg, Juulia Jylhävä, Nancy Pedersen, Yunzhang Wang**, **Yiqiang Zhan, Ida Karlsson, Jonathan Mak, Kristina Johnell, Ge Bai, KK, Karolina Kauppi, Xueying Qin, Máté Szilcz, Jonas**

Wastesson, Chenxi Qin, Laura Kananen, Adil Supiyev, Bowen Tang, Le Zhang, Peggy Ler, Malin Ericsson, Miriam Mosing, Kelli Lehto, Bojing Liu, Yasutake Tomata, Dylan Williams, Qi Wang, Xu Chen, Johanna Sieurin, Emma Raymond, Lucas Morin, and **Kathleen Bokenberger**. **Barbro Sandin** and **Camilla Palm** from the Swedish Twin Registry have provided trustful support to my STR studies. And how could I forget MEB?! **All MEBers** together have created such a synergetic system to make MEB an enjoyable place to work.

Friends give me spiritual support more than my words. I owe **Ji Zhang,** whom I have known and trusted for more than a decade, huge thanks for keeping me great company during the up and down moments in this foreign environment. **Tian Xie**, who holds the most optimistic attitude towards everything and everyone, always cheers me up and helps me see the directions when I lose my sight. Many friends from Sweden and China have made my life joyful and I am thankful to all the memories we shared together.

Despite all the joys, the past four years are not always easy and my family helps me make it toward the end. **My parents** give me unconditional love and have shaped my personality deeply. I cherish every moment I have spent and will spend with them. **Qifei He** has been bringing joy to my life over the years. Our communications have advanced my understanding of the philosophy of life and I can't wait to experience more of our life adventures.

Last but not least I sincerely appreciate are the funding bodies. **China Scholarship Council** provides the major financial support to my study and the funding granted to Sara Hägg and to the Swedish Twin Registry, including **the Strategic Research Program in Epidemiology and the Foundation for Geriatric Diseases from KI, the Swedish Research Council, the Swedish Research Council for Health, Working Life and Welfare, National Institutes of Health, and MacArthur Foundation**, all contribute to my research. **The Swedish National Graduate School for Competitive Science on Ageing and Health (SWEAH) funded by the Swedish Research Council** offers an annual allowance to my study activity and vastly increases my communication with PhD students outside of MEB and KI.

Chances are my relationships with the people aforementioned would not stop here. Look forward to the next adventure.

10 REFERENCES

1. United Nations, Department of Economic and Social Affairs, Population Division. World Population Prospects 2019: Highlights. 2019.
2. Galliot B. Hydra, a fruitful model system for 270 years. Int J Dev Biol. 2012;56(6-8):411-23.
3. Williams GC. Pleiotropy, Natural-Selection, and the Evolution of Senescence. Evolution. 1957;11(4):398-411.
4. Medawar PB. An unsolved problem of biology. London,: Published for the college by H. K. Lewis; 1952. 24 p. p.
5. Zella S. Handbook of Theories of Ageing (Third Edited Edition). J Popul Ageing. 2017;10(3):311-4.
6. Gladyshev VN. Aging: progressive decline in fitness due to the rising deleteriome adjusted by genetic, environmental, and stochastic processes. Aging Cell. 2016;15(4):594-602.
7. Kirkwood TB. Evolution of ageing. Nature. 1977;270(5635):301-4.
8. Khan SS, Singer BD, Vaughan DE. Molecular and physiological manifestations and measurement of aging in humans. Aging Cell. 2017;16(4):624-33.
9. Hayflick L. Handbook of the biology of aging, 5th edition. Gerontologist. 2002;42(3):416-21.
10. Kirkwood TB. Understanding the odd science of aging. Cell. 2005;120(4):437-47.
11. Pincus Z. Ageing: A stretch in time. Nature. 2016;530(7588):37-8.
12. Niccoli T, Partridge L. Ageing as a risk factor for disease. Curr Biol. 2012;22(17):R741-52.
13. Lopez-Otin C, Blasco MA, Partridge L, Serrano M, Kroemer G. The hallmarks of aging. Cell. 2013;153(6):1194-217.
14. Kennedy BK, Berger SL, Brunet A, Campisi J, Cuervo AM, Epel ES, et al. Geroscience: linking aging to chronic disease. Cell. 2014;159(4):709-13.
15. Sierra F. The Emergence of Geroscience as an Interdisciplinary Approach to the Enhancement of Health Span and Life Span. Cold Spring Harb Perspect Med. 2016;6(4):a025163.
16. Kubben N, Misteli T. Shared molecular and cellular mechanisms of premature ageing and ageing-associated diseases. Nat Rev Mol Cell Biol. 2017;18(10):595-609.
17. Menken J, Trussell J, Larsen U. Age and infertility. Science. 1986;233(4771):1389-94.
18. te Velde ER, Pearson PL. The variability of female reproductive ageing. Hum Reprod Update. 2002;8(2):141-54.
19. Rahmioglu N, Andrew T, Cherkas L, Surdulescu G, Swaminathan R, Spector T, et al. Epidemiology and genetic epidemiology of the liver function test proteins. PLoS One. 2009;4(2):e4435.
20. Spirduso WW. Physical dimensions of aging. Champaign, IL: Human Kinetics; 1995. xiii, 432 p. p.
21. Harada CN, Natelson Love MC, Triebel KL. Normal cognitive aging. Clin Geriatr Med. 2013;29(4):737-52.
22. Clegg A, Young J, Iliffe S, Rikkert MO, Rockwood K. Frailty in elderly people. Lancet. 2013;381(9868):752-62.
23. Franceschi C, Garagnani P, Morsiani C, Conte M, Santoro A, Grignolio A, et al. The Continuum of Aging and Age-Related Diseases: Common Mechanisms but Different Rates. Front Med (Lausanne). 2018;5:61.

24. Atella V, Piano Mortari A, Kopinska J, Belotti F, Lapi F, Cricelli C, et al. Trends in age-related disease burden and healthcare utilization. Aging Cell. 2019;18(1):e12861.
25. Zenin A, Tsepilov Y, Sharapov S, Getmantsev E, Menshikov LI, Fedichev PO, et al. Identification of 12 genetic loci associated with human healthspan. Commun Biol. 2019;2:41.
26. Ritchie K, Kildea D. Is Senile Dementia Age-Related or Aging-Related - Evidence from Metaanalysis of Dementia Prevalence in the Oldest-Old. Lancet. 1995;346(8980):931-4.
27. Ailshire JA, Beltran-Sanchez H, Crimmins EM. Becoming Centenarians: Disease and Functioning Trajectories of Older US Adults as They Survive to 100. J Gerontol a-Biol. 2015;70(2):193-201.
28. Kochanek KD, Xu J, Arias E. Mortality in the United States, 2019. NCHS Data Brief. 2020(395):1-8.
29. Xu J, Murphy SL, Kockanek KD, Arias E. Mortality in the United States, 2018. NCHS Data Brief. 2020(355):1-8.
30. Butler RN, Sprott R, Warner H, Bland J, Feuers R, Forster M, et al. Biomarkers of aging: from primitive organisms to humans. J Gerontol A Biol Sci Med Sci. 2004;59(6):B560-7.
31. Johnson TE. Recent results: biomarkers of aging. Exp Gerontol. 2006;41(12):1243-6.
32. Anstey KJ, Lord SR, Smith GA. Measuring human functional age: a review of empirical findings. Exp Aging Res. 1996;22(3):245-66.
33. Li WG, Vijg J. Measuring Genome Instability in Aging - A Mini-Review. Gerontology. 2012;58(2):129-38.
34. Montpetit AJ, Alhareeri AA, Montpetit M, Starkweather AR, Elmore LW, Filler K, et al. Telomere length: a review of methods for measurement. Nurs Res. 2014;63(4):289-99.
35. Horvath S, Raj K. DNA methylation-based biomarkers and the epigenetic clock theory of ageing. Nat Rev Genet. 2018;19(6):371-84.
36. Reynolds CA, Finkel D, McArdle JJ, Gatz M, Berg S, Pedersen NL. Quantitative genetic analysis of latent growth curve models of cognitive abilities in adulthood. Dev Psychol. 2005;41(1):3-16.
37. Sorlie T, Sexton HC, Busund R, Sorlie D. A global measure of physical functioning: psychometric properties. Health Serv Res. 2001;36(6 Pt 1):1109-24.
38. Searle SD, Mitnitski A, Gahbauer EA, Gill TM, Rockwood K. A standard procedure for creating a frailty index. BMC Geriatr. 2008;8:24.
39. Harrison C, Fortin M, van den Akker M, Mair F, Calderon-Larranaga A, Boland F, et al. Comorbidity versus multimorbidity: Why it matters. J Comorb. 2021;11:2633556521993993.
40. Levine ME. Modeling the rate of senescence: can estimated biological age predict mortality more accurately than chronological age? J Gerontol A Biol Sci Med Sci. 2013;68(6):667-74.
41. Ahadi S, Zhou W, Schussler-Fiorenza Rose SM, Sailani MR, Contrepois K, Avina M, et al. Personal aging markers and ageotypes revealed by deep longitudinal profiling. Nat Med. 2020;26(1):83-90.
42. Jylhava J, Pedersen NL, Hagg S. Biological Age Predictors. EBioMedicine. 2017;21:29-36.
43. Solovev I, Shaposhnikov M, Moskalev A. Multi-omics approaches to human biological age estimation. Mech Ageing Dev. 2020;185:111192.
44. Wagner KH, Cameron-Smith D, Wessner B, Franzke B. Biomarkers of Aging: From Function to Molecular Biology. Nutrients. 2016;8(6).

45. Xia X, Chen W, McDermott J, Han JJ. Molecular and phenotypic biomarkers of aging. F1000Res. 2017;6:860.
46. Ferrucci L, Gonzalez-Freire M, Fabbri E, Simonsick E, Tanaka T, Moore Z, et al. Measuring biological aging in humans: A quest. Aging Cell. 2019:e13080.
47. Jia L, Zhang W, Chen X. Common methods of biological age estimation. Clin Interv Aging. 2017;12:759-72.
48. Ferrucci L, Levine ME, Kuo PL, Simonsick EM. Time and the Metrics of Aging. Circ Res. 2018;123(7):740-4.
49. Mamoshina P, Kochetov K, Putin E, Cortese F, Aliper A, Lee WS, et al. Population specific biomarkers of human aging: a big data study using South Korean, Canadian and Eastern European patient populations. J Gerontol A Biol Sci Med Sci. 2018.
50. Guerville F, De Souto Barreto P, Ader I, Andrieu S, Casteilla L, Dray C, et al. Revisiting the Hallmarks of Aging to Identify Markers of Biological Age. J Prev Alzheimers Dis. 2020;7(1):56-64.
51. Kudryashova KS, Burka K, Kulaga AY, Vorobyeva NS, Kennedy BK. Aging Biomarkers: From Functional Tests to Multi-Omics Approaches. Proteomics. 2020;20(5-6):e1900408.
52. Kuo PL, Schrack JA, Shardell MD, Levine M, Moore AZ, An Y, et al. A roadmap to build a phenotypic metric of ageing: insights from the Baltimore Longitudinal Study of Aging. J Intern Med. 2020;287(4):373-94.
53. Xia X, Wang Y, Yu Z, Chen J, Han JJ. Assessing the rate of aging to monitor aging itself. Ageing Res Rev. 2021;69:101350.
54. Kristic J, Vuckovic F, Menni C, Klaric L, Keser T, Beceheli I, et al. Glycans are a novel biomarker of chronological and biological ages. J Gerontol A Biol Sci Med Sci. 2014;69(7):779-89.
55. Gonzalez-Freire M, de Cabo R, Bernier M, Sollott SJ, Fabbri E, Navas P, et al. Reconsidering the Role of Mitochondria in Aging. J Gerontol A Biol Sci Med Sci. 2015;70(11):1334-42.
56. Noren Hooten N, Evans MK. Techniques to Induce and Quantify Cellular Senescence. J Vis Exp. 2017(123).
57. Bartke A, Chandrashekar V, Dominici F, Turyn D, Kinney B, Steger R, et al. Insulin-like growth factor 1 (IGF-1) and aging: controversies and new insights. Biogerontology. 2003;4(1):1-8.
58. Horvath S. DNA methylation age of human tissues and cell types. Genome Biol. 2013;14(10):R115.
59. Ross R, Blair SN, Arena R, Church TS, Despres JP, Franklin BA, et al. Importance of Assessing Cardiorespiratory Fitness in Clinical Practice: A Case for Fitness as a Clinical Vital Sign: A Scientific Statement From the American Heart Association. Circulation. 2016;134(24):e653-e99.
60. Valderas JM, Starfield B, Sibbald B, Salisbury C, Roland M. Defining Comorbidity: Implications for Understanding Health and Health Services. Ann Fam Med. 2009;7(4):357-63.
61. Johnston MC, Crilly M, Black C, Prescott GJ, Mercer SW. Defining and measuring multimorbidity: a systematic review of systematic reviews. Eur J Public Health. 2019;29(1):182-9.
62. Olshansky SJ. From Lifespan to Healthspan. JAMA. 2018;320(13):1323-4.
63. Peters MJ, Joehanes R, Pilling LC, Schurmann C, Conneely KN, Powell J, et al. The transcriptional landscape of age in human peripheral blood. Nat Commun. 2015;6:8570.
64. Tanaka T, Biancotto A, Moaddel R, Moore AZ, Gonzalez-Freire M, Aon MA, et al. Plasma proteomic signature of age in healthy humans. Aging Cell. 2018;17(5):e12799.

65. Hertel J, Friedrich N, Wittfeld K, Pietzner M, Budde K, Van der Auwera S, et al. Measuring Biological Age via Metabonomics: The Metabolic Age Score. J Proteome Res. 2016;15(2):400-10.
66. Fried LP, Tangen CM, Walston J, Newman AB, Hirsch C, Gottdiener J, et al. Frailty in older adults: Evidence for a phenotype. J Gerontol a-Biol. 2001;56(3):M146-M56.
67. Levine ME, Lu AT, Quach A, Chen BH, Assimes TL, Bandinelli S, et al. An epigenetic biomarker of aging for lifespan and healthspan. Aging (Albany NY). 2018;10(4):573-91.
68. Greider CW. Telomeres and senescence: the history, the experiment, the future. Curr Biol. 1998;8(5):R178-81.
69. Aubert G, Lansdorp PM. Telomeres and aging. Physiol Rev. 2008;88(2):557-79.
70. Muezzinler A, Zaineddin AK, Brenner H. A systematic review of leukocyte telomere length and age in adults. Ageing Res Rev. 2013;12(2):509-19.
71. Chen W, Kimura M, Kim S, Cao X, Srinivasan SR, Berenson GS, et al. Longitudinal versus cross-sectional evaluations of leukocyte telomere length dynamics: age-dependent telomere shortening is the rule. J Gerontol A Biol Sci Med Sci. 2011;66(3):312-9.
72. Marioni RE, Harris SE, Shah S, McRae AF, von Zglinicki T, Martin-Ruiz C, et al. The epigenetic clock and telomere length are independently associated with chronological age and mortality. Int J Epidemiol. 2018;45(2):424-32.
73. Berglund K, Reynolds CA, Ploner A, Gerritsen L, Hovatta I, Pedersen NL, et al. Longitudinal decline of leukocyte telomere length in old age and the association with sex and genetic risk. Aging-Us. 2016;8(7):1398-415.
74. Dupont C, Armant DR, Brenner CA. Epigenetics: definition, mechanisms and clinical perspective. Semin Reprod Med. 2009;27(5):351-7.
75. Xiao FH, Kong QP, Perry B, He YH. Progress on the role of DNA methylation in aging and longevity. Brief Funct Genomics. 2016;15(6):454-9.
76. Rakyan VK, Down TA, Maslau S, Andrew T, Yang TP, Beyan H, et al. Human aging-associated DNA hypermethylation occurs preferentially at bivalent chromatin domains. Genome Res. 2010;20(4):434-9.
77. Hannum G, Guinney J, Zhao L, Zhang L, Hughes G, Sadda S, et al. Genome-wide methylation profiles reveal quantitative views of human aging rates. Mol Cell. 2013;49(2):359-67.
78. Lu AT, Quach A, Wilson JG, Reiner AP, Aviv A, Raj K, et al. DNA methylation GrimAge strongly predicts lifespan and healthspan. Aging (Albany NY). 2019;11(2):303-27.
79. Noroozi R, Ghafouri-Fard S, Pisarek A, Rudnicka J, Spolnicka M, Branicki W, et al. DNA methylation-based age clocks: From age prediction to age reversion. Ageing Res Rev. 2021;68:101314.
80. Simpson DJ, Chandra T. Epigenetic age prediction. Aging Cell. 2021:e13452.
81. Klemera P, Doubal S. A new approach to the concept and computation of biological age. Mech Ageing Dev. 2006;127(3):240-8.
82. Belsky DW, Moffitt TE, Cohen AA, Corcoran DL, Levine ME, Prinz JA, et al. Eleven Telomere, Epigenetic Clock, and Biomarker-Composite Quantifications of Biological Aging: Do They Measure the Same Thing? Am J Epidemiol. 2018;187(6):1220-30.
83. Zhong X, Lu Y, Gao Q, Nyunt MSZ, Fulop T, Monterola CP, et al. Estimating Biological Age in the Singapore Longitudinal Aging Study. J Gerontol A Biol Sci Med Sci. 2020;75(10):1913-20.

84. Chan MS, Arnold M, Offer A, Hammami I, Mafham M, Armitage J, et al. A Biomarker-based Biological Age in UK Biobank: Composition and Prediction of Mortality and Hospital Admissions. J Gerontol A Biol Sci Med Sci. 2021;76(7):1295-302.
85. Rockwood K, Mitnitski A. Frailty in relation to the accumulation of deficits. J Gerontol A Biol Sci Med Sci. 2007;62(7):722-7.
86. Mitnitski AB, Mogilner AJ, Rockwood K. Accumulation of deficits as a proxy measure of aging. ScientificWorldJournal. 2001;1:323-36.
87. Rockwood K, Mitnitski A. Frailty defined by deficit accumulation and geriatric medicine defined by frailty. Clin Geriatr Med. 2011;27(1):17-26.
88. Welstead M, Jenkins ND, Russ T, Luciano M, Muniz-Terrera G. A Systematic Review of Frailty Trajectories: Their Shape And Influencing Factors. Gerontologist. 2020.
89. Chen BH, Carty CL, Kimura M, Kark JD, Chen W, Li S, et al. Leukocyte telomere length, T cell composition and DNA methylation age. Aging (Albany NY). 2017;9(9):1983-95.
90. Banszerus VL, Vetter VM, Salewsky B, Konig M, Demuth I. Exploring the Relationship of Relative Telomere Length and the Epigenetic Clock in the LipidCardio Cohort. International Journal of Molecular Sciences. 2019;20(12).
91. Vetter VM, Meyer A, Karbasiyan M, Steinhagen-Thiessen E, Hopfenmuller W, Demuth I. Epigenetic clock and relative telomere length represent largely different aspects of aging in the Berlin Aging Study II (BASE-II). J Gerontol A Biol Sci Med Sci. 2018.
92. Hastings WJ, Shalev I, Belsky DW. Comparability of biological aging measures in the National Health and Nutrition Examination Study, 1999-2002. Psychoneuroendocrinology. 2019;106:171-8.
93. Araujo Carvalho AC, Tavares Mendes ML, da Silva Reis MC, Santos VS, Tanajura DM, Martins-Filho PRS. Telomere length and frailty in older adults-A systematic review and meta-analysis. Ageing Res Rev. 2019;54:100914.
94. Marioni RE, Suderman M, Chen BH, Horvath S, Bandinelli S, Morris T, et al. Tracking the Epigenetic Clock Across the Human Life Course: A Meta-analysis of Longitudinal Cohort Data. J Gerontol A Biol Sci Med Sci. 2019;74(1):57-61.
95. Grodstein F, Lemos B, Yu L, Iatrou A, De Jager PL, Bennett DA. Characteristics of Epigenetic Clocks Across Blood and Brain Tissue in Older Women and Men. Front Neurosci. 2020;14:555307.
96. Breitling LP, Saum KU, Perna L, Schottker B, Holleczek B, Brenner H. Frailty is associated with the epigenetic clock but not with telomere length in a German cohort. Clinical Epigenetics. 2016;8.
97. Gale CR, Marioni RE, Harris SE, Starr JM, Deary IJ. DNA methylation and the epigenetic clock in relation to physical frailty in older people: the Lothian Birth Cohort 1936. Clin Epigenetics. 2018;10(1):101.
98. Rockwood K, Andrew M, Mitnitski A. A comparison of two approaches to measuring frailty in elderly people. J Gerontol A Biol Sci Med Sci. 2007;62(7):738-43.
99. Cesari M, Gambassi G, van Kan GA, Vellas B. The frailty phenotype and the frailty index: different instruments for different purposes. Age Ageing. 2014;43(1):10-2.
100. Blodgett J, Theou O, Kirkland S, Andreou P, Rockwood K. Frailty in NHANES: Comparing the frailty index and phenotype. Arch Gerontol Geriatr. 2015;60(3):464-70.
101. Thompson MQ, Theou O, Yu S, Adams RJ, Tucker GR, Visvanathan R. Frailty prevalence and factors associated with the Frailty Phenotype and Frailty Index: Findings from the North West Adelaide Health Study. Australas J Ageing. 2018;37(2):120-6.

102. Houben JM, Giltay EJ, Rius-Ottenheim N, Hageman GJ, Kromhout D. Telomere length and mortality in elderly men: the Zutphen Elderly Study. J Gerontol A Biol Sci Med Sci. 2011;66(1):38-44.
103. Honig LS, Kang MS, Schupf N, Lee JH, Mayeux R. Association of shorter leukocyte telomere repeat length with dementia and mortality. Arch Neurol. 2012;69(10):1332-9.
104. Glei DA, Goldman N, Weinstein M, Risques RA. Shorter Ends, Faster End? Leukocyte Telomere Length and Mortality Among Older Taiwanese. J Gerontol A Biol Sci Med Sci. 2015;70(12):1490-8.
105. Wang Q, Zhan Y, Pedersen NL, Fang F, Hagg S. Telomere Length and All-Cause Mortality: A Meta-analysis. Ageing Res Rev. 2018;48:11-20.
106. Field AE, Robertson NA, Wang T, Havas A, Ideker T, Adams PD. DNA Methylation Clocks in Aging: Categories, Causes, and Consequences. Mol Cell. 2018;71(6):882-95.
107. McCartney DL, Hillary RF, Stevenson AJ, Ritchie SJ, Walker RM, Zhang Q, et al. Epigenetic prediction of complex traits and death. Genome Biol. 2018;19(1):136.
108. Zhang Y, Wilson R, Heiss J, Breitling LP, Saum KU, Schottker B, et al. DNA methylation signatures in peripheral blood strongly predict all-cause mortality. Nat Commun. 2017;8:14617.
109. Mitnitski A, Howlett SE, Rockwood K. Heterogeneity of Human Aging and Its Assessment. J Gerontol a-Biol. 2017;72(7):877-84.
110. Yoo J, Kim Y, Cho ER, Jee SH. Biological age as a useful index to predict seventeen-year survival and mortality in Koreans. Bmc Geriatrics. 2017;17.
111. Liu Z, Kuo PL, Horvath S, Crimmins E, Ferrucci L, Levine M. A new aging measure captures morbidity and mortality risk across diverse subpopulations from NHANES IV: A cohort study. PLoS Med. 2018;15(12):e1002718.
112. Kojima G, Iliffe S, Walters K. Frailty index as a predictor of mortality: a systematic review and meta-analysis. Age Ageing. 2017:1-8.
113. Kim S, Myers L, Wyckoff J, Cherry KE, Jazwinski SM. The frailty index outperforms DNA methylation age and its derivatives as an indicator of biological age. Geroscience. 2017;39(1):83-92.
114. Murabito JM, Zhao Q, Larson MG, Rong J, Lin H, Benjamin EJ, et al. Measures of Biologic Age in a Community Sample Predict Mortality and Age-Related Disease: The Framingham Offspring Study. J Gerontol A Biol Sci Med Sci. 2018;73(6):757-62.
115. Zhang Y, Saum KU, Schottker B, Holleczek B, Brenner H. Methylomic survival predictors, frailty, and mortality. Aging (Albany NY). 2018.
116. Organization WH. Healthy life expectancy (HALE) at birth (years) 2021 [Available from: https://www.who.int/data/gho/data/indicators/indicator-details/GHO/gho-ghe-hale-healthy-life-expectancy-at-birth.
117. World Health Organization, Department of Information, Evidence and Research. WHO methods and data sources for life tables 1990-2016. 2018.
118. Crimmins EM. Lifespan and Healthspan: Past, Present, and Promise. Gerontologist. 2015;55(6):901-11.
119. Kaeberlein M. How healthy is the healthspan concept? Geroscience. 2018;40(4):361-4.
120. Fuellen G, Jansen L, Cohen AA, Luyten W, Gogol M, Simm A, et al. Health and Aging: Unifying Concepts, Scores, Biomarkers and Pathways. Aging Dis. 2019;10(4):883-900.
121. Timmers P, Wilson JF, Joshi PK, Deelen J. Multivariate genomic scan implicates novel loci and haem metabolism in human ageing. Nat Commun. 2020;11(1):3570.

122. Terry DF, Pencina MJ, Vasan RS, Murabito JM, Wolf PA, Hayes MK, et al. Cardiovascular risk factors predictive for survival and morbidity-free survival in the oldest-old Framingham Heart Study participants. J Am Geriatr Soc. 2005;53(11):1944-50.
123. Newson RS, Witteman JCM, Franco OH, Stricker BHC, Breteler MMB, Hofman A, et al. Predicting survival and morbidity-free survival to very old age. Age. 2010;32(4):521-34.
124. Kluft C, de Maat MP. Genetics of C-reactive protein: new possibilities and complications. Arterioscler Thromb Vasc Biol. 2003;23(11):1956-9.
125. Pepys MB, Hirschfield GM. C-reactive protein: a critical update. J Clin Invest. 2003;111(12):1805-12.
126. Knight ML. The Application of High-Sensitivity C-Reactive Protein in Clinical Practice A 2015 Update. Us Pharm. 2015;40(2):50-3.
127. Sas AA, Vaez A, Jamshidi Y, Nolte IM, Kamali Z, T DS, et al. Genetic and environmental influences on stability and change in baseline levels of C-reactive protein: A longitudinal twin study. Atherosclerosis. 2017;265:172-8.
128. Pankow JS, Folsom AR, Cushman M, Borecki IB, Hopkins PN, Eckfeldt JH, et al. Familial and genetic determinants of systemic markers of inflammation: the NHLBI family heart study. Atherosclerosis. 2001;154(3):681-9.
129. Vickers MA, Green FR, Terry C, Mayosi BM, Julier C, Lathrop M, et al. Genotype at a promoter polymorphism of the interleukin-6 gene is associated with baseline levels of plasma C-reactive protein. Cardiovascular Research. 2002;53(4):1029-34.
130. Austin MA, Zhang C, Humphries SE, Chandler WL, Talmud PJ, Edwards KL, et al. Heritability of C-reactive protein and association with apolipoprotein E genotypes in Japanese Americans. Ann Hum Genet. 2004;68(Pt 3):179-88.
131. Fox ER, Benjamin EJ, Sarpong DF, Rotimi CN, Wilson JG, Steffes MW, et al. Epidemiology, heritability, and genetic linkage of C-reactive protein in African Americans (from the Jackson Heart Study). American Journal of Cardiology. 2008;102(7):835-41.
132. Schnabel RB, Lunetta KL, Larson MG, Dupuis J, Lipinska I, Rong J, et al. The Relation of Genetic and Environmental Factors to Systemic Inflammatory Biomarker Concentrations. Circ-Cardiovasc Gene. 2009;2(3):229-U102.
133. Hage FG, Szalai AJ. C-reactive protein gene polymorphisms, C-reactive protein blood levels, and cardiovascular disease risk. J Am Coll Cardiol. 2007;50(12):1115-22.
134. Melzer D, Perry JR, Hernandez D, Corsi AM, Stevens K, Rafferty I, et al. A genome-wide association study identifies protein quantitative trait loci (pQTLs). Plos Genet. 2008;4(5):e1000072.
135. Reiner AP, Barber MJ, Guan Y, Ridker PM, Lange LA, Chasman DI, et al. Polymorphisms of the HNF1A gene encoding hepatocyte nuclear factor-1 alpha are associated with C-reactive protein. American Journal of Human Genetics. 2008;82(5):1193-201.
136. Ridker PM, Pare G, Parker A, Zee RY, Danik JS, Buring JE, et al. Loci related to metabolic-syndrome pathways including LEPR,HNF1A, IL6R, and GCKR associate with plasma C-reactive protein: the Women's Genome Health Study. Am J Hum Genet. 2008;82(5):1185-92.
137. Dehghan A, Dupuis J, Barbalic M, Bis JC, Eiriksdottir G, Lu C, et al. Meta-Analysis of Genome-Wide Association Studies in > 80 000 Subjects Identifies Multiple Loci for C-Reactive Protein Levels. Circulation. 2011;123(7):731-U151.
138. Okada Y, Takahashi A, Ohmiya H, Kumasaka N, Kamatani Y, Hosono N, et al. Genome-wide association study for C-reactive protein levels identified pleiotropic associations in the IL6 locus. Hum Mol Genet. 2011;20(6):1224-31.

139. Reiner AP, Beleza S, Franceschini N, Auer PL, Robinson JG, Kooperberg C, et al. Genome-wide association and population genetic analysis of C-reactive protein in African American and Hispanic American women. Am J Hum Genet. 2012;91(3):502-12.
140. Wu Y, McDade TW, Kuzawa CW, Borja J, Li Y, Adair LS, et al. Genome-wide Association with C-Reactive Protein Levels in CLHNS: Evidence for the CRP and HNF1A Loci and their Interaction with Exposure to a Pathogenic Environment. Inflammation. 2012;35(2):574-83.
141. Williams SR, Hsu FC, Keene KL, Chen WM, Nelson S, Southerland AM, et al. Shared genetic susceptibility of vascular-related biomarkers with ischemic and recurrent stroke. Neurology. 2016;86(4):351-9.
142. Kocarnik JM, Richard M, Graff M, Haessler J, Bien S, Carlson C, et al. Discovery, fine-mapping, and conditional analyses of genetic variants associated with C-reactive protein in multiethnic populations using the Metabochip in the Population Architecture using Genomics and Epidemiology (PAGE) study. Human Molecular Genetics. 2018;27(16):2940-53.
143. Ligthart S, Vaez A, Vosa U, Stathopoulou MG, de Vries PS, Prins BP, et al. Genome Analyses of >200,000 Individuals Identify 58 Loci for Chronic Inflammation and Highlight Pathways that Link Inflammation and Complex Disorders. Am J Hum Genet. 2018;103(5):691-706.
144. Wojcik GL, Graff M, Nishimura KK, Tao R, Haessler J, Gignoux CR, et al. Genetic analyses of diverse populations improves discovery for complex traits. Nature. 2019;570(7762):514-8.
145. Raffield LM, Iyengar AK, Wang BQ, Gaynor SM, Spracklen CN, Zhong X, et al. Allelic Heterogeneity at the CRP Locus Identified by Whole-Genome Sequencing in Multi-ancestry Cohorts. American Journal of Human Genetics. 2020;106(1):112-20.
146. Sinnott-Armstrong N, Tanigawa Y, Amar D, Mars N, Benner C, Aguirre M, et al. Genetics of 35 blood and urine biomarkers in the UK Biobank. Nat Genet. 2021;53(2):185-94.
147. Momozawa Y, Mizukami K. Unique roles of rare variants in the genetics of complex diseases in humans. J Hum Genet. 2021;66(1):11-23.
148. O'Connor LJ, Schoech AP, Hormozdiari F, Gazal S, Patterson N, Price AL. Extreme Polygenicity of Complex Traits Is Explained by Negative Selection. American Journal of Human Genetics. 2019;105(3):456-76.
149. Lee S, Abecasis GR, Boehnke M, Lin X. Rare-variant association analysis: study designs and statistical tests. Am J Hum Genet. 2014;95(1):5-23.
150. Povysil G, Petrovski S, Hostyk J, Aggarwal V, Allen AS, Goldstein DB. Rare-variant collapsing analyses for complex traits: guidelines and applications. Nat Rev Genet. 2019;20(12):747-59.
151. Schick UM, Auer PL, Bis JC, Lin H, Wei P, Pankratz N, et al. Association of exome sequences with plasma C-reactive protein levels in >9000 participants. Hum Mol Genet. 2015;24(2):559-71.
152. Cirulli ET, White S, Read RW, Elhanan G, Metcalf WJ, Tanudjaja F, et al. Genome-wide rare variant analysis for thousands of phenotypes in over 70,000 exomes from two cohorts. Nat Commun. 2020;11(1):542.
153. Franceschi C, Garagnani P, Parini P, Giuliani C, Santoro A. Inflammaging: a new immune-metabolic viewpoint for age-related diseases. Nat Rev Endocrinol. 2018.
154. Lichtenstein P, De Faire U, Floderus B, Svartengren M, Svedberg P, Pedersen NL. The Swedish Twin Registry: a unique resource for clinical, epidemiological and genetic studies. J Intern Med. 2002;252(3):184-205.
155. Pedersen NL, Lichtenstein P, Svedberg P. The Swedish Twin Registry in the third millennium. Twin Res. 2002;5(5):427-32.

156. Lichtenstein P, Sullivan PF, Cnattingius S, Gatz M, Johansson S, Carlstrom E, et al. The Swedish Twin Registry in the third millennium: an update. Twin Res Hum Genet. 2006;9(6):875-82.
157. Magnusson PK, Almqvist C, Rahman I, Ganna A, Viktorin A, Walum H, et al. The Swedish Twin Registry: establishment of a biobank and other recent developments. Twin Res Hum Genet. 2013;16(1):317-29.
158. Zagai U, Lichtenstein P, Pedersen NL, Magnusson PKE. The Swedish Twin Registry: Content and Management as a Research Infrastructure. Twin Research and Human Genetics. 2019;22(6):672-80.
159. Finkel D, Pedersen NL. Processing speed and longitudinal trajectories of change for cognitive abilities: The Swedish Adoption/Twin Study of Aging. Aging Neuropsychol C. 2004;11(2-3):325-45.
160. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med. 2015;12(3):e1001779.
161. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. Nature. 2018;562(7726):203-9.
162. Van Hout CV, Tachmazidou I, Backman JD, Hoffman JD, Liu D, Pandey AK, et al. Exome sequencing and characterization of 49,960 individuals in the UK Biobank. Nature. 2020.
163. Wang Y, Karlsson R, Lampa E, Zhang Q, Hedman AK, Almgren M, et al. Epigenetic influences on aging: a longitudinal genome-wide methylation study in old Swedish twins. Epigenetics. 2018;13(9):975-87.
164. Horvath S. DNA Methylation Age Calculator 2021 [Available from: http://dnamage.genetics.ucla.edu/.
165. Finkel D, Sternang O, Jylhava J, Bai G, Pedersen NL. Functional Aging Index Complements Frailty in Prediction of Entry into Care and Mortality. J Gerontol A Biol Sci Med Sci. 2019.
166. Jiang M, Foebel AD, Kuja-Halkola R, Karlsson I, Pedersen NL, Hagg S, et al. Frailty index as a predictor of all-cause and cause-specific mortality in a Swedish population-based cohort. Aging (Albany NY). 2017;9(12):2629-46.
167. Neale Lab. UK Biobank GWAS results 2018 [Available from: http://www.nealelab.is/uk-biobank.
168. Li X, Ploner A, Wang Y, Zhan Y, Pedersen NL, Magnusson PK, et al. Clinical biomarkers and associations with healthspan and lifespan: Evidence from observational and genetic data. EBioMedicine. 2021;66:103318.
169. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, et al. The Ensembl Variant Effect Predictor. Genome Biol. 2016;17(1):122.
170. Ng PC, Henikoff S. SIFT: predicting amino acid changes that affect protein function. Nucleic Acids Research. 2003;31(13):3812-4.
171. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. Nature Methods. 2010;7(4):248-9.
172. Li X, Ploner A, Karlsson IK, Liu X, Magnusson PKE, Pedersen NL, et al. The frailty index is a predictor of cause-specific mortality independent of familial effects from midlife onwards: a large cohort study. BMC Medicine. 2019;17(1):94.
173. Mak JKL, Kuja-Halkola R, Wang YZ, Hagg S, Jylhava J. Frailty and comorbidity in predicting community COVID-19 mortality in the UK Biobank: The effect of sampling. Journal of the American Geriatrics Society. 2021;69(5):1128-39.
174. Williams DM, Jylhava J, Pedersen NL, Hagg S. A frailty index for UK Biobank participants. 2017.

175. Prins BP, Abbasi A, Wong A, Vaez A, Nolte I, Franceschini N, et al. Investigating the Causal Relationship of C-Reactive Protein with 32 Complex Somatic and Psychiatric Outcomes: A Large-Scale Cross-Consortium Mendelian Randomization Study. PLoS Med. 2016;13(6):e1001976.
176. Liu XR, Pawitan Y, Clements MS. Generalized survival models for correlated time-to-event data. Stat Med. 2017;36(29):4743-62.
177. Sjolander A, Lichtenstein P, Larsson H, Pawitan Y. Between-within models for survival analysis. Stat Med. 2013;32(18):3067-76.
178. Zhan Y, Liu XR, Reynolds CA, Pedersen NL, Hagg S, Clements MS. Leukocyte Telomere Length and All-Cause Mortality: A Between-Within Twin Study With Time Dependent Effects Using Generalized Survival Models. Am J Epidemiol. 2018.
179. Bakdash JZ, Marusich LR. Repeated Measures Correlation. Front Psychol. 2017;8.
180. Li B, Leal SM. Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. Am J Hum Genet. 2008;83(3):311-21.
181. Zhan XW, Hu YN, Li BS, Abecasis GR, Liu DJJ. RVTESTS: an efficient and comprehensive tool for rare variant association analysis using sequence data. Bioinformatics. 2016;32(9):1423-6.
182. Li X, Ploner A, Wang Y, Magnusson PK, Reynolds C, Finkel D, et al. Longitudinal trajectories, correlations and mortality associations of nine biological ages across 20-years follow-up. Elife. 2020;9.
183. Vetter VM, Meyer A, Karbasiyan M, Steinhagen-Thiessen E, Hopfenmuller W, Demuth I. Epigenetic Clock and Relative Telomere Length Represent Largely Different Aspects of Aging in the Berlin Aging Study II (BASE-II). J Gerontol a-Biol. 2019;74(1):27-32.
184. Belsky DW, Caspi A, Houts R, Cohen HJ, Corcoran DL, Danese A, et al. Quantification of biological aging in young adults. Proc Natl Acad Sci U S A. 2015;112(30):E4104-10.
185. An J, Li H, Tang Z, Zheng D, Guo J, Liu Y, et al. Cognitive Impairment and Risk of All-Cause and Cardiovascular Disease Mortality Over 20-Year Follow-up: Results From the BLSA. J Am Heart Assoc. 2018;7(15):e008252.
186. Ezzatvar Y, Ramirez-Velez R, Saez de Asteasu ML, Martinez-Velilla N, Zambom-Ferraresi F, Izquierdo M, et al. Physical Function and All-Cause Mortality in Older Adults Diagnosed With Cancer: A Systematic Review and Meta-Analysis. J Gerontol A Biol Sci Med Sci. 2021;76(8):1447-53.
187. Fan J, Yu C, Guo Y, Bian Z, Sun Z, Yang L, et al. Frailty index and all-cause and cause-specific mortality in Chinese adults: a prospective cohort study. Lancet Public Health. 2020;5(12):e650-e60.
188. Mitnitski AB, Graham JE, Mogilner AJ, Rockwood K. Frailty, fitness and late-life mortality in relation to chronological and biological age. BMC Geriatr. 2002;2:1.
189. Crimmins E, Vasunilashorn S, Kim JK, Alley D. Biomarkers related to aging in human populations. Adv Clin Chem. 2008;46:161-216.
190. Deelen J, van den Akker EB, Trompet S, van Heemst D, Mooijaart SP, Slagboom PE, et al. Employing biomarkers of healthy ageing for leveraging genetic studies into human longevity. Experimental Gerontology. 2016;82:166-74.
191. Sakaue S, Kanai M, Karjalainen J, Akiyama M, Kurki M, Matoba N, et al. Trans-biobank analysis with 676,000 individuals elucidates the association of polygenic risk scores of complex traits with human lifespan. Nat Med. 2020;26(4):542-8.
192. Emerging Risk Factors C, Sarwar N, Gao P, Seshasai SR, Gobin R, Kaptoge S, et al. Diabetes mellitus, fasting blood glucose concentration, and risk of vascular disease: a collaborative meta-analysis of 102 prospective studies. Lancet. 2010;375(9733):2215-22.

193. Markozannes G, Koutsioumpa C, Cividini S, Monori G, Tsilidis KK, Kretsavos N, et al. Global assessment of C-reactive protein and health-related outcomes: an umbrella review of evidence from observational studies and Mendelian randomization studies. Eur J Epidemiol. 2020.
194. Laufs U, Parhofer KG, Ginsberg HN, Hegele RA. Clinical review on triglycerides. Eur Heart J. 2020;41(1):99-109c.
195. Riaz H, Khan SU, Rahman H, Shah NP, Kaluski E, Lincoff AM, et al. Effects of high-density lipoprotein targeting treatments on cardiovascular outcomes: A systematic review and meta-analysis. Eur J Prev Cardiol. 2019;26(5):533-43.
196. Wang YJ, Lammi-Keefe CJ, Hou LF, Hu G. Impact of low-density lipoprotein cholesterol on cardiovascular outcomes in people with type 2 diabetes: A meta-analysis of prospective cohort studies. Diabetes Res Clin Pr. 2013;102(1):65-75.
197. Hamazaki T, Okuyama H, Ogushi Y, Hama R. Towards a Paradigm Shift in Cholesterol Treatment. A Re-examination of the Cholesterol Issue in Japan. Ann Nutr Metab. 2015;66 Suppl 4:1-116.
198. Collins R, Reith C, Emberson J, Armitage J, Baigent C, Blackwell L, et al. Interpretation of the evidence for the efficacy and safety of statin therapy. Lancet. 2016;388(10059):2532-61.
199. Ravnskov U, Diamond DM, Hama R, Hamazaki T, Hammarskjold B, Hynes N, et al. Lack of an association or an inverse association between low-density-lipoprotein cholesterol and mortality in the elderly: a systematic review. Bmj Open. 2016;6(6).
200. Ference BA, Ginsberg HN, Graham I, Ray KK, Packard CJ, Bruckert E, et al. Low-density lipoproteins cause atherosclerotic cardiovascular disease. 1. Evidence from genetic, epidemiologic, and clinical studies. A consensus statement from the European Atherosclerosis Society Consensus Panel. European Heart Journal. 2017;38(32):2459-72.
201. Sung KC, Huh JH, Ryu S, Lee JY, Scorletti E, Byrne CD, et al. Low Levels of Low-Density Lipoprotein Cholesterol and Mortality Outcomes in Non-Statin Users. J Clin Med. 2019;8(10).
202. Yi SW, Yi JJ, Ohrr H. Total cholesterol and all-cause mortality by sex and age: a prospective cohort study among 12.8 million adults. Sci Rep-Uk. 2019;9.
203. Fry A, Littlejohns TJ, Sudlow C, Doherty N, Adamska L, Sprosen T, et al. Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population. American Journal of Epidemiology. 2017;186(9):1026-34.
204. Swanson JM. The UK Biobank and selection bias. Lancet. 2012;380(9837):110-.
205. Batty GD, Gale CR, Kivimaki M, Deary IJ, Bell S. Comparison of risk factor associations in UK Biobank against representative, general population based studies with conventional response rates: prospective cohort study and individual participant meta-analysis. BMJ. 2020;368:m131.
206. Hutcheon JA, Chiolero A, Hanley JA. Random measurement error and regression dilution bias. BMJ. 2010;340:c2289.
207. Tripepi G, Jager KJ, Dekker FW, Zoccali C. Selection bias and information bias in clinical research. Nephron Clin Pract. 2010;115(2):c94-9.
208. Burgess S, Thompson SG. Use of allele scores as instrumental variables for Mendelian randomization. International Journal of Epidemiology. 2013;42(4):1134-44.
209. Davies NM, Holmes MV, Smith GD. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. Bmj-Brit Med J. 2018;362.
210. Brumpton B, Sanderson E, Heilbron K, Hartwig FP, Harrison S, Vie GA, et al. Avoiding dynastic, assortative mating, and population stratification biases in Mendelian randomization through within-family analyses. Nature Communications. 2020;11(1).

211. Hussien H, Nastasa A, Apetrii M, Nistor I, Petrovic M, Covic A. Different aspects of frailty and COVID-19: points to consider in the current pandemic and future ones. BMC Geriatr. 2021;21(1):389.