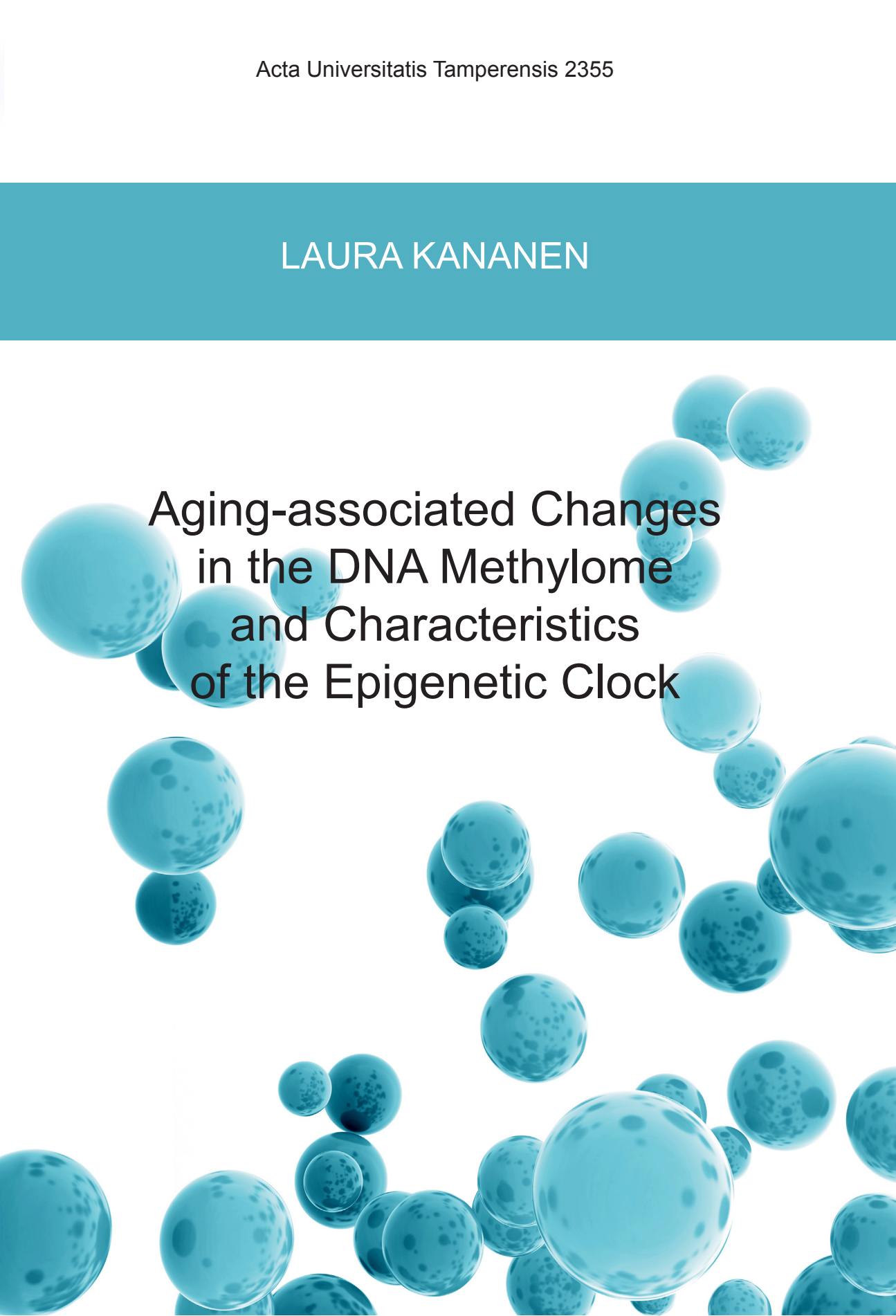


LAURA KANANEN

**Aging-associated Changes  
in the DNA Methylome  
and Characteristics  
of the Epigenetic Clock**





LAURA KANANEN

Aging-associated Changes  
in the DNA Methylome  
and Characteristics  
of the Epigenetic Clock



ACADEMIC DISSERTATION

To be presented, with the permission of  
the Faculty Council of the Faculty of Medicine and Life Sciences  
of the University of Tampere,  
for public discussion in the auditorium F115 of the Arvo building,  
Arvo Ylpön katu 34, Tampere,  
on 29 March 2018, at 12 o'clock.

UNIVERSITY OF TAMPERE

LAURA KANANEN

Aging-associated Changes  
in the DNA Methylome  
and Characteristics  
of the Epigenetic Clock

*Acta Universitatis Tamperensis 2355*  
*Tampere University Press*  
*Tampere 2018*

## ACADEMIC DISSERTATION

University of Tampere, Faculty of Medicine and Life Sciences  
Finland

*Supervised by*

Professor Mikko Hurme  
University of Tampere  
Finland

*Reviewed by*

Associate Professor Anu Kauppinen  
University of Eastern Finland  
Finland  
Docent Cilla Söderhäll  
Karolinska Institutet  
Sweden

The originality of this thesis has been checked using the Turnitin OriginalityCheck service in accordance with the quality management system of the University of Tampere.

Copyright ©2018 Tampere University Press and the author

Cover design by  
Mikko Reinikka

Acta Universitatis Tamperensis 2355  
ISBN 978-952-03-0668-7 (print)  
ISSN-L 1455-1616  
ISSN 1455-1616

Acta Electronica Universitatis Tamperensis 1861  
ISBN 978-952-03-0669-4 (pdf)  
ISSN 1456-954X  
<http://tampub.uta.fi>

Suomen Yliopistopaino Oy – Juvenes Print  
Tampere 2018



*This PhD thesis is dedicated to my family.*



# CONTENTS

List of Original Communications .....	9
Abbreviations .....	11
Abstract .....	13
Tiivistelmä.....	17
1 Introduction .....	21
2 Review of the literature .....	25
2.1 Human lifespan .....	25
2.2 Biology of aging .....	26
2.2.1 Biomarkers of aging.....	27
2.2.2 Longevity genetics.....	29
2.2.3 The hallmarks of aging.....	31
2.2.3.1 Causes of damage in aging.....	32
2.2.3.2 Responses to aging-associated cellular damage .....	37
2.2.3.3 The integrative hallmarks of aging.....	40
2.3 The human immune system.....	40
2.3.1 Aging of the immune system .....	42
2.3.2 Cytomegalovirus infection.....	43
2.4 DNA Methylation.....	44
2.4.1 Maintenance and de novo methylation .....	48
2.4.2 Demethylation .....	49
2.4.3 DNA methylation and environmental factors .....	50
2.4.4 DNA methylome changes in diseases .....	54
2.4.4.1 Cancer.....	54
2.4.4.2 Imprinting disorders .....	55
2.4.4.3 Brain disorders and DNA methylation.....	55
2.4.4.4 Environment, DNA methylation and disorders .....	56
2.4.4.5 Genetic variation, DNA methylation and disorders.....	57
2.4.5 Aging-associated changes in the DNA methylome.....	58
2.4.5.1 Epigenetic clock.....	62
3 Aims of the thesis.....	67
4 Subjects and Methods.....	68
4.1 Subjects.....	68
4.1.1 The Young Finns Study (I, III).....	68
4.1.2 Vitality 90+ study (II, III, IV).....	68

4.2	Methods .....	70
4.2.1	DNA methylation analysis .....	70
4.2.1.1	Blood sample collection and sample preparation (I-IV) .....	70
4.2.1.2	Illumina methylation assay (I-IV).....	70
4.2.1.3	Preprocessing of methylation data (I-IV) .....	71
4.2.1.4	DNA methylome age calculation (II-IV).....	72
4.2.1.5	Estimation of the blood cell counts (I, III).....	72
4.2.1.6	Predicting functional roles for regions of interest (I, IV) .....	72
4.2.2	Gene expression data (IV) .....	73
4.2.2.1	Origin of RNA .....	73
4.2.2.2	Illumina gene expression assay .....	73
4.2.3	Fluorescence-activated cell sorting analysis (II, III, IV) .....	74
4.2.4	Cytomegalovirus titer (II).....	74
4.2.5	Other covariate data (IV) .....	74
4.2.6	Availability of raw data (I-IV) .....	76
4.2.7	Testing hypotheses (I-IV) .....	76
5	Results .....	77
5.1	The Studies.....	77
5.2	Aging-associated DNA methylation changes in middle-aged individuals (I) .....	77
5.2.1	Comparison analysis of different studies.....	81
5.3	DNA methylome ages (II, III, IV).....	84
5.3.1	Follow-up of DNA methylome age values (III) .....	84
5.3.2	Cytomegalovirus-association with epigenetic aging (II) .....	88
5.4	The effect of blood cell subtypes (I-IV).....	89
5.5	Methylomic predictors of mortality (IV).....	93
6	Discussion.....	96
6.1	The Studies.....	96
6.2	Aging-associated DNA methylation changes at single CpG sites .....	96
6.2.1	Comparison analysis of different studies.....	97
6.3	Methylomic mortality predictors .....	99
6.4	Longitudinal analysis of the epigenetic age.....	102
6.5	Cytomegalovirus infection.....	105
6.6	The effect of blood cell subtypes .....	105
6.7	$\beta$ value vs DNA methylation status .....	108
6.8	Limitations.....	110
7	Summary and Conclusions.....	112
8	Acknowledgements .....	114



9	References .....	116
10	Original communications.....	138



# List of Original Communications

The thesis is based on the following original publications, which are referred to in the text by their Roman numerals (I-IV):

- I **Kananen L**, Marttila S, Nevalainen, Jylhävä J, Mononen N, Kähönen M, Raitakari OT, Lehtimäki T, Hurme M. Aging-associated DNA methylation changes in middle-aged individuals: the Young Finns study. *BMC Genomics*. 2016 Feb 9;17:103.1.
- II **Kananen L**<sup>1</sup>, Nevalainen T<sup>1</sup>, Jylhävä J, Marttila S, Hervonen A, Jylhä M, Hurme M. Cytomegalovirus infection accelerates epigenetic aging. *Exp Gerontol*. 2015 Dec;72:227-9.
- III **Kananen L**, Marttila S, Nevalainen T, Kummola L, Junttila I, Mononen N, Kähönen M, Raitakari OT, Hervonen A, Jylhä M, Lehtimäki T, Hurme M, Jylhävä J. The trajectory of the blood DNA methylome ageing rate is largely set before adulthood: evidence from two longitudinal studies. *Age (Dordr)*. 2016 Jun;38(3):65.
- IV Jylhävä J, **Kananen L**, Raitanen J, Marttila S, Nevalainen T, Hervonen A, Jylhä M, Hurme M. Methylomic predictors demonstrate the role of NF- $\kappa$ B in old-age mortality and are unrelated to the aging-associated epigenetic drift. *Oncotarget*. 2016 Apr 12;7(15):19228-41.

<sup>1</sup>Equal contribution

The original publication in Study I was also partly used in the doctoral thesis of Saara Marttila (University of Tampere, 2016). The original publications are reproduced in this thesis with the permission of the copyright holders or under the terms of the Creative Commons Attribution Licence.



# Abbreviations

5caC	5-carboxycytosine
5fC	5-formylcytosine
5hmC	5-hydroxymethylcytosine
5hmU	5-hydroxymethyluracil
5mC	5-methylcytosine
AID	Activation-induced deaminase
APE1	Apurinic/aprimidinic endonuclease 1
APOBEC	Apolipoprotein b mRNA editing enzyme catalytic polypeptide
BER	Base excision repair
BMI	Body mass index
cAge	Chronological age
CD	Cluster of differentiation
cf-DNA	Cell-free DNA
CGI	CpG island
CMV	Cytomegalovirus
CpG site	Region in DNA where cytosine and guanine are separated by one phosphate
CRP	C-reactive protein
DHEAS	Dehydroepiandrosterone sulfate
DNA	Deoxyribonucleic acid
DNAmAge	DNA methylome age; epigenetic age
DNMT	DNA methyltransferase
FACS	Fluorescence-activated cell sorting analysis
GO	Gene ontology
HR	Hazard ratio
IL	Interleukin
MAD	Median absolute deviation
MED	Median
MBD4	Methyl binding domain protein 4
MMSE	Mini-mental state examination
mRNA	messenger RNA
mtDNA	mitochondrial DNA
NA	Not available
PBMC	Peripheral blood mononuclear cell
PCA	Principal component analysis
PRC2	Polycomb Repressive Complex 2
SEM	Standard error of mean
SNP	Single nucleotide polymorphism
TDG	Thymine DNA glycosylase
TET	Ten-eleven-translocation protein
V90	Vitality 90+ study
YFS	Young Finns study
QTL	Quantitative trait loci
$\Delta$ -cAge-DNAmAge	Difference between chronological age and DNAmAge



# Abstract

**Background:** Aging is a continuous process, whereby the physiological integrity of an organism gradually diminishes. Eventually, disabilities in functioning evolve, and the risk of developing pathologies and death increases. The rate of aging differs between species, organisms and tissues, and inheritable components and environmental factors contribute to the aging process.

DNA methylation has a critical role in genomic stability, cell differentiation and development. The genomes of mammals are abundant in 5-methylcytosines i.e. methylated cytosines (5mCs), which are primarily located in the sequence CpG. In general, the interindividual differences between genome-wide DNA methylation profiles increase with age. During aging, random as well as systematic, clocklike-behaving DNA methylation changes occur, and genetic and environmental factors contribute to these changes. Here, clocklike-behaviour refers to change where DNA methylation levels increase or decrease linearly with age, and these changes constitute *the epigenetic clock*. DNA methylation is a marker that disentangles details in human well-being at a more detailed level compared to conventional biomarkers. Importantly, DNA methylation has dynamic nature and is prone to the effect of environmental input.

The clocklike-behaving CpG sites can be used in a sophisticated manner to estimate calendar age of “a blind sample” of DNA with a high level of accuracy. The difference between chronological age and this *epigenetic age* estimate ( $\Delta$ -cAge-DNA<sub>m</sub>Age) is considered a highly promising biomarker of the aging rate, and many accounts of its significance in human fitness, morbidity, mortality and longevity have been reported.

**Aims:** The specific hypotheses for this thesis comprise the following. The longitudinal within-subject behavior of  $\Delta$ -cAge-DNA<sub>m</sub>Age as well as the blood cell subtype landscapes was indefinite. In addition, discrepancies in the lists of reported aging-associated methylation sites have emerged, thus requiring further evaluations. Another question concerned methylomic mortality predictors: were there DNA methylation sites that could be used as survival predictors in the elderly (+90 years of age) and could those predictors even overcome conventional biomarkers of aging in the survival analysis? Moreover, is epigenetic aging associated with biomarkers that are related to immune system aging such as the cytomegalovirus or blood cell distribution?

**Methods:** Two study populations, the Young Finns and the Vitality 90+, together representing a wide age range (from 15 years to 94 years of chronological age) were used in the thesis. The samples were utilized in cross-sectional (Study I, II) and longitudinal (Study III, IV) manners, and the sample sizes ranged between 111 and 183 participants in Studies I-IV. The epigenomes of the participants were characterized using microarray technology-based technology, and the subjects’

DNAmAges were determined using Horvath's calculator of epigenetic age. Blood cell compositions were determined using either DNA methylation profile-based estimation algorithm or fluorescence-activated cell sorting analysis. The hypotheses were tested using regression and correlation analyses.

In the thesis, aging-associated DNA methylation level changes (I), and mortality predictors (IV) at single CpG site resolution were characterized. The longitudinal behavior of difference between calendar age and epigenetic age ( $\Delta$ -cAge-DNAmAge) (III), and its role as mortality predictor (IV) were explored. Furthermore, the  $\Delta$ -cAge-DNAmAge association with cytomegalovirus infection was analyzed (II). In all sections in the thesis, the role of blood cell sample heterogeneity was considered (I-IV). Specifically, the relevance of controlling sample heterogeneity in epigenome-wide association studies (EWASs) was evaluated through analysis where results from multiple studies were compared (I), the association between  $\Delta$ -cAge-DNAmAge and blood cell subtype counts was explored (III), and the overall longitudinal behavior of the blood cell subtypes was investigated (III).

**Results and conclusions:** The results from Study I highlighted that clocklike-behaving CpG site methylation may be reliably detected from a cross-sectional sample with an age range of only nine years. The data showed that aging-associated hypermethylation and hypomethylation at single CpG site resolution were related to different cellular functions and are enriched in different ways in single genes. The study underlined that, in order to obtain replicative results from EWASs with heterogenic tissue samples such as whole blood, cell count adjustment appears to be essential.

In the survival analysis (IV), the mortality-predicting methylomic signature performed better than the conventional aging biomarkers and was independent from the aging-associated epigenetic drift. The methylomic prediction signature supported the genomic-level role of NF- $\kappa$ B at the very end of the human lifespan.

In Study II, the increased epigenetic age of the blood cells was associated with latent cytomegalovirus infection in the populations of young adults and nonagenarians; however, this finding may be a reflection of changes in the blood cell composition. In studies II and III, the cell subtype counts correlated clearly with epigenetic aging ( $\Delta$ -cAge-DNAmAge). The most significant correlate linked with DNA methylomic data were CD28<sup>-</sup> T cells, which are markers of immune system exhaustion and aging.

The longitudinal data of blood cell composition (III) provided evidence that, in parallel with aging-associated shifts in the immune cell composition, intra-individual changes in blood cell subtype proportions are relatively small during young adulthood and middle age, as well as in the advanced age for several years or even decades. These results suggest that the major shifts in the blood cell composition might occur somewhere after middle age and before advanced ages. Thus, the blood cell composition-related issues discussed here warrant careful consideration when



interpreting blood cell based results concerning individuals with varying chronological ages.

Longitudinal methylomic data (III) provided evidence that the difference between chronological and epigenetic age is surprisingly stable over several years or even decades, and, when accompanied with previous reports, it may be hypothesized that the main trajectory of the blood DNA methylome aging rate ( $\Delta$ -cAge-DNA<sub>m</sub>Age) is largely set before adulthood. This hypothesis is thought provoking, because older epigenetic age is also associated with increased mortality rate in adulthood. Therefore, in this light, epigenetic changes in the beginning of life gain an even more crucial meaning for the entire human lifespan.



# Tiivistelmä

**Tausta:** Ikääntyminen on jatkuva prosessi, jonka aikana organismin fysiologinen eheys heikkenee. Tämän seurauksena yksilön toimintakyky alentuu ja sairauksien sekä kuoleman riski kasvaa. Ikääntymisprosessiin vaikuttavat periytyvät komponentit ja ympäristö. Eri eliölajeilla, yksilöillä ja kudoksilla ikääntymisen etenemisnopeus on erilainen.

DNA:n metylaatiolla on merkittävä rooli genomisen stabiiliuden ylläpitäjänä, solujen erilaistumisessa ja yksilön kehityksessä. Nisäkkäiden genomissa esiintyy paljon 5-metyylietyosiineja eli metyloituja sytosiineja (5mC), jotka ovat useimmiten sekvenssissä CpG. Ikääntymisen aikana CpG-kohtien metylaatioissa tapahtuu sekä sattumanvaraisesti että systemaattisen kellomaisesti käyttäytyviä metylaatiotasomuutoksia, joihin molempiin vaikuttavat ympäristö- ja geneettiset tekijät. Kellomaisessa muutoksessa metylaatiotaso nousee tai laskee hyvin lineaarisesti ajan funktiona läpi elämän. Nämä kellomaisesti käyttäytyvät metylaatiokohdat muodostavat ns. *epigeneettisen kellon*. Yleisesti ottaen DNA-metylaatioissa nähtävät yksilöiden väliset erot kasvavat ikääntyessä, ja metylaatio on markkeri joka kuvastaa hyvinvointia eri tasolta kuin perinteiset ikääntymisen biomarkerit. DNA-metylaation merkittävyys vanhenemiseen liittyvässä tutkimuksessa korostuu sen vuoksi, että metylaatiotasot muuttuvat ympäristötekijöiden johdosta, ja näillä DNA-metylaatiomuutoksilla on myös kyky palautua.

Kellomaisesti käyttäytyviä DNA:n metylaatiokohtia voidaan käyttää kronologisen iän estimointiin DNA-sokkonäytteestä huomattavalla tarkkuudella. Tätä epigeneettisen iän estimaattia eli *epigeneettistä ikää* on luonnehdittu biologisen iän mittariksi, joka voisi mahdollisesti kuvastaa ihmisten erilaista biologista ikääntymisnopeutta. Useat aiemmat tutkimukset ovat puoltaneet tätä oletusta yhdistäessään epigeneettisen iän fyysiseen toimintakykyyn, sairastavuuteen, kuolleisuuteen sekä pitkäikäisyyteen.

**Tutkimustavoitteet:** Tässä työssä hypoteesit koostuivat seuraavista kohdista. On huomattu, että raportoiduissa ikääntymiseen liitettyissä DNA:n metylaatiokohdissa on huomattavia eroja ja tätä epäyhdenmukaisuutta oli tarpeen tarkastella (I). Tämän lisäksi DNA-metylaatioprofiilin potentiaali kuolleisuuden ennustajana yli 90-vuotiailla on ollut epäselvää, ja lisäksi sen tehokkuus perinteisiin ennustajiin verrattuna oli tarpeen arvioida (IV). Väitöskirjatutkimuksessa keskityttiin myös siihen, miten epigeneettinen ikä sekä verisolualatyypijakaumat käyttäytyvät pitkittäisseurannassa (III). Tutkimuksessa tarkasteltiin lisäksi epigeneettisen ikäestimaatin yhteyttä immuunijärjestelmän ikääntymiseen liittyviin biomarkkereihin, joita olivat sytomegalovirusinfektio (II) ja verisolualatyyppien osuuksien muutokset (III).

**Metodit:** DNA:n metylaatiotasot mitattiin genomilaajuisella mikrosiruanalytiikalla ja tutkimushypoteesejä testattiin käyttäen muun muassa erityyppisiin regressiomalleihin ja korrelaatioon perustuvia analyysimenetelmiä. Epigeneettinen ikä estimoitii Horvathin laskenta-algoritmillä. Aineistoina käytettiin kahta populaatiopohjaista tutkimusaineistoa, jotka yhdessä kattavat ikävuodet 15-94. Tutkimustietoa analysoitiin yhden (I, II) ja kahden aikapisteen (III, IV) perusteella. Seuranta-ajat olivat neljä tai 25 vuotta ja näytekoot vaihtelivat tutkimuskysymyksestä riippuen N=111 ja N=183 välillä. Väitöskirjatyön kaikissa osatöissä huomioitiin tutkimuksessa käytettyjen verisolunäytteiden heterogeenisyys.

Tarkemmin kuvattuna, väitöskirjatyössä kartoitettiin koko epigenomin laajuisesti ikääntymiseen (I) ja suurentuneeseen kuolleisuusriskiin (IV) liittyviä yksittäisissä DNA:n metylaatiokohdissa tapahtuvia metylaatiotasojen muutoksia. Tämän lisäksi tarkasteltiin laskennallisesti määritetyn DNA-metylomi-ian (kalenteri-ian ja epigeettisen ian erotus;  $\Delta$ -cAge-DNA<sub>m</sub>Age) sekä verisoluosuusjakaumien käyttäytymistä pitkittäisseurannassa (III). Väitöskirjatyössä selvitettiin myös epigeneettisen ikäestimaatin vaihtelun yhteyttä sytomegalovirusinfektioon (II) ja verisolualatyypin osuusjakaumiin (III) sekä potentiaalia kuolleisuuden ennustajana (IV). Soluosuusjakaumien vaihtelun merkitystä koko epigenomin laajuisessa ikäassosiaatioanalyysissä analysoitiin tarkemmin peilaamalla osatyössä I paikannettuja ikäassosiaatioita ja analyysimetodeja aiemmin julkaistuihin vastaaviin tutkimuksiin.

**Tulokset ja johtopäätökset:** Osatyössä I selvisi, että kellomaisesti käyttäytyviä DNA:n metylaatiokohtia voi havaita keski-ikäisistä koostuvassa poikkileikkaustutkimusaineistossa, jonka ikäjakauma on vain yhdeksän vuotta. Riippuen siitä oliko DNA:n metylaatiotason muutos kasvavaa vai vähenevää iän funktiona, muutokset olivat erilaisia päteiltään, kun tarkasteltiin metylaatiokohtiin liitettyjä solun perustoimintoja tai niiden sijoittumista yksittäisiin geeneihin. Huomattiin myös, että assosiaatioanalyysien toistettavuuden kannalta oli merkittävää, että mahdollinen verisolunäytteen heterogeenisyys on huomioitu.

Osatyössä IV havaittiin, että kuolleisuutta ennustava DNA-metylaatioprofiili on tehokkaampi ennustaja kuin perinteiset kuolleisuuteen ja vanhenemiseen yhdistetyt biomarkerit. Epigeneettinen ikä ei ollut yhteydessä kuolleisuuteen ja kuolleisuutta ennustava DNA:n metylaatioprofiili erosi yleisesti tunnetuista ikääntymiseen liitettyistä DNA:n metylaatiomuutoksista. Tässä työssä selvitetty kuolleisuutta ennustava DNA-metylomien profiili tukee genomisella tasolla NF- $\kappa$ B-geenin roolia ihmisen elinkaaren loppupäässä.

Osatyössä II huomattiin, että elämänsä aikana sytomegalovirusinfektion saaneilla tutkittavilla oli korkeampi epigeneettinen ikä verrattuna ei-infektoituneisiin tutkittaviin. Tämä tulos saattaa selittyä muutoksilla verisolu populaatioissa, koska suuri osa eri verisoluosuuksista oli myös yhteydessä epigeneettisen iän vaihteluun (III). Merkittävimpiä DNA-metylaation variaatioon kytkeytyneitä soluosuukskorrelaatteja näyttivät olevan CD28<sup>-</sup> -tyyppiset T-solut (II, III), joita

pidetään yleisesti immuunijärjestelmän kuormitustilan ja vanhenemisen markkereina. Havainnot osatyössä III osoittivat myös, että poikkileikkaustarkastelussa 90-vuotiailla verisoluosuusjakauma on selvästi erilainen kuin nuorilla aikuisilla. Seurantatuloksista nähtiin kuitenkin, että aikuisiässä ennen keski-ikää ja elinkaaren loppupäässä solualatyypijakaumat ovat suhteellisen muuttumattomia. Tämä voisi viitata siihen, että suurin muutos solujakaumissa tapahtuu jossain keski-ian jälkeen ennen 90 vuoden ikää. Nämä tulokset korostavat entisestään verisolupopulaatioiden heterogeenisyyden huomioimisen merkitystä tämän tyyppisissä tutkimuksissa.

Seurantatutkimusaineisto antoi viitteitä, että kalenteri- ja epigeneettisen iän välinen erotus ( $\Delta$ -cAge-DNA<sub>m</sub>Age) muuttuu aikuisiässä suhteellisen vähän useamman vuoden tai vuosikymmenen aikana (III). Kun tähän havaintoon yhdistetään aiemmat epigeneettisen iän seurantalutkimustulokset, missä on keskitytty erityisesti lapsuusikään, vaikuttaa sille, että tämä erotus asettuu ja vakiintuu pääasialliselle tasolle jo ennen aikuisikää. Nämä havainnot ovat ajatuksia herättäviä, koska kiihtynyt epigeneettinen ikääntyminen on liitetty kasvaneeseen kuolleisuusriskiin aikuisilla. Nämä havainnot yhdessä korostavat entisestään lapsuudessa tapahtuvien epigeneettisten muutosten merkitystä myöhemmälle terveydelle elinkaaren aikana.



# 1 INTRODUCTION

Aging is a continuous process, which influences all humans. During this process, the physiological integrity of an organism gradually diminishes. Eventually, this leads to disabilities in functioning, increases the risk of developing pathologies, and death. The factors contributing to the aging process at a cellular level have been studied extensively and are reviewed by, for example, López-Otin et al. (2013) and Carmona et al. (2016) and Pan et al. (2017). These factors include genomic instability, telomere shortening, epigenetic changes, and unbalanced protein homeostasis, decayed nutrient sensing regulation, mitochondrial dysfunction, cellular senescence, stem cell exhaustion and changes in intercellular communication. These determinates of aging are presented in details in the thesis section 2.2 under heading Biology of aging. The aging process and the aging rate are influenced by genetic and environmental factors, and the aging rate varies between species, organisms and tissues. The non-random aspect in the aging process is demonstrated in the high evolutionary conservation levels of the aging-associated biochemical pathways. When considering different species, the average lifespan expectancy is species-specific. Within species-specific limits, the varying length of a lifespan is inheritable for the offspring, and in parallel, is modified by random factors. Within one specie and organism, there are even intra-individual tissue-specific differences in the aging rates. (Lopez-Otin et al. 2013; Carmona and Michan 2016; Kaeberlein et al. 2016; Pan et al. 2017)

In this thesis, the focus is set to DNA methylation changes that occur during aging (presented in section 2.4). The genomes of mammals and many other organisms are abundant in 5-methylcytosines, i.e. methylated cytosines (5mCs), which are primarily located in the sequence CpG. Typically, through the enzymatic methylation of cytosine in DNA sequence, the gene activities may be switched on and off, and the gene activity may be inherited from a cell to a daughter cell. (Holliday and Pugh 1975; Riggs 2002; Holliday 2006) From a wider perspective, DNA methylation is a mechanism, which is often associated with long-term gene expression silencing in X chromosome inactivation, imprinting and tissue-specific gene expression. Overall, it has a critical role in genomic stability, cell differentiation and development. (P. A. Jones and Liang 2009; B. C. Christensen et al. 2012; Smith and Meissner 2013; W. Xu et al. 2016; Rasmussen and Helin 2016; Elhamamsy 2016)

In general, the interindividual differences between genome-wide DNA methylation profiles increase with age (Zampieri et al. 2015), and genetic and environmental factors contribute to this change (van Dongen et al. 2016). The aging-related methylomic changes at the CpG sites are termed as the epigenetic drift (Teschendorff et al. 2013). Specifically, cell-type-specific genome-wide DNA methylation profiles are changed, and these changes are associated with various health conditions. Changes in the DNA methylome have been associated with diseases with aging-resembling features and diseases that are frequent in the aged population. Moreover, many of the aging-associated DNA methylation changes

share characteristics in common with DNA methylomic changes associated with pathologies (i.e. cancer). (Zampieri et al. 2015)

Interestingly, in addition to more random DNA methylation changes, clocklike-behaving DNA methylation changes occur during aging. In clocklike-behaving methylation sites, DNA methylation level decreases or increases with almost constant rate as a function of age. These changes have been considered as the base for the epigenetic clock. The DNA methylation levels in these clock-CpG sites correlate highly with the chronological age of an individual, and interestingly, the correlation is even better than between telomere length and chronological age. (Zampieri et al. 2015; M. J. Jones et al. 2015) Due to growing interest and feasible technology, DNA methylation has been studied intensively in numerous large-scale epidemiological human cohorts (Schübeler 2015). Based on the extensive research, it has been discovered that clocklike-behaving CpG sites can be used in a sophisticated manner to estimate the age of “a blind sample” from a DNA. (Zampieri et al. 2015; Jylhava et al. 2017). This *epigenetic age* estimation might at least be employed in practice in forensic investigations in order to assess age of an unknown individual with DNA sample (Freire-Aradas et al. 2017).

A popular epigenetic age estimate is Horvath’s DNAmAge, which is fitted for multiple human tissues (Horvath 2013). The DNAmAge estimate has emerged as a highly promising biomarker of the aging rate, and many supporting reports of its significance in human fitness, morbidity, mortality and longevity have been announced following the DNAmAge calculator being published in 2013 (Zampieri et al. 2015; Jylhava et al. 2017).

Nevertheless, the literature provides only hints of the complete picture regarding epigenetics. To what extent does the epigenetic drift have its own functionally relevant role, or is it merely a reflection of varying chromatin states? For example, at present, there is no causal experimental evidence supporting lifespan extension through changes in DNA methylation profile alone (Lopez-Otin et al. 2013). There is also the question of how the epigenetic clock is functioning. In spite of the open questions, DNA methylation and epigenetic age estimates have proven to be strong biomarkers in numerous pathologies and conditions (Schübeler 2015; M. J. Jones et al. 2015; Zampieri et al. 2015; Leenen et al. 2016; Elhamamsy 2017).

The hypotheses in this thesis comprise the following detailed issues. The longitudinal behavior of the epigenetic age (i.e. Horvath’s DNAmAge) in adults has been indefinite. In addition, discrepancies in the lists of reported aging-associated methylation sites in human samples have emerged, thus requiring further evaluation. The second question concerns methylomic mortality predictors: are there methylation sites that could be used as survival predictors in the elderly (+90 years of age) and could these predictors even overcome conventional biomarkers of aging in the survival analysis? Moreover, are factors that are related to immune system aging, cytomegalovirus (CMV) and blood cell distribution associated with epigenetic aging? Therefore, in studies I-IV, aging-associated DNA methylation level changes (I), and mortality predictors (IV) at single CpG site resolution are characterized. The



longitudinal behavior of the DNA methylome age (III), and its role as a mortality predictor (IV) is explored. Furthermore, the epigenetic age change associated with CMV infection is analyzed (II). In all sections in the thesis, the role of blood cell sample heterogeneity is considered (I-IV). Specifically, the relevance of controlling sample heterogeneity in EWASs is evaluated (I), the association between the DNA methylome age and blood cell subtype counts is explored (III), and the overall longitudinal behavior of the blood cell subtypes is investigated (III).



## 2 REVIEW OF THE LITERATURE

### 2.1 Human lifespan

Normal aging and disease in combination contribute to functional ability, survival and longevity. The numbers in longevity, disease incidence and fatality rates have been improved considerably during the past decades, and the change is continuing in developed countries with reasonable welfare (E. M. Crimmins 2015).

At population level, aging-associated transitions in health condition often follow a sequential of events: 1) changes in biological risk factors occur; 2) pathologies are diagnosed; 3) frailty and disabilities arise, and 4) these are followed by death. Overall, morbidity, i.e. disease burden concentrates close to the time of death. Nonetheless, studies have shown that, in addition to the obvious concordance of morbidity and mortality, these two issues follow their own trends. That is, for instance, people are living longer than previously with chronic diseases such as cancer. (E. Crimmins et al. 2008; E. M. Crimmins and Beltran-Sanchez 2011; E. M. Crimmins 2015) A large-scale evaluation of the average population health changes in 187 countries during 1990-2010 has demonstrated that for every 1-year increase in lifespan expectancy, the length of healthy life is extended for ca. 10 months (Salomon et al. 2012). Women reach 90 years of age more often than men do. At the same time, when compared to nonagenarian men (i.e. men 90-99 years of age), women of the same age more often have multiple diseases and disparities in functioning (Jylhava 2014).

The worldwide number of individuals who have reached an extremely old age, i.e. 110 years of age, is between few tens and less than thousand (Cournil et al. 2010; Adams Jun 13, 2017). The number varies depending on source and validation level provided by longevity researchers. In history, the longest-lived person with confirmed birth and death dates was Jeanne Calment, who lived for 122 years. These facts underline the existence of high mortality rates at very old ages, and the aging researchers debate the possible limits for the biological maximum of human lifespan. (Dong et al. 2016; Shkolnikov et al. 2016)

The life expectancy of an individual at the time of birth or later is the expected average time in years between that time and time of death, and thus, it is different from the maximum of human lifespan. The expectancy estimates are dependent on birth cohort, chronological age, ethnicity and gender. (Rector et al. 2016) Human life expectancy has changed dramatically during the last century in many countries worldwide, and at present, the average age in these regions with reasonable welfare systems is increasing due to decreasing birth rates and extending life expectancies. Earlier, the life expectancy used to be approximately 50 years, yet recently, it has shifted to close to 80 years. Transition in the age distribution has had different stages: at the beginning, mortality was higher at the young population due to infectious diseases, and the disappearance of such deaths increased life expectancy. Following this, cardiovascular diseases and cancer in older individuals were the main causes of

mortality; for example, in the United States in the 2000s, these diseases accounted for more than half of all deaths. Nowadays, given that fatality rates of heart diseases are decreasing, the lifespan expectancy continues to grow, especially at the old ages. (K. Christensen et al. 2009; E. M. Crimmins and Beltran-Sanchez 2011; E. M. Crimmins 2015)

In Japan, the life expectancy is the highest in the world. After the 1960s, the number of Japanese reaching very old age, i.e. reaching 100 years of age, has doubled in every successive decade. *Population projection 2015–2065* estimations by Statistics Finland exemplifies also the general trend of population structure changes in developed countries with reasonable welfare systems. The projection estimations suggests that of total Finnish population ( $N=5861\,491$ ) for example in 2040, the number of individuals who reach 90 years of age or more is 140 632 (2.4%). The number is clearly higher than it was in 1980, when the corresponding number of individuals who had reached 90 years of age was 6107. This is only 0.12% of total population at that time. (Official Statistics of Finland, OSF 2015) At the oldest ages, the mortality rate increases exponentially through the gained years, and the survival percentages at population level are relatively low. Meaning that only ~1% of individuals alive in 2010 in, for instance, the United States, are likely to be a centenarian (E. M. Crimmins 2015). In summary, nowadays, the number of nonagenarians and older individuals is still small but the number is clearly growing.

The growing rate of life expectancy at population level is difficult to predict, and thus, different scenarios exist. For instance, at the beginning of 2000, the U.S. Social Security Administration has estimated that, by 2050, in every survived decade before 65 years of age, people will gain at least one year of lifespan extension and half a year extension per each subsequent decade (E. M. Crimmins 2015). More optimistic predictions do exist: researchers have estimated even a two-times larger increase in the gained years/survived decade, if the lifespan expectancy remains to grow as fast as it has been growing (K. Christensen et al. 2009). Thus, using these lifespan expectancy values it may even be estimated that as much as 50% of the birth cohort of the year 2000 in the United States might survive to the age of 100 years. (E. M. Crimmins 2015)

## 2.2 Biology of aging

In organismal aging, the physiological integrity is gradually lost. This leads to disabilities in functioning, increases the risk of developing many pathologies such as cancer, diabetes, cardiovascular disorders and neurodegenerative diseases, and the risk of death is greater. The rate of aging demonstrates both random and non-random features. Aging-associated biochemical processes and genetic pathways show high evolutionary conservation levels. The average life expectancy is species-specific. However, within one specie, there are variations in the aging rates between individual organisms, and intra-individual tissue-specific differences in the aging

rates exist. The rate of aging is modified by random factors but is also inheritable. Through comprehensive research evidence, it is now obvious that the aging process is not driven by a single factor. In contrast, the aging process is a complex trait, which is influenced by genetic and environmental factors. (Lopez-Otin et al. 2013; Carmona and Michan 2016; Kaeberlein et al. 2016)

### 2.2.1 Biomarkers of aging

The biological (risk) factors, i.e. biomarkers or “biomeasures”, may indicate elevated mortality risk and transitions in health status (pathologies, loss of functioning, disability, frailty) and they can even be used to describe one’s biological age (Arbeev et al. 2016). The National Institute of Health (NIH) has defined a biomarker as: “a characteristic that is objectively measured, and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacological responses to a therapeutic intervention”. In general, some of the aging biomarkers are specific to certain diagnoses or prognoses of pathologies, and some are descriptive of the aging process itself. (E. Crimmins et al. 2008) The American Federation of Aging Research has suggested that a good biomarker of aging: (1) predicts the rate of aging in such a way that it predicts lifespan better than chronological age; (2) is a non-pathological phenomena associated with aging process; (3) is easily and repeatedly assessable in such a way that it does not harm the study subject, and (4) is also suitable for analysis in laboratory animals. The biological age predictors that describe one’s health status better than chronological age have been extensively explored, but this search has been demanding. It has been also questioned whether such biomarkers or biomarker combinations can be found due to strong correlation between normal aging and chronic diseases. (Burkle et al. 2015; Jylhava et al. 2017)

The conventional biomarkers of healthy aging may be divided into categories of physiological, immune system, endocrine system, physical capability, and cognitive functions (Table 1) (E. Crimmins et al. 2008; Lara et al. 2015). Other established aging biomarkers exist, including DNA methylation (Zampieri et al. 2015), and telomere attrition (Bojesen 2013; Deelen et al. 2014b). Moreover, promising candidate aging biomarkers or biomarker combinations such as gene expression patterns (J. Yang et al. 2015), plasma cell free DNA (Jylhava et al. 2011), micro-RNAs (Thum 2014), long non-coding RNAs (Thum 2014) and histone modifications (Han and Brunet 2012) have emerged. Even “the brain age” through neuroimaging can be utilized to predict chronological age, and this measure has also shown mortality-predicting capability (Cole et al. 2017).

**Table 1.** The conventional human aging biomarkers and their domains (modified from Crimmins et al. 2008; Lara et al. 2015). All these factors change during aging, and they are manifesting gradual weakening in regulation of cellular pathways and physiological integrity of an organism. Direction of change by increasing chronological age is marked with ↑ or ↓ if the biomarker is systematically observed as increased or decreased in the elderly. The biomarkers may be obtained through different sources including blood, urine, saliva, spirometry, physical, or cognitive examinations.

Functions	Domain	Biomarkers
<b>Physiological</b>	Cardiovascular system	↑Systolic blood, Diastolic blood and ↑Pulse pressure, Resting pulse rate, ↑Total homocysteine, ↑Total cholesterol, High-density Lipoprotein cholesterol, Triglycerides
	Organs	↓Creatinine, ↓Cystatin C, Electrocardiogram, ↓Forced expiratory volume
	Glucose metabolism	↑Fasting glucose, ↑Glycated hemoglobin (HbA1C)
	Body composition	Body Mass Index, Waist-to-hip ratio, ↓Muscle mass, ↓Bone density
<b>Immune system</b>	Inflammatory factors	↑C-reactive protein (CRP), ↑Interleukin-6 (IL-6), ↑Fibrinogen, ↓Albumin, ↑Tumor necrosis factor $\alpha$ , ↑Serum amyloid A, T cell counts, Blood cell subtype counts and phenotypes
<b>Endocrine</b>	Hypothalamic-pituitary axis, growth hormones, sex hormones and other	Cortisol, ↓Dehydroepiandrosterone sulfate (DHEA-S), Insulin-like growth factor-1 (IGF-1), Norepinephrine, Epinephrine (adrenaline), Ghrelin, Melatonin, Somatostatin, Thyroid hormones, Leptin, Adiponectin, Testosterone, Estrogen
<b>Physical capability</b>	Strength, Balance, Dexterity, Locomotion	↓Grip strength, ↓Standing balance, Pegboard test, ↓Chair raising
<b>Cognitive</b>	Memory, Processing speed, Executive function	Digit symbol coding, Verbal and visual memory/learning

Arbeev et al. (2016) have underlined in their review that the best practice for evaluating the human aging rate in population studies is to use large sample sizes and multiple biomarkers measured in different time points (Arbeev et al. 2016). A Europe-wide population-based study MARK-AGE comprising 3200 participants with age range of 35-74 years is an example of such a project (Burkle et al. 2015). Instead of exploring single aging biomarkers, the project was founded in order to

establish weighted biomarker combination of human aging (i.e. a biological age score) from hundreds of aging-associated biomeasures. The study exemplifies the standard procedure of such a complicated analysis with a highly heterogenic human sample. For instance, the procedure includes considerations of anthropometric, clinical and demographic information of the study subjects. Specifically, functional (e.g. daily activities), cognitive, mood (e.g. Zung depression scale) and health (diseases, medications) statuses, lifestyle indicators (e.g. alcohol and tobacco consumption, exercise activity) and sociodemographic (marital status, occupation, education, household income) factors were also surveyed and adjusted for. (Burkle et al. 2015). Another large-scale biomarker-profiling example is a population-based study in Northern Europe, which showed that circulating alpha-1-acid glycoprotein, albumin, very low-density lipoprotein (VLDL) particle size, and citrate have clear short-term all-cause mortality predictor capability. The analysis was performed using high-throughput nuclear magnetic resonance spectroscopy characterization of 106 lipids, proteins and metabolites from non-fasting plasma samples of 17 345 individuals. (Fischer et al. 2014)

## 2.2.2 Longevity genetics

Survival analyses where long-living families are compared to the general population living in the same area have revealed that siblings of long-living subjects have higher survival rates at all ages. Aging-associated pathologies are more infrequent in these families. Centenarians are also geographically enriched in general population in several specific regions worldwide, such as Okinawa, Nicoya, Icaria, Sardinia, Loma Linda, Costa Rica, Greece, Italy and Mexico. All these observations suggest that longevity and healthy aging is influenced by a hereditary component, and this component may comprise genetics, culture, lifestyle and environment in general. Dissecting experimentally the magnitude of effect brought by environment and genetics and studying environment-gene-interactions in humans is highly challenging or even virtually impossible. For instance, the complexity of gene-environment-interaction is demonstrated in lifestyle (e.g. dietary habits), which may be transmitted to children in addition to the actual genes of the parents. (Dato et al. 2017) However, through lifespan manipulation experiments in animal models where the environment and genetical background is controlled, interaction between environment (dietary restriction) and genetics (IIS network) is a shown phenomena. (Lopez-Otin et al. 2013; Carmona and Michan 2016; Kaeberlein et al. 2016) In summary, genetics and longevity are linked together, but gathering experimental evidence of that link in human population is more challenging.

Studies have proposed that heritability of age at the time of death is between 15 and 30 percentages in adulthood (Brooks-Wilson 2013; Jylhava 2014; Dato et al. 2017). Furthermore, in humans, genetics appear to have a greater impact on the survival rate at ages above 60 years (Herskind et al. 1996; Skyttthe et al. 2003;

Hjelmborg et al. 2006), and be in the highest level at the oldest ages (Brooks-Wilson 2013). There are implications that genetics promoting longevity in age groups close to 90 years is different from that which is enhancing extreme longevity (>100 years) (Jylhava and Hurme 2010; Jylhava 2014).

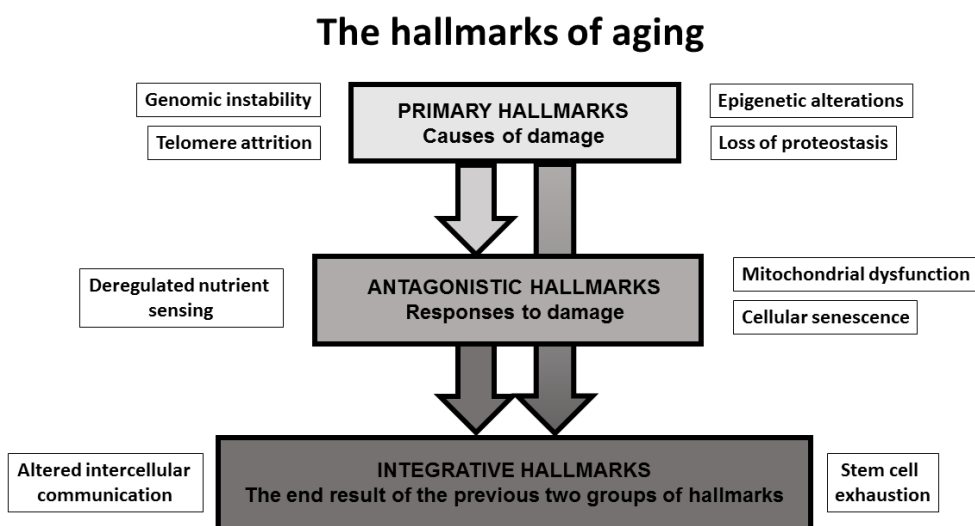
Kaerberlein et al. (2016) have summarized that, using a large variety of animal models, hundreds of genes have been associated with longevity or aging-resembling phenotypes. Based on these results, the genes may be clustered to certain pathways: *proteostasis, insulin and insulin-like growth factor 1 (IIS) pathway, mitochondrial metabolism, sirtuins, chemosensory function or dietary restriction*. In contrast, in human gene variant mapping studies, a substantially low number of variants have been even modestly associated with longevity. That is to say, the statistical significance is not strong enough and/or the longevity-association is not replicative. Moreover, rare gene variants are assumed to be contributing to human longevity and aging, and may sometimes be falsely undetected by genome-wide methods. (Kaerberlein et al. 2016) Commonly, risk gene variants are associated with lower survival, but there are also other aspects in this phenomenon. For instance, recent genome-wide genetic analyses have suggested that long-living individuals might also have risk gene variants for pathologies. Consequently, there might be additional factors (gene variants, epigenetics, lifestyle, environment) protecting these individuals from the impact of risk alleles. (Brooks-Wilson 2013; Jylhava 2014)

Through accordance with multiple linkage analyses using human families and genome-wide association studies, a set of promising polymorphic regions, which are located in genes *RBMS3, CTDSPL, MB21D2, CAMK4, RCBTB1, RLN1, RLN2, MAPKAP1, PRKCB, GAS7, APO E*, and *ANGPT4* (according to genome assembly GRCh37.p13) have been reported (summarized in Table 1.3 in a textbook by Kaerberlein et al. 2016). Linkage analyses using long-living families have pointed out longevity-associated locus on chromosome 3. In addition, reviews by Brooks-Wilson 2013 and Jylhava 2014 summarize that polymorphisms located in *FOXO3A* and *AKT1* in the IIS pathway are reported to be longevity-associated repeatedly. The most acknowledged single-nucleotide polymorphism associated with successful aging is located in gene *APO E* (rs4420638 on chromosome 19). The gene codes apolipoprotein E, which participates in cholesterol metabolism. (Jylhava 2014; Kaerberlein et al. 2016) The *APO E*-association has been confirmed in a recent genome-wide association meta-analysis using tens of thousands of people of European descent. In addition to *APO E*, only one novel locus (rs2149954) in chromosome 5 was identified as longevity-associated in the study. (Deelen et al. 2014a)



### 2.2.3 The hallmarks of aging

In the fields of cell biology and aging research, Lopez-Otin et al. (2013) have proposed the concept of *hallmarks of aging*, which are the common biological determinates contributing to organismal aging (Figure 1). The hallmarks may be categorized hierarchically, but they also overlap each other. Primary damage-causing factors are *genomic instability, telomere shortening, epigenetic changes and unbalanced protein homeostasis*. Factors that respond to the primary damage include *loss of regulation in nutrient sensing, mitochondrial dysfunction and cellular senescence*. The outcome hallmarks that are the consequences of the latter hallmarks and comprise *stem cell exhaustion and altered intercellular communication*.



**Figure 1.** The hallmarks of aging are suggested and described by Lopez-Otin et al. (2013). In this proposal, epigenetics is one of the hallmarks that is causing cellular damage in aging. Categories and the hierarchy of them is constructed here in highly artificial manner to provide framework for the aging research. For instance, these determinates overlap each other.

The primary cause(s) of aging is difficult to differentiate from the consequences of aging. In order to categorize the biological key features contributing to aging, Lopez-Otin et al. (2013) made strict rules for the hallmarks shown in Figure 1. A precise definition for these aging hallmarks should encompass the following: “(1) it should manifest during normal aging; (2) its experimental aggravation should accelerate aging; and (3) its experimental amelioration should retard the normal aging process and hence increase healthy lifespan.” (Lopez-Otin et al. 2013) It is worth noting that DNA methylation, which is the main topic of this thesis, does not fill these criteria of being a hallmark of aging alone, and is presented separately in Section 2.4.

### 2.2.3.1 Causes of damage in aging

In the hallmarks of aging (Lopez-Otin et al. 2013), the damage-causing factors are *genomic instability, telomere shortening, epigenetic changes and unbalanced protein homeostasis* (Figure 1). Numerous factors threaten genomic stability. Exogenous threatening factors include physical, chemical and biological mediators, while endogenous factors include DNA replication errors, reactive oxygen species, and DNA hydrolysis. Point mutations, translocations, deletions, telomere attrition, and virus or transposon insertions are examples of manifestations of genomic instability. To prevent these events from occurring, targeted repair machineries (e.g. base excision repair, homologous recombination, nucleotide excision repair, non-homologous end-joining, mismatch repair, DNA methylation maintenance, telomerase) have been developed. Studies have shown that malfunctioning repair systems and accelerated aging are linked. For instance, deficient DNA repair mechanisms are associated with progeroid syndromes such as Werner, Bloom Cockayne and Seckel syndromes with premature aging. (Lopez-Otin et al. 2013) A transgenic mice experiment has underlined the causality of this link between repair systems and accelerated aging. The *BubR1* gene participates to check points of mitosis, and thus the gene contributes to precise chromosome segregation. When *BubR1* was over-expressed, the lifespan of the mice was prolonged and their cancer incidence was decreased. (Baker et al. 2013) Studies with other progeroid syndromes, for instance, Hutchinson-Gilford and Néstor-Guillermo syndromes have demonstrated that these pathologies are affected by mutations in genes encoding protein components for nuclear lamina. As reviewed by Lopez-Otin et al. (2013), during normal aging, expression of an aberrant prelamin A isoform, called progerin, is increased, and, in addition, a study has shown that progerin is over-expressed in cells with telomere malfunction (K. Cao et al. 2011). Moreover, manipulation of prelamin A and progerin levels, or somatotrophic axis through e.g. NF- $\kappa$ B may extend the lifespan of the progeroid mice (Marino et al. 2010; Osorio et al. 2012).

Telomere attrition is a special case of DNA damage. In many cell types, the lifespan of a cell culture is limited because of the limited number of proliferations. This limitation, called Haylick's limit or replicative senescence, is caused by telomere shortening, which occurs gradually alongside each cell division. Telomeres shorten because the DNA replication machinery is unable to appropriately copy the linear telomere sequence at the end of the chromosomes (the end replication problem). An enzyme, telomerase, is responsible for maintaining or even increasing telomere length, and the enhanced activity of telomerase may be used to create an immortal cell line. In addition, in healthy situations, the shelterin complex is sealing the telomeres, and thus the DNA repair machinery does not recognize telomeres as problematic breaks in DNA. Deficient telomerase or shelterin complex may increase the risk of developing pathologies (e.g. dyskeratosis congenital or aplastic anemia) where the regenerative capacity of different tissues is lost. Moreover, the causal relationship between normal aging, telomere shortening and cellular senescence has

been shown using genetically engineered mice; their lifespan is extended when telomeres are lengthened or shortened when telomeres are shortened. (Lopez-Otin et al. 2013; Bojesen 2013; Cerella et al. 2016) In humans, shorter leukocyte telomere length has been associated with a higher risk of death in younger adults as well as in nonagenarians in the Leiden Longevity Study (1580 subjects 30-80 years of age; 870 subjects >90 years of age). Interestingly, telomere length was linked to mortality independent of the immune system-related markers (e.g. IGF-1, CRP, IL-6, CMV or blood cell counts). (Deelen et al. 2014b)

The proteostasis in the cell is impaired in aged organisms and is one of the aging hallmarks (Figure 1). The normal situation of proteostasis comprises the well-controlled homeostasis of protein quantities, conformation, binding interactions, and protein localization in the cell. Polypeptides that are unfolded due to cellular stress are either refolded by chaperone-mediated system, or degraded by ubiquitin-proteasome or lysosomal pathways. (Lopez-Otin et al. 2013) With these mechanisms, the expression of unfolded, misfolded or aggregated proteins are prevented. In aging-associated pathologies such as Alzheimer's and Parkinson's diseases, the expression of mistakenly folded proteins is observed. (Basaiawmoit and Rattan 2010; Lopez-Otin et al. 2013) For instance, chaperone-mediated protein folding weakens with age, and in flies and worms, the overexpressed chaperones have been linked to extended lifespan (Walker and Lithgow 2003; Morrow et al. 2004). However, the chaperone-connection to longevity appears to be more complex in higher animals, as shown in the long-living dwarf mouse model (Swindell et al. 2009).

### 2.2.3.1.1 *Epigenetics*

Epigenetics (Figure 2) may be categorized as one of the primary hallmarks of aging that are causing cellular damage in aging (Figure 1). In its entirety, Lopez-Otin et al. (2013) have described the hallmarks of aging, and these hallmarks are the biological determinates that are contributing to human aging.

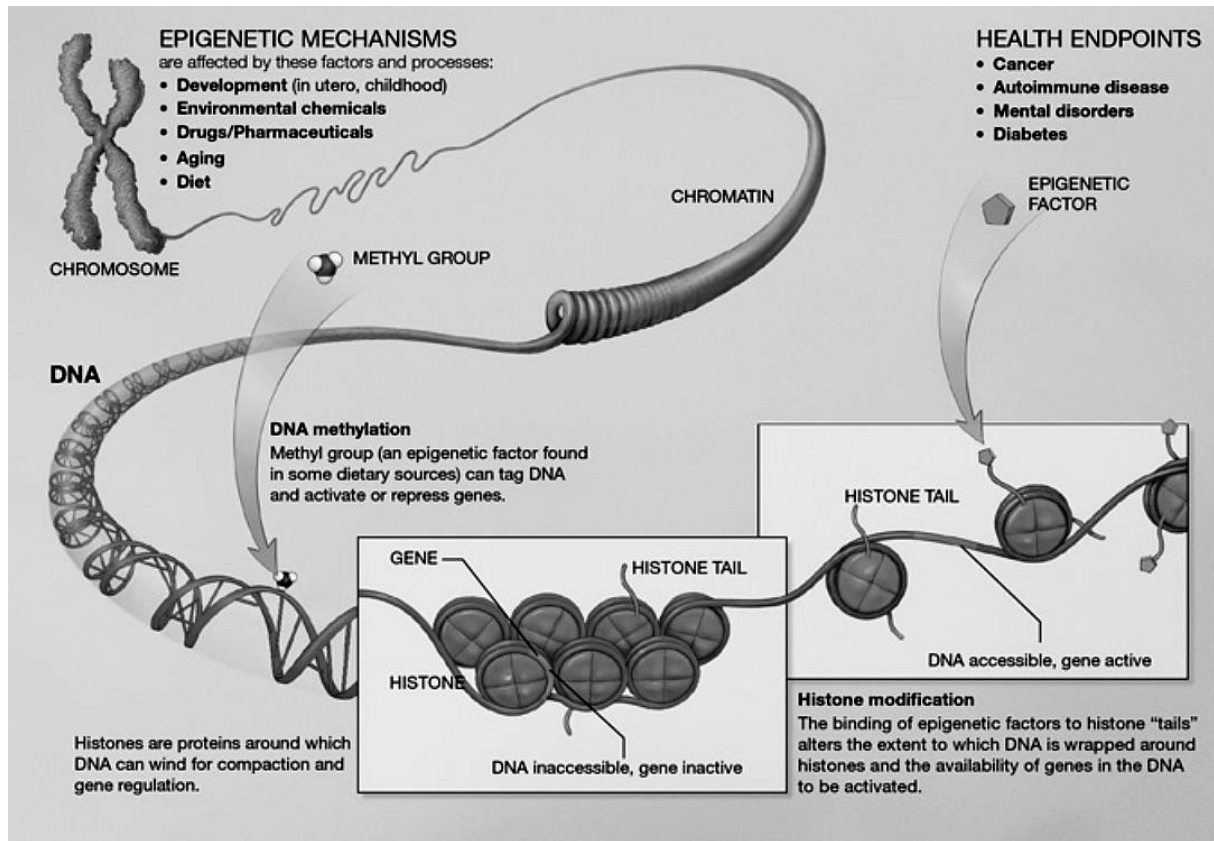
Epigenetics is a field in science, which, through the Greek prefix *epi*, is literally “on top of” traditional genetics. Epigenetic regulators provide an inheritable connection link between genome and surrounding environment. Overall, epigenetic regulation is manifested in gene expression. Distinct functions of different cell types are enabled with the precise activation of cell type specific gene expression patterns even though the genomic DNA sequence can be identical from one cell type to another. Gene expression may be initiated when the transcription factors bind to regulatory DNA sequences. In order to control the output of genome appropriately, interaction between transcription machinery and different epigenetic regulators is needed. Epigenetic regulation comprises mechanisms such as DNA methylation, which is the primary focus of this thesis and presented in Section 2.4, chemical modifications of the histones, operation of noncoding RNAs and nucleosomal remodeling. Epigenetics alters DNA packaging and modifications on DNA without

changing the DNA sequence itself, and importantly, all these epigenetic modifications are either inheritable to the daughter cells or possibly even to the offspring. (P. A. Jones and Liang 2012; Spruijt and Vermeulen 2014; M. J. Jones et al. 2015; Schübeler 2015; W. Xu et al. 2016)

Changes in chromatin structure are manifestations of epigenetic regulation (Figure 2). The chromatin strand is largely packed tightly with nucleosomes, which are complexes of DNA and histone proteins (dimers of H2A, H2B, H3 and H4 units). Positively charged amino acids on tails of these histone domains are modified with specific chemical modifications such as acetylation, methylation, phosphorylation, and ubiquitination. In addition, histone GlcNAc-ylation, butyrylation, malonylation, and crotonylation have been recently presented as novel chemical modifications. The histone code of chemical modifications defines which DNA regions are “open” for transcription machinery initiation. Analogously, locally “closed” chromatin structure represses transcription. For example, acetylation can neutralize positively charged amino acids (most commonly lysine and arginine), and thus weaken the tightness between DNA and histones which enables genes to be transcribed. In addition to acetylation, methylation of histone residues has been intensively studied. This modification may enable up- or downregulation of gene expression by trimethylation of lysine 4 on H4 or trimethylation of lysine 27 on H3, respectively. (P. A. Jones and Liang 2012; Schübeler 2015; W. Xu et al. 2016)

Longevity research has shown that epigenetics is a hallmark of aging (Lopez-Otin et al. 2013). In aged cells, histone H4K16 acetylation, H4K20 trimethylation or H3K4 trimethylation are increased, and H3K9 methylation or H3K27 trimethylation are decreased (Fraga and Esteller 2007; Han and Brunet 2012). In *C. elegans* and *Drosophila*, components in histone methylation complexes (H3K4 and H3K27, respectively) are regulating lifespan (Siebold et al. 2010; Greer et al. 2010). In addition, a study with *C. elegans* demonstrated that histone demethylase *Utx-1* inhibition, which is targeted for H3K27, decreased IIS activity in parallel with increasing survival (Jin et al. 2011). In contrast, DNA methylation has not, as such, been linked with experimental evidence to the extended lifespan of an organism and thus does not fulfill the criteria of being an independent hallmark of aging (Lopez-Otin et al. 2013).

Noncoding RNAs are also epigenetic regulators, and interplay with other epigenetic modifiers (Figure 2). The noncoding RNA species are transcribed from DNA but are not used as a template for protein synthesis; all these species are networking in many processes in time and in cellular space. In many organisms, it has been directly shown that DNA methylation and histone modifications are manipulated by noncoding RNAs. Noncoding RNAs involved in epigenetics are categorized into two classes: the long (>200 nt) and the short ones (<200 nt; micro RNAs, small interfering RNAs and Piwi-interacting RNAs). The latter group of RNAs regulate at least messenger RNA transcript usage, control of transposons and chromatin modifications.



**Figure 2.** Epigenetic mechanisms including DNA methylation, histone post-translational modifications, and chromatin structure, are shown in this schematic presentation (NIH 2017). Possible outcomes in health condition and factors contributing to epigenetic landscapes are also presented. Epigenetics is changed in developmental stages, in aging process, and due to environmental factors.

A well-known example of long noncoding RNA function is involved in X chromosome inactivation. The inactivation process compensates the expression levels of X-chromosome-linked genes in female mammals. This occurs because females carry two X chromosomes while males have Y chromosome in addition to one X chromosome. When compared to males, without this compensation procedure, females would have twofold expression levels of X-linked genes. In female mammals, two large RNA molecules, Xist and Tsix, work together to ensure that the other randomly chosen X chromosome is packed tightly during development, while the other X chromosome remains active. The Xist RNA-coated X chromosome is further transformed to an inactive and condensed chromatin state through different chromatin modifications, including DNA methylation of promoters located in X-linked genes. (Mohandas et al. 1981; Avner and Heard 2001; Collins et al. 2011; P. A. Jones and Liang 2012)

Studies in mammals and other organisms have implied that intergenerational or even transgenerational epigenetic inheritance mechanisms exist. That is to say, a phenotype is transmitted to following generations (from parents to first, second, third or more generations of offspring) with epigenetic regulators (DNA methylation, histone modification and regulatory RNAs) that are not written to the DNA sequence. (Sharma 2013; Martos et al. 2015) In plants, the evidence for the existence of transgenerational epigenetic inheritance and its usage as adaptation mechanism to the changing environment is strong. Mobile regulatory RNAs and hormones have been shown to participate in the plant transgenerational inheritance system. (Hauser et al. 2011; Sharma 2013) In animals, intergenerational transfer of effects through epigenetics is reported in many cases in which exposure to environmental toxic materials or maternal malnutrition during pregnancy are involved. In humans, *in utero* exposure to smoking (Chatterton et al. 2017) or arsenic (Kaushal et al. 2017) have been linked to changes in fetal DNA methylomes and future phenotypes of the progeny. A study with Agouti mouse model underlined that the environment during the embryogenesis determines phenotype of cubs rather than uterine environment (described in more details in Chapter 2.4). (Morgan et al. 1999; Leenen et al. 2016) In another example of intergenerational inheritance, the effect of Dutch famine in 1944-45 during the first trimester of pregnancy is associated with offspring obesity in adulthood (Heijmans et al. 2008; Roseboom et al. 2011). The intergenerational inheritability of the human lifespan has been also connected to epigenetics through manifestations in DNA methylation profiles in the offspring of long-living fathers in a EWAS (Marttila, Kananen, Jylhava et al. 2015).

Some evidence from association studies support the existence of transgenerational epigenetic inheritance in animals (i.e. changed epigenetic profile is transferred from a parent over multiple generations via germ cells). For instance, in rats, vinclozolin, an anti-androgenic fungicide has been reported to disrupt male fertility in multiple following generations, in parallel with observed changes in the DNA methylome in the germline. Inheritance is possible when the epigenetic reprogrammings (including methylation erasures) during developmental stages

operate in such way that the the acquired epigenomic landscapes of previous generations is somehow “remembered”. (Sharma 2013; Sharma 2015) A recent review (Sharma 2017), suggests that mechanisms mediating environmental exposure through germline to the progeny involve epigenetic regulators such as small noncoding RNAs, DNA methylation, and histone modifications. However, this complex inheritance system is still poorly understood.

### 2.2.3.2 Responses to aging-associated cellular damage

The aging hallmarks (Figure 1), *loss of regulation in nutrient sensing, mitochondrial dysfunction and cellular senescence* are responding in a compensatory or antagonistic manner to the aging-associated cellular damage. Their intensity levels define whether the effect is beneficial or accelerates aging. This means that, at low level, these hallmarks are beneficial systems, but overscaled intensity may be harmful. (Lopez-Otin et al. 2013)

Environmental factors such as nutrition, radiation, temperature and osmotic levels may influence the aging process (Lopez-Otin et al. 2013; Kaeberlein et al. 2016). An approach called caloric (or dietary) restriction is the most effective known lifespan manipulation method, and has been shown to modulate all the aging hallmarks at the molecular level. The lifespan-stimulating effect of the caloric restriction method in mammals had already been shown by the 1930s, using rats with 30% lowered daily calories. Since then, numerous species have been examined with caloric restriction diets. The lifespan of rodents has been extended by more than 50%, and the lifespan of lower animals has even been multiplied. Nonetheless, in humans, the lifespan extension through diet manipulation is yet to be shown. Instead of extension, many experiments or interventions are associated with improved health conditions. (Carmona and Michan 2016)

The evolutionarily conserved nutrient-sensing system is deregulated in aged organisms. This IIS network includes mTOR, AMPK and sirtuin (histone deacetylase) signaling pathways. The AMPK system detects high AMP concentration and sirtuins detect nicotinamide adenine dinucleotide (NAD<sup>+</sup>) concentrations, which are both implications of energy-deficiency in a cell; AMPK and sirtuins promote catabolism in low-energy situation. mTOR detects changes in amino acid quantities and responds to growth factor or environmental stress. In a high nutrient state, the role of IIS and mTOR pathways is to enhance anabolism. Increased anabolic activity accelerates the aging process. Lopez-Otin et al. (2013) concluded in their review that aging-associated cellular DNA damage is minimized by lowering organismal growth and metabolism. This is supported by the fact that growth hormone and insulin-like growth factor 1 are downregulated in aged organisms. Dietary restriction acts on IIS in similar way: it decreases the expression levels of anabolism-enhancing molecules in the ISS network. Multiple studies have shown that the IIS network contributes to longevity benefits gained from caloric restriction diet. (Lopez-Otin et al. 2013; Kaeberlein et al. 2016; Pan and Finkel 2017)

Genetic variability of the members in the glucose-sensing IIS pathway and their downstream targets are associated with longevity, both in humans and in animal models. Specifically, based on animal models, manipulated activities of the growth hormone, insulin-like growth factor 1 receptor, insulin receptor, mechanistic target of rapamycin complex (mTOR complex), protein kinase B (known as Akt) and members in FOXO family are all associated with extended lifespan. The research breakthroughs in the 1980s and 1990s showed that the lifespan of *C. elegans* may be doubled by blocking the function of *age-1* and *daf-2* genes only. The genes code human-resembling gene products, phosphatidylinositol-3-OH kinase and the insulin-like receptor in the IIS pathway, respectively. (Lopez-Otin et al. 2013; Jylhava 2014; Pan and Finkel 2017) Later, it was shown that the blockage of mTOR pathway using rapamycin-treatment increases the lifespan of laboratory mice (Harrison et al. 2009). Rapamycin is an immunosuppressive drug. In yeast, flies and worms, including rapamycin-addition to caloric restriction treatment extends lifespan more effectively than caloric restriction treatment without rapamycin. Yang et al. (2012) showed that aging-associated weight gain is related to hypothalamic mTOR expression, which is increased in aged mice. Injecting rapamycin directly to the hypothalamus reversed these changes (S. B. Yang et al. 2012).

Sirtuin enzymes (i.e. histone deacetylases) modify histones and other proteins post-translationally. The enzymes catalyze several reactions including NAD-dependent deacetylation as well as the demalonylation, desuccinylation and depropionylation of the target proteins. Sirtuins are able to mediate the environmental signal (e.g. energy-deficiency) to the epigenome through the removal of an acetyl group from a lysine residue on histones. (Lopez-Otin et al. 2013; Carmona and Michan 2016) In mammals, sirtuins are linked to the aging process. The mammalian sirtuins (*SIRT1-7*) localize in cytoplasm, nucleus and/or mitochondria. Increasing the expression of *Sir2* gene (near homolog of mammalian *SIRT1*) in *S. cerevisiae* or its orthologs in *C. elegans* and *D. melanogaster* increases their lifespan. Studies in yeast have also shown that, in order to gain advantages from a caloric restriction treatment, *Sir2* gene expression is essential. Nowadays, NAD<sup>+</sup>-dependent sirtuin family is set to the focus on anti-aging research, and sirtuin-activating compounds have been explored. Resveratrol, which is found in e.g. red grapes, is extensively studied and is an example of a sirtuin-mediated potential anti-aging compound. (Kaeberlein et al. 2016; Pan and Finkel 2017)

In higher organisms, such as mice or humans, longevity-associated experimental observations from sirtuins and other partners in nutrient-sensing are inconsistent. Many findings of lifespan-extending effects observed in worms or yeast appear to be more or less species-specific and replication attempts in higher animals have been unsuccessful. For instance, in spite of the high evolutionary conservation-percentage, there are major differences in the IIS pathway; in worms, large groups of insulin- and insulin-like polypeptides simultaneously regulate body size, cellular growths and metabolism, while in mammals, insulin growth factor 1 (IGF-1) and insulin signaling separately control growth and nutrient sensing. Therefore, the



replication of the experimental findings from one species to another is not a straightforward process. Interestingly, instead of a longer lifespan, some of the replication attempts have resulted healthier lifespan. (Lopez-Otin et al. 2013; Pan and Finkel 2017)

The initial role of cellular senescence is suggested to be a cellular checkpoint in tissue homeostasis and cancer-prevention. Based on this idea, in healthy young tissue, cellular senescence is a compensatory mechanism, which inhibits the proliferation of cells with profound damage in genomic stability or with oncogenic activity. For instance, telomere dysfunction leads to replicative senescence. In addition, cells may be induced to telomere-independent senescence via stressors such as oxidative stress. In aged organisms, the compensation mechanism that clears and replaces the senescent cells appears to dysfunction, and thus senescent cells accumulate during aging. The senescence program initiates when a certain damage threshold is reached. The program proceeds in parallel through the p53 and p16-Rb tumor suppressor pathways, and they can both independently lead to cell cycle arrest. (Lopez-Otin et al. 2013; Aravintan 2015; Cerella et al. 2016)

Mitochondria are also central players in aging. Mitochondria produce the majority of energy consumed in a cell (Carmona and Michan 2016). Mitochondrial DNA is considered to be more sensitive for violating agents than nuclear DNA because of insufficient repairing mechanisms and oxidative conditions in the mitochondria (Lopez-Otin et al. 2013). The free reactive oxygen species (ROS), which originate from mitochondrion, are more frequently observed in aged organisms. Already by the 1950s, Harman D. proposed that oxidative damage in the cell increases during aging. This is supported by numerous studies performed during the past decades. Later, it has has been discovered that oxidative stress iduces deletions and point mutations to mtDNA and these radically accumulate with age in wide variety of species. Single-cell experiments have shown that the main weight in mitochondrial heteroplasmy is switched to a mutated mitochondrial genome in the aging cell, i.e. the mutated genome is dominant over the original mitochondrial genome. (Harman 1992; Lopez-Otin et al. 2013; Vina et al. 2013)

Nowadays, however, the free radical theory has been partially questioned. Unexpectedly, studies with yeast and flies in the 2000s implied that increased levels of ROS may actually extend lifespan. Moreover, multiple studies in mouse models have shown similar results. (Lopez-Otin et al. 2013) Nonetheless, mitochondria are strongly affected by aging. Specifically, Dufour et al. (2000) have underlined in their study the causal link between longevity and mitochondrial DNA. Disruption of *COX5* gene coding for the fifth subunit of cytochrome c oxidase in mitochondrial DNA was shown to contribute longevity of the fungus *Podospora anserine*. The lifespan extension of the model organism was tenfold and coupled with alternative oxidase pathway usage. (Dufour et al. 2000)

### 2.2.3.3 The integrative hallmarks of aging

The integrative aging hallmarks (Figure 1, *stem cell exhaustion* and *altered intercellular communication*) are the consequences of various types of aging-associated damage that initially cause the organismal aging. In aged organisms, the regenerative potential of tissues is weakened, because stem cell mobilization from the niches is decelerated. The stem cell exhaustion is also seen in the immunosenescence, in which hematopoiesis decreases with age, and this causes lower production of adaptive immune cells.

Analogous to hematopoietic inhibition, tissue regeneration is essentially changed in all aged tissues. For instance, lowered tissue regeneration in the mouse forebrain (Molofsky et al. 2006), the bone (Gruber et al. 2006), and the muscle fibers (Conboy and Rando 2012) is reported. Moreover, changes in telomere length and telomerase activity have been associated with altered epidermal stem cell behavior (Flores et al. 2005). Balanced stem cell production appears to be essential for healthy aging. Research with *Drosophila* has demonstrated that stem cell exhaustion and accelerated aging are caused by increased proliferation activity. (Lopez-Otin et al. 2013) Fascinatingly, Conboy et al. reported in 2005 that through heterochronic parabiosis pairing in mice, tissue-specific stem cells originating from the muscle or liver may be rejuvenated. The stem cells of a young animal adopted the older functional phenotype in the microenvironment of an older animal, and similarly, molecular signatures in stem cells of old mice were restored to a younger state through the younger microenvironment. (Conboy et al. 2005)

## 2.3 The human immune system

Decline in the *cell-to-cell signaling* systems including endocrine, neuroendocrine, and neuronal signaling is one of the integrative hallmarks of aging (Figure 1) (Lopez-Otin et al. 2013). Signaling from cell to another in space and time in orchestrated manner is crucial for the immune responses, and this system is altered during aging (Weiskopf et al. 2009).

The immune system protects the organism from extrinsic (bacteria, viruses) and intrinsic (e.g. tumoric cells) factors using innate and adaptive mechanisms. These cascade mechanisms use antigen-recognition system in order to differentiate between healthy and infected or otherwise stressed tissues to generate appropriate responses. This is the basic concept of immunology. (Murphy et al. 2008) In rough generalization, the innate mechanism operates mainly through the acts of epithelial cells, monocytes, macrophages, natural killer (NK) and dendritic cells, while the adaptive mechanism operates through B and T lymphocytes. Granulocytes (polymorphonuclear leukocytes) in the innate system include neutrophils, eosinophils and basophils. Before differentiation and specialization to certain functions, all white and red blood cells originate from pluripotent hematopoietic stem cells in the bone

marrow, and some of these cells differentiate and mature there, as well. The progenitor cells migrate from the bone marrow to the peripheral tissues and either reside within the tissue, circulate in the blood or circulate in the lymphatic system. The site of differentiation distinguishes B and T lymphocytes from each other (thymus and bone marrow, respectively) and their antigen-properties differentiate them from other leukocytes. (Murphy et al. 2008; Schroder and Tschopp 2010; Hawse and Morel 2014)

In a textbook example, a pathogen may invade human body through e.g. leaking intestinal cell membrane. The recognition site on a pathogen is bound to the germline-encoded pattern recognition receptors, such as Toll-like, scavenger, or mannose receptors on the innate immune cells. The binding with pathogen initiates cytokine and chemokine secretion (e.g. IL-1, IL-6, IL-8, IL-12, TNF- $\alpha$ ), which in turn activates more inflammatory pathways and cell-to-cell signaling cascades. As a result, more immune cells such as neutrophils and monocytes arrive to the infected tissue, and continue pathogen destruction through lysosomal degradation and secretion of reactive oxygen and nitrogen compounds. Monocytes, macrophages (mature form of monocytes), dendritic cells and neutrophils may eliminate the pathogen (or other debris in a sterile inflammation where the inflammation exists without influence of pathogens) via phagocytosis. Activated macrophages and dendritic cells produce and secrete chemokines and cytokines in order to invite leukocytes including eosinophils and lymphocytes. Specifically, virus invaders cause infected cells to produce interferons, which are needed to prevent viral replication. Furthermore, the interferone secretion recruits NK cells to destroy the infected cells. (Murphy et al. 2008; Schroder and Tschopp 2010; Hawse and Morel 2014)

Antigen-presenting cells such as dendritic cells and macrophages connect the innate and adaptive immune systems. In the adaptive immune system, memory B and T cells in the blood and lymph node assure long-term immunity. Antigen-presenting cells may activate naïve T cells by presenting novel antigen in association with major histocompatibility complex (MHC). Peptide-MHC molecule is recognized by T cell receptor (TCR), which is located on the surface of CD4<sup>+</sup> or CD8<sup>+</sup> T cells. (Li et al. 2013) Activated CD4<sup>+</sup> (helper) T cells begin secretion of IL-2, IL-4 and IFN- $\gamma$ . Next, the T cells differentiate to Th subpopulations that recognize the same cytokine than previously. The helper-Th populations recruit cytotoxic CD8<sup>+</sup> T cells, modulate antibody production of B cells, activate macrophages and prevent autoimmunity. (Geginat et al. 2013; Hawse and Morel 2014; Tu and Rao 2016)

As a whole, the number of circulating T cells appears to be constant over time in an individual. The main categories of the circulating T cells differ functionally and phenotypically from each other, and are termed as central memory T cells and effector memory T cells. Majority of the central memory T cells are enriched with helper CD4<sup>+</sup> T cells while the latter category comprises mostly cytotoxic CD8<sup>+</sup> T cells. (Geginat et al. 2013; Hawse and Morel 2014; Tu and Rao 2016)

### 2.3.1 Aging of the immune system

The innate and adaptive immune systems are altered with age. In practice, this change is demonstrated by weakened responsiveness to vaccination in the elderly (Goodwin et al. 2006). The vaccination challenge is linked to the progressive decline in the immune system functioning, which is also termed as *immunosenescence* (Pawelec 2017). Immune cell distribution-related data from other tissues than blood is sparse and thus, here, the focus is set to blood.

In aging, adaptive immunity is weakened while higher-level innate immune functions are preserved or increased. In the innate immunity, the inflammatory background and the number of NK cells are increased, while functions of macrophages, the antigen presentation to T cells and delivery to the lymph node are decayed. The adaptive immunity is modified with age, as well. Thymic function declines, number of naïve T cells on the periphery is reduced and primary T cell responses to foreign, novel antigens are decreased. Effector memory T cells becomes more abundant and the T cell repertoire in general is restricted. (Weiskopf et al. 2009; Pawelec et al. 2010; Franceschi and Campisi 2014; Tu and Rao 2016) The most well known change in blood cell composition seen in the elderly is the decreased percentage of naïve CD8<sup>+</sup> T cells in the blood (Pawelec 2017). Specifically, extensively differentiated CD28<sup>-</sup> T cells become more abundant. CD28 is a co-stimulatory molecule, which is essential for T cell activation. Loss of CD28 is associated with infection vulnerability, and the lowered responsiveness to vaccination is linked to CD28 loss. In addition to T cells, the aging-associated changes include shifts in B cell proportions and the B cell functions are altered. (Weiskopf et al. 2009; Pawelec et al. 2010; Franceschi and Campisi 2014; Tu and Rao 2016)

Aging-associated changes in intercellular communication is one of the integrative aging hallmarks (Figure 1) (Lopez-Otin et al. 2013). Failed intercellular communication is implicated also in nonresolving low-grade inflammation state termed *inflammaging*. This condition is a phenotype, which occurs without microbial influence. At the beginning in 2000s, Franceschi et al. (2000) suggested that inflammaging might predispose to aging-associated diseases and adverse health conditions, and this has now been widely accepted. Increased serum cytokine (e.g. IL-6, TNF- $\alpha$  or CRP) levels that are usually at low level are the typical characteristics of inflammaging state (Krabbe et al. 2004; Singh and Newman 2011). Studies have linked inflammaging with increased mortality and pathologies such as Alzheimer's disease, atherosclerosis, heart disease, type II diabetes and cancer. (Franceschi and Campisi 2014; Xia et al. 2016).

The factors that are causing inflammaging are still unknown due to lack of solid experimental evidence. However, extensive research is pointing to various mechanisms that might underlie inflammaging. Stress, oxidation-inflammation, cytokines, DNA damage, autophagy and stem cell aging are the main contributors in these theoretical models (Xia et al. 2016). Specifically, characteristics of

inflammaging show over-expression of genes related to inflammation and immune responses (de Magalhaes et al. 2009; Swindell 2009), and NF- $\kappa$ B activation (Helenius et al. 1996; Adler et al. 2007; Salminen et al. 2008). In addition to NF- $\kappa$ B, which controls inflammatory responses as a master regulator, the proposed mechanisms of inflammaging include Ras, Notch, TGF- $\beta$ , RIG-I, mTOR, and sirtuin signalling pathways (Xia et al. 2016).

### 2.3.2 Cytomegalovirus infection

Latent persistent human cytomegalovirus (CMV) infection appears to be contributing to the aging of the immune system. The cytomegalovirus is a DNA-based  $\beta$ -herpesvirus that has co-evolved alongside human evolution. CMV infects most humans at some stage of life, and in healthy individuals, the CMV infection is mainly subclinical. Primary CMV infection evokes both innate and adaptive immune responses, and CMV has developed survival mechanisms in which MHC functioning is restricted. This enables the virus to escape from antigen presentation. During primary infection, the main response to CMV is formed by CD4<sup>+</sup> T cells which secrete Th1 cytokines (e.g. IFN- $\gamma$  and TNF- $\alpha$ ). B cells are also recruited during the response reaction. After days of infection, CD8<sup>+</sup> T cells populate the blood, and are responsible for protecting the host from CMV re-infection and re-activation. (Tu and Rao 2016)

CMV has been linked to aging-related pathogenesis in epidemiological studies. The quantity of CMV titer is associated with frailty and functional impairment of the body (Wang et al. 2010; Moro-Garcia et al. 2012), as well as all-cause and cardiovascular mortality (Strandberg et al. 2009; Roberts et al. 2010; Gkrania-Klotsas et al. 2013). In the long-term, even in healthy humans, CMV infection seem to drive notable changes in the T cell repertoire and functionality, and the absolute anti-CMV titer is shown to correlate with the T cell proportions (Wertheimer et al. 2014). In CMV<sup>+</sup> humans, the proportion of CMV-specific memory CD8<sup>+</sup> T cells in the peripheral blood increases during aging in such way that, in the elderly, half of the memory CD8<sup>+</sup> T cells appear to be CMV-specific. CD4<sup>+</sup> T cells are also shown to be affected, but to a lesser extent, as  $\sim$ 30% of those cells show CMV-responsiveness at the old ages. This major change in the T cell proportions has been considered to be the driving force in immunosenescence. Frequently, the CMV-specific T cells become terminally differentiated to CD45RA<sup>+</sup>CD57<sup>+</sup>CD28<sup>-</sup>CCR7<sup>-</sup> cells, and this transformation is a characteristic phenotype for aging-related senescent T cells. (Sylwester et al. 2005; Pourghesari et al. 2007; Pawelec et al. 2012; Alonso Arias et al. 2013; Tu and Rao 2016)

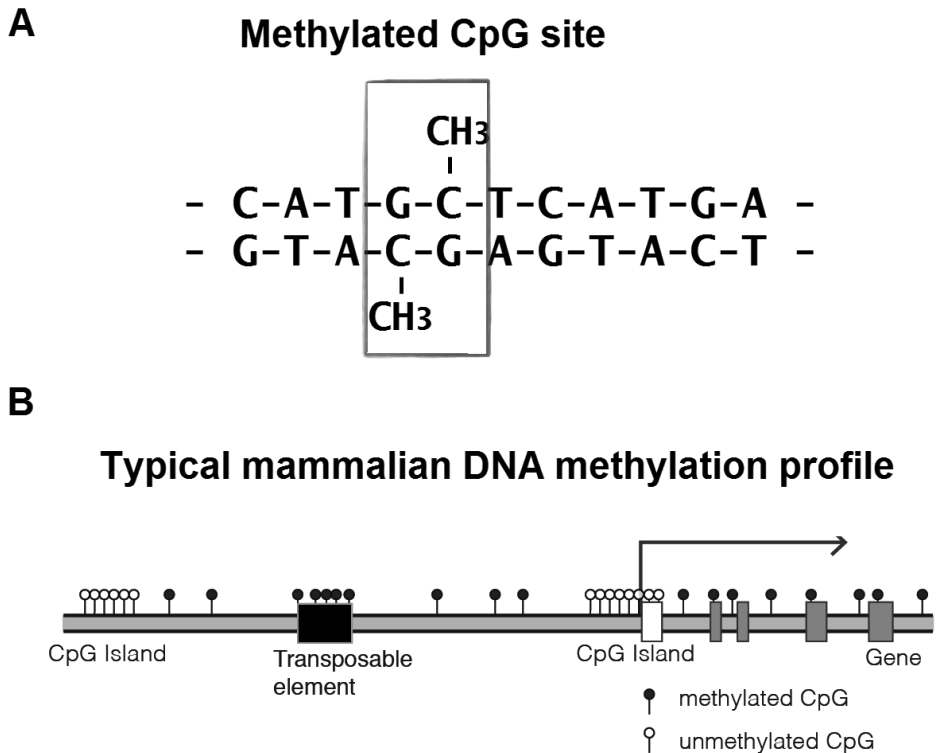
## 2.4 DNA Methylation

The role of DNA methylation in developmental biology came to the fore in the late 1960s and 1970s. During this era, it was proposed that gene activities are switched on and off, and the inheritability of this gene activity is based on the enzymatic methylation of cytosine in the DNA sequence. (Holliday and Pugh 1975; Riggs 2002; Holliday 2006) Nowadays, these suggestions are supported by a wide variety of evidence, and many functions of this epigenetic mark have been unraveled in greater detail. DNA methylation is a mechanism which is capable of blocking transcription initiation, and is now associated with long-term gene expression silencing (such as X chromosome inactivation, imprinting and tissue-specific gene expression). Overall, DNA methylation has a critical role in cell differentiation and development. (P. A. Jones and Liang 2009; P. A. Jones 2012; B. C. Christensen et al. 2012; Spruijt and Vermeulen 2014; Rasmussen and Helin 2016; Elhamamsy 2016)

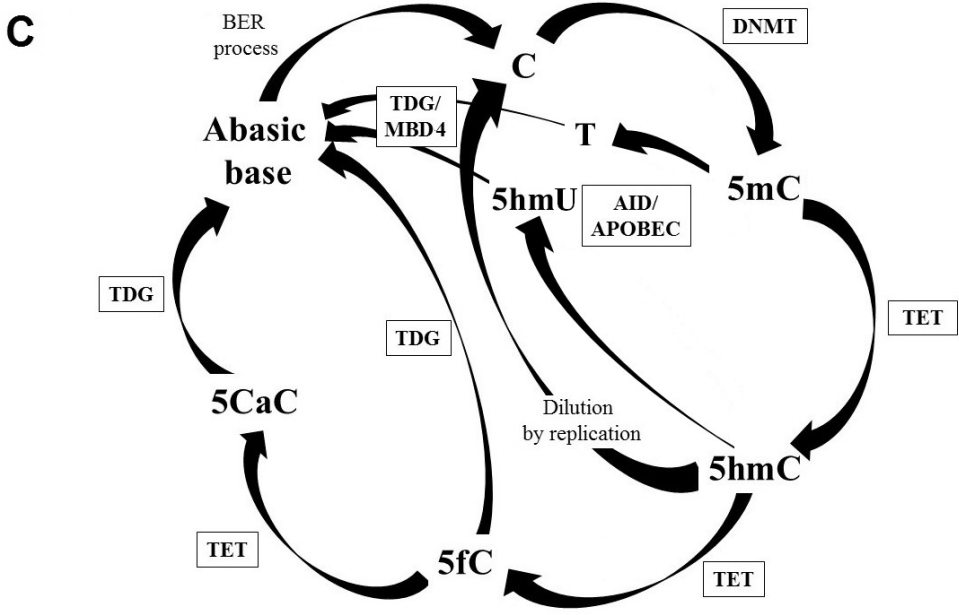
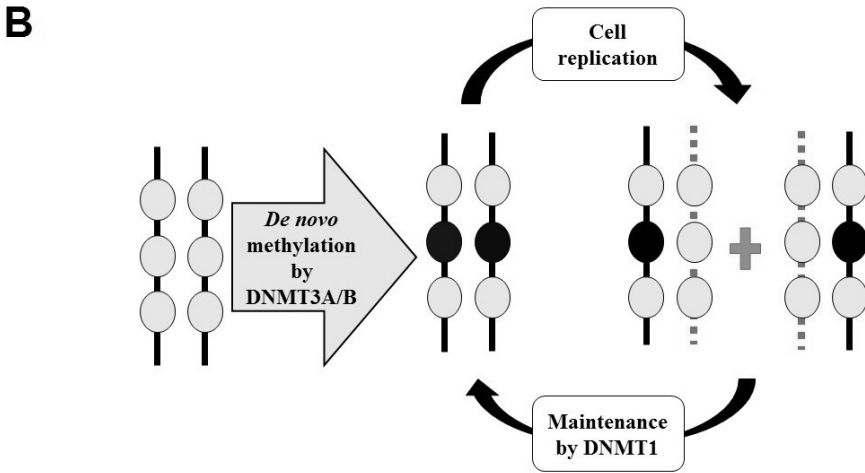
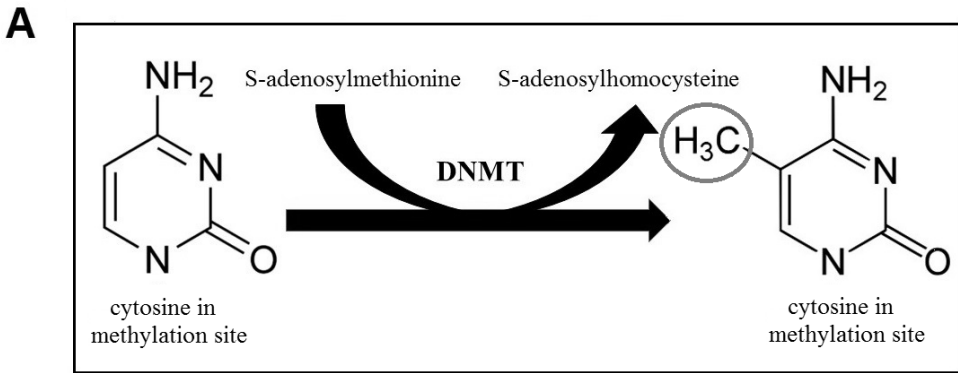
In general, the genome-wide DNA methylation profile is stable, and DNA methylation profiles differ between cell types and tissues. DNA methylation contributes to the overall genomic stability through centromeric, pericentromeric, subtelomeric and telomere length-controlling systems. There are approximately 28 million CpG sites in the human genome, and 60-80% of those are methylated. In mammals and many other organisms, the DNA sequence contains high amounts of 5-methylcytosines, i.e. methylated cytosines (5mCs), which are primarily located in the sequence CpG (Figure 3A). The notion CpG stands for 5'—C—phosphate—G—3', where cytosine and guanine are separated by one phosphate. (Holliday 2006; P. A. Jones and Liang 2009; B. C. Christensen et al. 2012; P. A. Jones 2012; Smith and Meissner 2013; Spruijt and Vermeulen 2014; Zampieri et al. 2015; W. Xu et al. 2016; Rasmussen and Helin 2016; Elhamamsy 2016) The observed occurrence of CpG sites is only ~25% of the expected frequency in mammalian genomes (Simmen 2008; Elhamamsy 2016). 5mC is prone to conversion to thymine spontaneously or enzymatically (Figure 4C), and it is suggested that a large proportion of the CpGs have been lost via mutations during evolution. However, the DNA sequence comprises CpG islands (CGIs), the CpG-rich DNA sequences with GC content >50% spanning ~500 – 1000 base pairs (Figure 3B). (P. A. Jones 2012; Smith and Meissner 2013; Rasmussen and Helin 2016; Elhamamsy 2016)

A large proportion of CGIs locate at gene promoters, and they are often hypomethylated (meaning that they are less methylated). In a textbook example of functional consequences, DNA methylation located in a gene promoter typically represses gene transcription (Figure 3B). An increase in methylation level is termed as hypermethylation. Most of the gene bodies are methylated and contain fewer CpG sites, but CGIs located to the gene bodies do exist. The gene bodies may contain repetitive sequences and transposable elements (such as *Alu* and *LINE1*), and methylation is considered as being the mechanism that silences these elements while enabling appropriate genes to be expressed (Figure 3B). (Holliday 2006; P. A. Jones and Liang 2009; B. C. Christensen et al. 2012; P. A. Jones 2012; Smith and Meissner

2013; Spruijt and Vermeulen 2014; Zampieri et al. 2015; W. Xu et al. 2016; Rasmussen and Helin 2016; Elhamamsy 2016)



**Figure 3.** Principles of DNA methylation. A) Methylated CpG site in DNA double strand. B) Typical DNA methylation profile includes GC-rich regions termed CGIs (Mariuswalter 2016). Promoter-associated CGIs are usually unmethylated and promoter region is available for transcription initiation. Transposable elements are repressed by dense DNA methylation. Symbols: CH<sub>3</sub> = methyl group, CGI = CpG island





**Figure 4.** DNA methylation and demethylation pathways. A) A methyl group ( $\text{CH}_3$ ) is transferred from S-adenosylmethionine that is provided by one-carbon cycle.  $\text{CH}_3$  is bound covalently to cytosine in DNA stand. DNA methylation is a trans-methylation event; B) A simplified model summarizes the different principal functions of DNA methylation transferase enzyme 1 and 3A/B (DNMT1 and DNMT3A/B, respectively) in DNA methylation (modified from Jones and Liang 2009). Typically, de novo methylation establishes new methylation marks to the DNA while maintenance methylation with DNMT1 takes place during cell replication. By maintenance methylation, hemimethylated sites in newly synthesized DNA double strands are transformed to symmetrically methylated positions to match the situation in original template DNA; C) Suggested demethylation pathways and restorations to cytosine (modified from Kohli and Zhang 2013 and Elhamamsy 2016). Details of these pathways are described in the main text. Symbols in Figure 4B: black circle = methylated CpG site, grey circle = unmethylated CpG site, black line = original template DNA strand, dashed line = newly synthesized DNA strand; Symbols in Figure 4C: BER = base excision repair process, TET = ten-eleven translocation enzyme, TDG = thymine-DNA-glycosylase, AID = activation-induced cytidine deaminase, APOBEC = apolipoprotein B mRNA editing enzyme, MBD4 = methyl-CpG binding protein domain 4, DNMT = DNA methylation transferase enzyme, 5mC = methylcytosine, 5hmC = 5-hydroxymethylcytosine, 5fC = 5-formylcytosine, 5caC = 5-carboxylcytosine

However, in spite of the extensive research, there are still many DNA methylation-related issues to be discovered. For instance, the role of DNA methylation at non-CGI regions and gene body methylation in alternative transcript splicing are unclear. The significance of other DNA methylation target regions such as C-X-G (where X is adenine, cytosine or thymine) are less studied. The dynamics between active demethylation and methylation pathways (shown in Figure 4) in gene expression regulation are unknown. The role of methylation in enhancer and insulator control remains to be solved. (P. A. Jones 2012) Moreover, the role of demethylation pathways and their intermediate products (e.g. 5-hydroxymethylcytosine) have questions that remain to be answered (G. Xu and Wong 2015).

In mammals, dynamic, highly organized methylation is crucial for normal ontogenesis. Genome-wide methylation profiles are reprogrammed, i.e. the methylation marks are erased and re-established during developmental stages (fertilization, implantation, blastocyst), and the global DNA methylation level fluctuates greatly (Schuermann et al. 2016). Another epigenetic reprogramming stage occurs in the development of germ cells. Nevertheless, some of the DNA methylation site-specific patterns are maintained completely or partially during epigenetic reprogramming in early embryogenesis and gametogenesis. Repetitive elements often maintain their methylation profiles untouched. A fraction of genes in the genome are imprinted, and these genes are reset in a parent-of-origin-specific manner at the later stage of epigenetic reprogramming, namely at the stage of primordial germ cell development. Therefore, imprinting enables monoallelic gene expression, in which only one of the alleles inherited from the paternal or maternal side are expressed. Imprinting is established by a complex process involving DNA methylation, histone modifications and non-coding RNAs. Many of the imprinted

genes are evolutionarily conserved and clustered together in the genome. Moreover, imprinting is tissue-specific. The mechanism controlling imprinted genes is unique, and depends on imprinting control regions, which are targeted to specific genes. In a textbook example, a control region is maternally methylated and paternally unmethylated. In this manner, the target gene expression is maternally silenced, but paternally expressed. (Sharma 2013; Martos et al. 2015; G. Xu and Wong 2015; Elhamamsy 2017)

#### 2.4.1 Maintenance and de novo methylation

The methylation pattern on DNA may be changed via de novo methylation or via demethylation; demethylation may occur either passively or actively with the help of enzymes, while addition of a methyl group (CH<sub>3</sub>) is always an active process (Riggs 2002; Xu and Wong 2015; Rasmussen and Helin 2016). 5-methylcytosines (5mCs) in CpG sites are generated symmetrically to both strands in the double stranded DNA by the DNA methylation machinery. DNA methyltransferases, DNMT1 and *de novo* DNMT3A/B are responsible for adding methyl groups to cytosine bases, such as presented in Figures 4A and 4B. Typically, *de novo* methylation establishes new methylation marks to the DNA while maintenance methylation takes place during cell replication (Figure 4B). DNMT3A/B shows equal affinity for hemimethylated and unmethylated DNA. In a hemimethylated CpG site, only the other cytosine in opposite strands is methylated. In other words, hemimethylated site is partially methylated position in the DNA double strand, where only opposite strand is lacking methyl group in that particular methylation site (such as shown in Figure 4B). The DNMT3A/B enzymes are expressed the most at the embryonic stage and downregulated but not shut down in somatic cells. Blocking DNMT3A/B in embryonic or neonatal mice is lethal. (P. A. Jones and Liang 2009; Smith and Meissner 2013)

Most of the genome-wide methylation profile is maintained during mitosis by the maintaining methylation machinery, a large complex of molecules including chromatin-associated enzymes. DNMT1 is a key player in the machinery, and it is expressed the most in the S phase of mitosis when the methylation maintenance event takes place. DNMT1 has higher affinity for hemimethylated site than for symmetrically un-methylated site in the DNA sequence. In the process of methylation maintenance, DNMT1 interacts directly with assisting molecules, proliferating cell nuclear antigen (PCNA) and Ubiquitin-like, containing plant homeodomain and RING finger domains-containing protein 1 (UHRF1). UHRF1 binds to target site located in original DNA strand, recruits the DNMT1, and guides the enzyme to operate on newly synthesized strand; DNMT1 adds the missing methyl group from the methyl group donor, S-adenosylmethionine (SAM) to the new strand in order to generate an identical copy of methylation to the new strand that mimics the old strand. (Lu 2000; P. A. Jones and Liang 2009) The importance

of this DNMT1/PCNA/UHRF1-complex was underlined by a recent study where it was shown in several (non-tumorigenic) cell lines that global hypomethylation, which is considered as hallmark of tumorigenesis is induced by disturbance of DNMT1/PCNA/UHRF1 (Pacaud et al. 2014).

Figure 4B illustrates straightforward models of introducing and maintaining DNA methylation patterns, but these processes are much more complex, and many details are unsolved. Functions of DNMT1 and DNMT3 enzymes appear to overlap to some extent. For example, it is suggested that DNMT3A/B enzymes also participate in maintenance methylation during mitosis. DNMT1 is the main maintenance methylator for most of the CpG sites, but DNMT3 may also repair mistakenly un-methylated sites, which are located inside nucleosomes (P. A. Jones and Liang 2009). In this suggestion, DNMT3 is bound stably to chromatin in nucleosomes which contain methylation sites; the enzyme behaves in similar way as other well-known chromatin-modifying enzymes, such as the enhancer of zeste homolog 2 (EZH2) (P. A. Jones and Liang 2009). EZH2 is a methyltransferase that is localized close to the methylation site and catalyzes histone H3 lysine 27 methylation (R. Cao et al. 2002).

## 2.4.2 Demethylation

Demethylation pathways (Figure 4C) have the potential to alter systematically DNA methylation profiles as a response to environmental stimuli. During embryogenesis, active and ordered demethylation is essential. Currently, detailed knowledge of demethylation pathways is lacking and the research is actively ongoing. However, many models have already been suggested. The enzymes which are associated with the active demethylation pathways are called the ten-eleven translocation (TET) enzymes, and were discovered fairly recently (Tahiliani et al. 2009; Kohli and Zhang 2013; Rasmussen and Helin 2016). Soon after, it was shown that oxidation of 5-methylcytosine (5mC) with 2-oxoglutarate, Fe (II) and catalyzation by TETs produces 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) (He et al. 2011; Ito et al. 2011; Schuermann et al. 2016). These cytosine modifications base pair in a normal way with guanine. After 5mCs, 5hmCs are the most common cytosine modifications found in the human genome. TET enzymes have a fast rate constant to produce 5hmCs while their catalytic activity is much lower to produce 5fCs and 5CaCs. In addition to TETs, the thymine-DNA-glycosylase (TDG) and base-extension-repair (BER) enzyme are important catalyzation factors in suggested model(s) of active demethylation pathway(s) (Figure 4C). TDG is able to recognize the more oxidized forms of cytosine, 5fCs and 5CaCs. The glycosylase brakes their N-glycosidic bonds and releases the base with an abasic site. Following this, the BER process repairs the cytosine to its unmodified state, which may be methylated to 5mC.

At present, characterization of binding partner proteins for the oxidized cytosines are explored. Even though 5hmC is shown to be enriched in post-mitotic brain tissue, the independent roles of 5hmCs, 5fCs and 5CaCs as epigenetic regulators are yet to be fully determined and confirmed. (Kohli and Zhang 2013; Rasmussen and Helin 2016; Schuermann et al. 2016)

Another suggested active deamination pathway is based on the deamination of 5-methylcytosines and 5-hydroxymethylcytosines, which are transformed to thymine and 5-hydroxymethyluracils, respectively (Figure 4C). The process is catalyzed by activation-induced deaminase (AID) and apolipoprotein B mRNA editing enzyme (APOBEC). At the end of this pathway, the abasic site is formed when deamination-generated base pair mismatch is identified and cleaved by TDG and methyl-CpG binding protein domain 4 (MBD4). Finally, as in TET-oriented demethylation pathway, BER (base excision repair) process repairs abasic sites and restores them to cytosines. (Elhamamsy 2016)

The passive demethylation results from a spontaneous and unintended failure of DNA methylation maintaining machinery, which is responsible for transforming hemimethylation to symmetrically methylated cytosines. An additional suggested model for DNA demethylation is a partially active pathway where TET is first generating 5hmCs, which are then passively diluted away during following replications (Figure 4C). (Rasmussen and Helin 2016)

### 2.4.3 DNA methylation and environmental factors

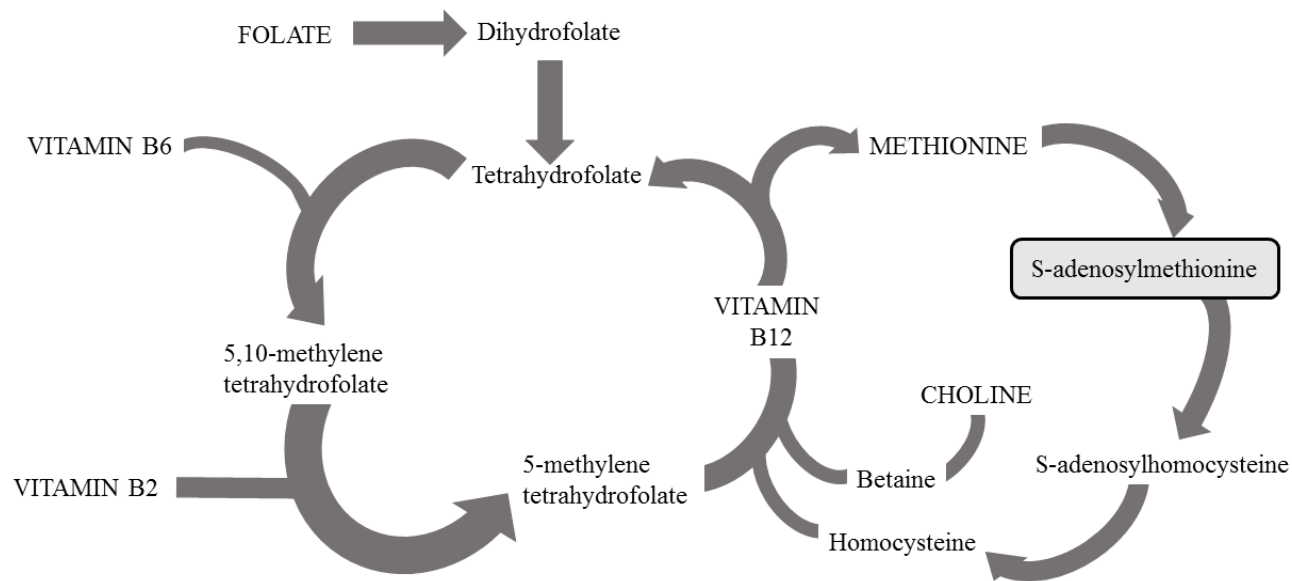
Environmental factors such as diet, smoking, toxins and stressful environment shape the DNA methylome (Zampieri et al. 2015). The causal connection between environment, epigenome and phenotypic variation during development has been shown in a functional study in the Agouti ( $A^{vy}$ ) mouse model. The mice coat color ranges between yellow and brown because of hypomethylation of the A locus in Agouti gene, where methylation-sensitive intracisternal-A particle retrotransposon has been inserted. Methylation level change from 70% to 20% shifts the coat color from yellow to brown. In the study, it was shown that the environment during the embryogenesis determines the coat color rather than uterine environment. The oocyte was transferred to a surrogate mother with opposite epigenetic phenotype, yet, despite this, the coat color depended on the phenotype and the epigenome of the biological mother. (Morgan et al. 1999; Leenen et al. 2016)

Regarding humans, one-carbon metabolism is an example of connections between diet and epigenetics. Dietary folate, vitamins B2, B6 and B12, choline and methionine act as methyl group donors and/or co-enzymes in the one-carbon metabolism (Figure 5), and therefore changes in their intake may alter epigenetic modifications. SAM is an intermediate in the one-carbon cycle and universal methyl group ( $\text{CH}_3$ ) donor for DNA, RNA, and many other biomolecules. Aging-associated deficiency of methyl group donors (folate and vitamin B12) in the one-carbon cycle

(Guéant 2016) is a phenomenon that highlights the relevance of the one-carbon metabolism regarding research focusing on aging-associated DNA methylation changes. The lack of methyl group donors leads to the accumulation of homocysteine and the decreased synthesis of methionine and S-adenosylmethionine (SAM). (P. A. Jones and Liang 2012; B. C. Christensen et al. 2012; Bacalini et al. 2014; W. Xu et al. 2016)

SAM provides the CH<sub>3</sub> to be added to the 5-position of the cytosine ring on histones and DNA in the trans-methylation event (Figure 4A). The production of SAM is localized to the cytosol in every cell, but the liver has an important role in SAM synthesis and degradation. A significant proportion of daily-consumed methionine is converted to SAM in the liver in a reaction, which consumes ATP and is catalyzed by methionine adenosyltransferase (Figure 5). The synthesized SAM is constantly restored to S-adenosylhomocysteine (SAH) through the trans-methylation where a biomolecule (such as DNA) receives the CH<sub>3</sub>. SAH may be a competitive inhibitor of trans-methylation reaction, and thus the increased quantity of SAH or decreased quantity of SAM (seen in SAM-SAH ratio) may disturb trans-methylation reactions. Consequently, it is important to remove excess SAH through a reaction that converts SAH to homocysteine and adenosine. In addition, SAM acts in other crucial biochemical pathways, in trans-sulfuration and in aminopropylation, and thus the one-carbon cycle is highly essential for numerous biochemical reactions in the cell. (P. A. Jones and Liang 2012; B. C. Christensen et al. 2012; Bacalini et al. 2014; W. Xu et al. 2016) For instance, homocysteine from the one-carbon cycle is a biomarker of the cardiovascular system, and functions in lipid metabolism (E. Crimmins et al. 2008).

Folate is an especially important substrate in the one-carbon cycle as shown in Figure 5. Humans are unable to synthesize it and it has to be supplied in a nutrient or in the form of synthetic folic acid. In a well-known example, adequate maternal folate supply decreases risk for neural tube abnormalities, and thus it is essential for fetal development. (W. Xu et al. 2016) Human studies imply that B vitamin absorption is influenced by genetics and a diet poor in folate is associated with global DNA hypomethylation (Axume et al. 2007; Bacalini et al. 2014). B12 vitamin deficiency is linked to DNA damage, DNA methylation changes, reduced quantity of SAM, and increased quantity of homocysteine (Bacalini et al. 2014). Ethanol is an example of a substance with a negative influence on the one-carbon cycle. It disturbs the absorption of folate in the intestine, and the release of folate from the liver to receiving tissues. Ethanol interferes with the one-carbon metabolism in multiple ways: it blocks the reaction where homocysteine is converted to methionine, the reaction where methionine is converted to SAM, and the methylation pathways. (B. C. Christensen et al. 2012; Varela-Rey et al. 2013)



**Figure 5.** One-carbon metabolism (cycle) provides universal methyl group donor, S-adenosylmethionine (SAM) for cellular epigenetic modifications (methylation of DNA and histones) (modified from Xu et al. 2016). In the cycle, vitamin B (2, 6, and 12), choline, and methionine (in capital letters) may originate from dietary sources. Folate (in capital letters) originates only from diet. The process of SAM production involves sequence of reactions. First, folate is transformed to dihydrofolate, and further to tetrahydrofolate (THF), which enters the one-carbon cycle. This is followed by reaction where THF is converted to 5,10-methylene tetrahydrofolate with vitamin B6 as a cofactor. Next process where 5,10-methylene tetrahydrofolate is transformed to 5-methyl tetrahydrofolate depends on vitamin B2. In de novo methionine biosynthesis, vitamin B12-dependent methionine synthase catalyses conversion of 5-methyltetrahydrofolate and homocysteine to tetrahydrofolate and methionine. Finally, methyl group is transferred from methionine to SAM. Transfer of CH<sub>3</sub> from SAM to biomolecules (i.e. DNA, histones or others) is a trans-methylation event. S-adenosylhomocysteine (SAH) is the product released after methyl group transfer. SAH may be used in homocysteine production. Then, the one-carbon circle has reached full circle because homocysteine may be used again in the methionine synthesis. Alternatively, dietary choline, which is converted to betaine may be used as a source of methionine (and SAM) production. (Xu et al. 2016)

The SAM/SAH ratio is influenced by multiple environmental factors in addition to nutritional substances. Arsenic from soil, water, and food may enter the cell and participate methylation reaction in which the arsenite methyltransferase enzyme catalyzes production of methylated arsenic compounds. This reaction consumes SAM, and it is therefore a competitive reaction for DNA methylation and this process may account for DNA hypomethylation. Tobacco smoking also alters DNA methylation profiles, and arsenic can be found in cigarettes. (Zampieri et al. 2015) In addition to arsenic, other compounds such as bismuth, selenium, a polyphenol from green tea (epigallocatechin gallate, EGCG), caffeic acid and chlorogenic acid may participate in enzymatic methylation and consume the methyl group donor, SAM. Moreover, expression or activity of DNMTs may be altered. Competitive inhibition of the enzyme activity for cytosine has been observed in the presence of EGCG, soy-bean isoflavon genistein and curcumin. Moreover, the gene-expression levels of DNMTs are shown to be influenced by the same nutritional and environmental factors (e.g. arsenic, alcohol, curcumin). (M. Z. Fang et al. 2003; M. Z. Fang et al. 2005; M. Fang et al. 2007; Z. Liu et al. 2009; Xie et al. 2014; Zampieri et al. 2015)

Smoking-associated, CpG site-specific DNA methylation changes have been observed in human cohorts. For instance, smoking induces DNA hypomethylation at a specific CpG site (cg03636183) located in gene *F2RL3*. This change appear to be reversible because DNA methylation levels in cg03636183 were recovered close to the level of never-smokers (>20 years) after smoking had been ceased. (Breitling et al. 2011; Zhang et al. 2014) Of the large number of EWASs focusing on smoking-associated effects in the blood, in addition to *F2RL3*, DNA methylation changes in CpG sites located in genes *AHRR* (cg05575921), *GPR15* (cg19859270) as well as intergenic regions 2q37.1 and 6p21.33 have been reported frequently (Gao et al. 2015). Furthermore, a very recent case-control study with strong evidence through validation and replication cohorts by Baglietto et al. (2017) showed that smoking-associated changes in DNA methylation are also associated with lung cancer risk (Baglietto et al. 2017). In another study, van Dongen et al. (2016) coupled trajectories of aging-associated methylation level changes with own smoking in a sample comprising monozygotic twins (van Dongen et al. 2016). Nonetheless, despite this, the full picture of smoking-related methylomic changes is much more complicated because another very recent study demonstrated that smoking has distinct effects on DNA methylomes and transcriptomes of purified leukocyte subtypes. For example, methylation levels in genes *AHRR* (cg05575921), *ALPPL2* (cg21566642), *GFI1* (cg09935388), *IER3* (cg06126421) and *F2RL3* (cg03636183) all showed different smoking-associated profiles in granulocytes, monocytes and B cells. (Su et al. 2016)

Some human DNA methylomic association studies have highlighted that the epigenetic profile adapts environmental signals during the developmental stages. Prenatal exposure to large changes in environmental circumstances has been linked to the future health of the offspring. For instance, the effect of Dutch famine in 1944-45 during the first trimester of pregnancy has been associated with offspring obesity in adulthood. Moreover, when compared to their same-sex sibling, the

offspring showed hypermethylation in insulin-like growth factor 2 receptor gene from the IIS network. (Heijmans et al. 2008) *In utero* exposure to smoking (Chatterton et al. 2017) or arsenic (Kaushal et al. 2017) has been linked to changes in fetal DNA methylomes in human studies, and these two substances are linked to altered brain development and health conditions in the after-birth life, respectively. In another study with mothers exposed to a heavy winter storm in Quebec in 1998, prenatal maternal stress was associated with adiposity as well as changed methylomic profiles of the children born (Cao-Lei et al. 2015). The perceived maternal stress during pregnancy may also contribute to immune function development (Veru et al. 2014). A recent report from the study of the winter storm in Quebec has provided signs that maternal stress which is mediated through changed DNA methylation, manifests in cytokine production in accordance with T helper cell type 2 promotion in the children (Cao-Lei et al. 2016).

#### 2.4.4 DNA methylome changes in diseases

It has been expected that epigenetics would easily provide the missing mechanistic link between environment and human pathologies, but this aim has been only partially reached, at least regarding DNA methylation. However, shown connections between epigenetics and diseases do exist (Figure 2). In many cases, such as in some cancers, the role of DNA methylation in the development of pathologies through a straightforward on-off switch system has been demonstrated. Large changes in DNA methylation levels in oncogenes belong to this paradigm. Disease associations exhibiting lower methylation level changes (ranging between 1% and 10%) in DNA methylation are still a grey area. Such diseases or health conditions are often complex disorders (hypertension, type 2 diabetes, coronary heart diseases, and cardiovascular disorders). (P. A. Jones 2012; Leenen et al. 2016)

##### 2.4.4.1 Cancer

Methylation changes in aging and progression of cancer share some characteristics; in both cases, promoter-specific CpG dinucleotides are often hypermethylated (Ahuja and Issa 2000; Zampieri et al. 2015). Cancer development follows separate crossing phases: the cells become precancerous, start to grow without control, and ultimately form a malignant tumor. Strong evidence suggests that alterations in methylation of CGIs located in tumor-suppressing genes (e.g. *p16INK4a*, *PTEN*, *GSTP1*, *TP53*, *BRCA1*, *p14*, *p16*, among many others) contribute to the development of various cancers. Genes *VIM*, *SEPT9*, *SHOX2*, *GST1*, *APC*, and *RASSF1A* are considered to be the principal diagnostic epigenetic cancer biomarkers. Their methylation levels may differ more than 60% in the on/off-switch manner between healthy and malignant situations. (Juo et al. 2014; Leenen et al. 2016; Elhamamsy



2016) CpG island methylator phenotype (CIMP) is a tool that is used in prognosis and tumor classifications. It is a CGI-methylation profile, which utilizes methylation statuses from groups of genes, and the CIMP statuses are classified as positive and negative or high, low and zero. The specific pool of genes in CIMP classification varies. According to a review by Juo et al. 2014 genes *MINT1*, *MINT2*, *MINT31*, *CKKN2A(p16)*, *bMLH1* or *CACNA1G*, *IGF2*, *NEUROG1*, *RUNX3*, *SOCS1* have been most often employed by colorectal cancer prognosis. (Juo et al. 2014) The prognostic power of cancer-specific CIMP has been demonstrated, for instance, in meta-analyses with colorectal cancer (Juo et al. 2014) and neuroblastoma (Asada et al. 2013) patients.

#### 2.4.4.2 Imprinting disorders

Many severe pathologies such as Transient neonatal diabetes 1 or Silver–Russell, Beckwith–Wiedemann and Prader–Willi Syndromes have been associated with imprinting faults in single or multiple loci in the human genome. Causes of these disorders include deletions, mutations, uniparental disomy, and epimutations. (Elhamamsy 2017)

Nevertheless, altered imprinting in some genes, such as *H19* and *IGF2*, has been observed in healthy newborns, and thus this modification appears to also exist in non-pathologic phenotypes. In other words, altered imprinting may result a healthy phenotype with minor effects in health condition. (Elhamamsy 2017)

#### 2.4.4.3 Brain disorders and DNA methylation

During the human lifespan, multiple sensitivity periods occur, when the brain plasticity is high, and is the most capable of developing various functions, such as language and behavior. Long-term memory requires complex synapse-to-nuclear signal transduction cascades to be initiated in such a steady manner that it leads to controlled and firm changes in the transcriptome. Epigenetic regulation has been shown to contribute to these processes in the brain. The gene BDNF contributes to neurodevelopment, neuroplasticity, the onset of psychiatric disorders, and suicidal behavior. The extensive research with BDNF using animal models for early-life adversities has shown that long-term memory formation is associated with locus-specific bidirectional DNA methylation changes. (Leenen et al. 2016) In stressed rodents, BDNF promoter sites were hypermethylated (methylation change 10-15 percentages) when compared to controls. This change was accompanied by downregulated gene expression. (Roth et al. 2009; Leenen et al. 2016) Regarding previous examples, analogous changes in BDNF methylation profile have been

shown in human post-mortem brain samples from adults who have committed suicide. (Keller et al. 2010; Leenen et al. 2016)

As DNA methylation is in a central role of nervous system organogenesis, aberrant methylation is suggested to be a significant contributor to neurodegenerative disorders. However, the evidence from many such disorders is limited. (H. Lu et al. 2013; Heyward and Sweatt 2015; Leenen et al. 2016; Elhamamsy 2016) In Alzheimer's disease, global DNA hypomethylation and CpG site-specific methylation changes on certain genes (MAPT, APP, GSK3B, FKBP5, CRTCL1) have been reported (Blair et al. 2013; H. Lu et al. 2013; Mendioroz et al. 2016; Elhamamsy 2016). In neurons, changes in levels of DNMTs and 5mC quantities in the mitochondrial DNA have been associated with amyotrophic lateral sclerosis (Iacobazzi et al. 2013). Changes in DNA methylation and DNMT levels have also been linked to other diseases (such as Parkinson's and Huntington's), in which neurons are also progressively lost through endogenous apoptotic pathways (Elhamamsy 2016).

#### 2.4.4.4 Environment, DNA methylation and disorders

Environmental factors including parental diet, maternal mental state and experienced early-life stress/adversity and early-life diet have been connected to DNA methylation changes accompanied by various complex disorders or pathologies in adulthood (obesity, hypertension, type 2 diabetes, coronary heart diseases, and cardiovascular disorders) (Leenen et al. 2016). The Dutch famine case is particularly indicative of such connections (Heijmans et al. 2008; Roseboom et al. 2011). As reviewed by Leenen et al. (2016), in general, in EWASs with focus on complex disorders, the observed DNA methylation changes between cases and controls has typically been low. The magnitude of methylation level change ( $\Delta$ -change) has ranged from one to ten percentages. In the Dutch famine study, those individuals who had exposed to malnutrition *in utero*, showed a 5.0 and 1.6 percentage decrease in methylation levels in genes *IGF2* and *INSIGF*, respectively, in later adulthood. A 1.1-2.4 percentage increase in DNA methylation levels was observed in *IL10*, *ABCA1*, *GNASAS* and *LEP*. The individuals were more obese, and more often had impaired glucose tolerance, hypertension and a higher risk rate for coronary heart disease. (Heijmans et al. 2008; Tobi et al. 2009) Moreover, maternal severe depression during gestation has been associated with a higher tendency of these same negative health effects of the offspring at their adult age. The offspring of depressed mothers were also born with lower birth weights. The genes where aberrant methylation was observed in parallel with these defects were *SLC6A4*, *MEG3*, *IGF2*, *PLAG1* and *PEG10* (magnitude of methylation change 1-10 percentages). (Y. Liu et al. 2012)

The findings as in the Dutch famine study support the paradigm that organism adjusts its phenotype to the environment through epigenetic processes *in utero* and

during childhood. The biological plasticity in the early stages of lifespan has been considered as having the potential to guide lifelong health trajectories. Nevertheless, Leenen et al. (2016) concluded in the review that the studies have often only reported associations, not functionally shown mechanisms, and thus these findings should be re-evaluated with experiments showing the cascades of events connecting and environmental factor(s), the epigenome and the phenotype. It is still uncertain whether changes in the methylome are contributing to the mechanism, which increases the prevalence of disorders, or if these changes are simply an independent side effect. (Leenen et al. 2016)

#### 2.4.4.5 Genetic variation, DNA methylation and disorders

The complex relationship between genetic variation and DNA methylation is another aspect regarding epigenetics in aging, health condition and disease progressions. Estimations in twin data by van Dongen et al. (2016) showed that genome-wide heritability of DNA methylation levels is on average 0.19, and common genetic variation captured by single nucleotide polymorphisms (SNPs) explain ~7% of the methylomic variation. The study also highlighted that there are regions (i.e. CpG sites) in the genome where aging-associated interindividual DNA methylation changes are explained more strongly by genetic differences. On the contrary, some methylation sites appear to be more prone to the environmental effects during aging. (van Dongen et al. 2016)

Moreover, quantitative trait loci (QTL) mapping studies of *trans*-gene expression effects have shown the potential to unravel the downstream biological effects impacted by genetic variation (Westra et al. 2013). The effect of nearby or long-distance genetic variation on methylation levels may be explored through either *cis*-meQTLs or *trans*-meQTLs. It means that the genomic distance between a CpG site and a SNP in the exploration is set to certain limits such as distance < 250 kilo bases and distance > 5 mega bases, respectively. A very recent study using whole-genome DNA methylation (Illumina 450Beadchip) and RNA-seq data of 2101 individuals focused on known disease-associated genetic variants, i.e. 6111 manually selected SNPs associated with, among others, immune, cancer, cardiovascular, neurological and metabolic system-related diseases. The study demonstrated the overall complexity in gene expression regulation and the countless possibilities of functional connections in the regulatory networks. They reported a correlation between methylation and expression levels of 12809 CpG sites and 3842 transcripts, respectively, in *cis*. 70 percentages of these methylation sites exhibited inverse correlation with gene expression in accordance with textbook expectations. The remainder showed a positive correlation with the expression levels. (Bonder et al. 2017) The phenomenon that DNA methylation has both an enhancing and repressing effect on gene expression has been acknowledged elsewhere (Maunakea et al. 2010; P. A. Jones 2012). Moreover, Bonder et al. (2017) observed 1907 of the

tested 6111 disease-associated SNPs having an effect on methylation levels in 10141 methylation sites *in trans*. In addition, they included in the analysis ChIP-seq data-based information of transcription binding to the CpG sites. While *cis*-meQTLs showed no enrichment, the *trans*-meQTLs exhibited functional enrichment around transcription binding sites, while they were infrequent in heterochromatin. The study highlighted that disease-associated genetic variants *in trans* may change transcription factor gene expression (e.g. *NFKB1*, *CTCF*, *NKX2-3*) and DNA methylation levels at their target sites. The downstream effects of the disease-associated variants showed uniform features: DNA hypomethylation is often coupled with enhanced transcriptional activator binding. (Bonder et al. 2017)

#### 2.4.5 Aging-associated changes in the DNA methylome

During aging, DNA methylation levels change in a high number of CpG sites (K. Christensen et al. 2009; Boks et al. 2009; Bollati et al. 2009; B. C. Christensen et al. 2009; Rakyan et al. 2010; Bocklandt et al. 2011; Bell et al. 2012; Garagnani et al. 2012; Heyn et al. 2012; Horvath et al. 2012; Hannum et al. 2013; Gentilini et al. 2013; Horvath 2013; Weidner et al. 2014; Florath et al. 2014; McClay et al. 2014; Marttila, Kananen, Hayrynen et al. 2015; Gervin et al. 2016; Slieker et al. 2016). In general, interindividual differences in genome-wide DNA methylation profiles increase with age (Zampieri et al. 2015), and genetic and environmental factors contribute to this change (van Dongen et al. 2016). Specifically, unpredictable variability in genome-wide DNA methylation patterns increases with age, but there are also individual CpG sites in which DNA methylation level decreases or increases with almost constant rate as a function of age. These two phenomena contribute to epigenetic drift and epigenetic clock, respectively. The epigenetic drift is demonstrated in changed cell-type specific DNA methylation profiles that are transmitted forward to daughter cells. (Zampieri et al. 2015).

With age, DNA methylation located in Alu and LINE-1 elements becomes more variable, and, in addition, those elements become globally hypomethylated. Failed methylation maintenance due to downregulated or inhibited DNMTs and passive demethylation due to, for example, insufficient SAM pools (Figure 4A and 5) may contribute to the epigenetic drift. (Zampieri et al. 2015; Jones et al. 2015; Bonder et al. 2017) Interestingly, very recent results from a large cross-sectional Europe-wide sample (named as MARK-AGE) have coupled changes in the active DNA demethylation pathway (Figure 4C) to aging. In the study, they showed that *TET1*, *TET3* and *TDG* gene expression was downregulated by age in parallel with 5hmC and 5caC decrease and increase, respectively. (Valentini et al. 2016) These findings underline the loss of control in DNA methylation profile maintenance in aged humans.

Regarding global methylation, individual sites and their genomic locations, the DNA methylation patterns change in different ways during the human lifespan. The

global level of DNA methylation is lower in neonatal blood when compared to the early years of life and the global methylation level increases most rapidly during the first year. These changes appear to concentrate on specific genomic locations (CGI shores and CGI shelves) and regulatory elements such as enhancers and non-CGI promoters. (M. J. Jones et al. 2015) In addition, longitudinal studies in buccal as well as in blood cells of monozygotic twins have shown that the initially similar DNA methylomes become highly divergent (i.e. inter-individual variability increases) during the first years (D. J. Martino et al. 2011; D. Martino et al. 2013). The *in utero* environment may be the contributor to the similarities in the epigenomes of the newborn twins while the post-natal life with fast development together with the effects by variable environmental factors drive the epigenetic divergence as time passes. In adults, methylomic deviation continues at single CpG site resolution but with a lower rate when compared to children. (M. J. Jones et al. 2015) Divergence of DNA methylomes during aging in adults is demonstrated by several monozygotic twin studies (Fraga et al. 2005; Talens et al. 2012; van Dongen et al. 2016). However, in spite of the continuous divergence, global methylation level is, on average, relatively stable during the middle stage of a lifespan (i.e. during adulthood). At older ages, variation in methylation patterns is at its highest, and at that time, global hypomethylation is an observable phenomenon. (M. J. Jones et al. 2015) Based on array technology experiments in aged human populations, aging-associated changes typically comprise site-specific hypermethylation at CGIs located at proximity of gene promoters and hypomethylation outside of CGIs (Zampieri et al. 2015).

Until the most recent years, EWASs have been performed without cell subtype considerations even though the samples have been heterogenic, for example, whole-blood samples. A recent study by Gervin et al. (2016) aimed to explore longitudinally intra-individual aging-associated DNA methylation changes that are independent from cell subtype variation. The study comprised 57 whole-blood samples of 2-, 10- and 16-year-old subjects and the methylomes were analyzed using reduced representation bisulfite sequencing. The final filtered dataset comprised 635 899 CpG sites. They concluded that the aging-associated changes seen in methylomic landscape are largely mediated through variation in cell subtypes. In these young individuals, less than 10 percentages of the 346 aging-associated CpG sites located in 178 genes demonstrated an association with age directly, without the cell type effect as the mediator. (Gervin et al. 2016) In adults, similarly, the cell subtype effect and methylomic changes are shown to manifest in parallel with aging (Marttila et al. 2015).

The impact of genetic and environmental effectors on the aging-associated shifts in the methylomic landscape has been demonstrated in a study by van Dongen et al. (2016) using a population-based twin cohort (age range 17-79 years) and EWAS approach with microarray technology. As a detailed example from the study, the correlations of DNA methylation levels in 65 aging-associated CpG sites between smoking-discordant monozygotic twin pairs were clearly lower than the correlations between smoking-concordant twin pairs. In other words, environmental effectors

such as smoking appeared to accelerate aging-associated changes in these sites. van Dongen et al. (2016) reported that the heritability ( $h^2$ ) was, on average, 0.19 for the site-specific DNA methylation levels genome-wide, but the overall variation in  $h^2$  values was considerable. Furthermore, detailed region-specific exploration in the study demonstrated that highly inheritable sites are enriched in CGIs and DNase I hypersensitivity regions. In contrast, CpG sites where the variation in DNA methylation by age was most likely to be impacted by random, environmental contributors were lacking from those regions, and instead, they were enriched in CGI shores. These findings suggest that genetics has a more relevant role in explaining the aging-associated DNA methylation variability in promoter-associated CGI and DNase I hypersensitivity regions. In contrast, environmental factors seemed to have a stronger impact on DNA methylation located in gene bodies, shores, shelves and non-CGI sites. (van Dongen et al. 2016) These results underline that, in addition to CGIs, up to two kilo bases distant CGI shores exhibit dynamically and functionally relevant DNA methylation changes. These characteristics of CGI shores during development and in different tissues have been shown in other studies, as well (Irizarry et al. 2009; Doi et al. 2009; Bockmühl et al. 2015).

In order to accurately interpret reported results, it should be mentioned that the commonly used array technology (i.e. Illumina EPICBeadchip, 450Beadchip or 27Beadchip) covers only small fraction of the ~27 million CpG sites in the human genome, and the technology was originally designed to focus more on promoter-associated regions. (Zampieri et al. 2015) Analyses using full genome scanning methods (i.e. whole genome bisulphite-sequencing) with complete genomic coverage of the 27 million methylation sites have indicated that, during aging, hypomethylation appears to dominate over hypermethylation. For example, the latter events have been reported in only 13 percentages of the aging-associated methylation regions in a study by Heyn et al. (2012) in which methylomes in CD4<sup>+</sup> T cells of a newborn and a centenarian were compared. The authors found in total, 617 338 CpG sites as differentially methylated by age across the human genome. These sites were located in promoters (10%), exons (10%), introns (45%) and intergenic (34%) regions. Among the intergenic regions, methylation differences in repetitive sequences including LINE-1 and LINE-2, LTR retrotransposons and Alu elements were abundant. (Heyn et al. 2012)

Another whole-epigenome-scanning study (McClay et al. 2014) covering the 27 million autosomal CpG sites in blood samples of 700 individuals aged 25-92 years demonstrated some overlap but also some discrepancy with previously reported results before the year 2014. Analogously to Heyn et al. (2012) using technology with higher resolution, the authors concluded that hypomethylation was a more abundant aging-associated event than hypermethylation. In accordance to array technology-based results, hypermethylated regions were enriched to CpG islands and shores. The study by McClay et al. (2014) showed that hypomethylated regions were enriched in genomic locations associated with polycomb/regulatory proteins (such as EZH2, CTCF, histone 2A member Z) or histone modifications (such as

H3K27ac, H3K4me1- H3K4me3, H3K9ac) (McClay et al. 2014). Elsewhere, quantities of H3K4me3 have been linked to the longevity of model animals (Han and Brunet 2012). Some of the top-ranking aging-associated methylation regions in the study by McClay et al. (2014) were annotated to previously aging-linked protocadherins, homeobox genes and MAPKs. In addition, the study was first to report aging-associated changes in regions annotated to ryanodine receptors (McClay et al. 2014). However, even though these two example studies provide high genomic coverage of CpG sites, the sample size was minimal (Heyn et al. 2012) or the cell type heterogeneity was considered only moderately (McClay et al. 2014). For instance, McClay et al. (2014) controlled all the unwanted variation, i.e. unmeasured confounding factors, in a robust manner, using a method where the variation is captured to the principal components produced by PCA and these components are regressed out.

A study by Sliker et al. (2016) reported aging-associated variably methylated positions that are related to central aging mechanisms. They used Illumina 450Beadchip analysis to characterize methylomic profiles of 3295 blood samples from participants at the ages of 18-88 years. Of these samples, RNA-seq gene expression profiles were also characterized from 2044 samples. In their hypothesis, methylation changes which correlate most closely with chronological age, are not the best descriptive of human biological age. Consequently, in addition to the EWAS, where differentially methylated positions by age are explored, the study focused on CpG sites where the divergence in DNA methylation levels increases gradually as a function of chronological age rather than correlates with it. The age-related variably methylated sites were considered to be a reflection of the health status in the elderly. The variability was determined as significant when it increased more than 5% in 10 years, and it was calculated independent of major blood cell subtype distribution, and was validated in separate cohorts. (Sliker et al. 2016) As a result, the most variable sites appeared to represent a unique cluster of aging-associated regions. Of the ~400 000 CpG sites in the analysis, they reported 25% as differentially and 1.6% (i.e. 6366) as variably methylated sites during aging. The lists of variably methylated sites and the previously reported sites which correlate most closely with age (Rakyan et al. 2010; Heyn et al. 2012; Bell et al. 2012; Hannum et al. 2013; Horvath 2013; Florath et al. 2014; Z. Xu and Taylor 2014) were compared, and the lists overlapped with each other only slightly. The variably methylated sites by age were categorized based on their direction of methylation change: in relation to (1) increase or (2) decrease in average methylation change or (3) independent of average methylation level change by age. Variably (1) hypermethylated and (2) hypomethylated sites by age showed enrichment bias to CGIs and non-CGI-regions in a similar manner to that reported previously. Over 60% of the variably methylated CpG sites were enriched in regions associated with Polycomb Repressive Complex 2 (PRC2) binding sites and chromatin state markers (based on Epigenomics Roadmap) that are signals of repressed genome. (Sliker et al. 2016)

To date, in spite of the great interest, only a small fraction of aging-associated methylation sites which correlate with chronological age have been shown to be associated with changes in gene expression (Hannum et al. 2013; Steegenga et al. 2014; Reynolds et al. 2014; Marttila et al. 2015; Dozmorov 2015). Nevertheless, Dozmorov (2015) has reported that when comparing the functional roles of the aging-associated genes and methylation regions, they appear to share some associated-features such as repressing (H3K27me3) and activating (H3K4me1) histone marker combinations, as well as binding of PRC2 co-factors. PRC2 signature seem to be consistent in different cell types. Both aging-associated transcripts and methylation sites show enrichment in chromatin regions termed as bivalent promoters. (Dozmorov 2015)

Correlations between gene expression and methylation levels have been analyzed often locally. That is to say, hypothetical associations are tested in the overlapping genomic region of a CpG site and the canonical transcript. However, Slieker et al. (2016) correlated DNA methylation levels in age-related variably methylated positions against expression levels of all expressed and coding genes determined by RNA-seq analysis. The results showed that the variably methylated sites associate with gene expression. Of the 6366 variably methylated positions, a significant correlation between levels of gene expression *in cis* (i.e. target CpG site and the gene within 500 base pairs) and methylation was observed in 1549 genes. The *cis*-genes showed enrichment in GO term categories of *neuron differentiation* and *neuron development*. Of the 6366 variable sites, 854 were associated with expression changes in *trans*-genes. These genes were located in a different chromosome or were more than five mega bases apart from the methylation site. A significant proportion of variably methylated sites correlated positively with the expression of *trans*-genes encoding components (such as EED, SUZ12, EZH2) for PRC2. The pathway analysis revealed that *trans*-genes which increased gene expression levels and were accompanied by changed methylation levels in variably methylated regions belong to fundamental aging-associated mechanisms, including *DNA repair and apoptosis*. In contrast, downregulated *trans*-genes mapped to intracellular metabolic pathways (such as *pentose metabolism and regulation of CDC42 activity*), and thus exhibited different phenomenon. (Slieker et al. 2016)

#### 2.4.5.1 Epigenetic clock

Numerous DNA methylation sites appear to change their methylation levels as a function of gained years with the same direction of change and with highly constant rates. These DNA methylation sites demonstrate the paradigm of the epigenetic clock. Furthermore, certain CpG sites can be used in age predictor models, which accurately predict the chronological age of an individual. (Zampieri et al. 2015; M. J. Jones et al. 2015; Jylhava et al. 2017) In addition to estimation of chronological age, the epigenetic clock is considered to be descriptive of the biological age of an



individual. In this hypothesis, the difference between estimated epigenetic age and chronological age is the measure representing level of biological aging. Epigenetic age and other biological age predictors are hypothesized to provide more detailed information for disease risk estimations when compared to chronological age alone. (Zampieri et al. 2015; M. J. Jones et al. 2015; Jylhava et al. 2017)

Multiple epigenetic clock estimators have been proposed. The most well-known blood methylomic predictors for humans are by Horvath (2013) using 353 CpG sites, Hannum et al. (2013) using 71 CpG sites and Weidner et al. (2014) using 99 CpG sites. The predictors mentioned here for humans are all based on DNA methylation measurement by Illumina BeadChip platform. A total of 99 CpG sites in the Weidner predictor were selected using bivariate correlation analysis. The CpG sites used in epigenetic clock algorithms by Horvath and Hannum have been fetched using elastic net algorithm (i.e. penalized regression model) in such a way that the combination of individual CpG sites predict the best, not the single CpG sites. These two predictors share only six overlapping CpG sites, intelligibly. The predictor CpG sites in Horvath's and Weidner's clocks have 22 overlapping CpG sites. (Hannum et al. 2013; Horvath 2013; Weidner et al. 2014; Lin et al. 2016; Jylhava et al. 2017)

Horvath (2013) generated the multi-tissue age predictor algorithm (available at <https://dnamage.genetics.ucla.edu/>) using 21000 probes existing in both the HumanMethylation450 and HumanMethylation27 BeadChips. The predictor was trained and tested with 8000 samples comprising various tissue types and chronological ages. The different tissues and organs included whole blood, brain, breast, kidney, liver, lung, saliva as well as sorted cell samples including CD4<sup>+</sup> T cells, monocytes, B cells, glial cells and neurons. (Horvath 2013) In contrast, other epigenetic age predictors are often trained in blood samples (Hannum et al. 2013; Weidner et al. 2014). To date, the most widely used predictors appear to be those of Horvath and Hannum. Recent independent studies have shown that, even though estimated epigenetic age by these two age predictors correlate strongly with chronological age ( $r > 0.9$ ), they correlate less with each other ( $r = 0.37 - 0.76$ ). (B. H. Chen et al. 2016; Jylhava et al. 2017)

Multi-tissue DNA methylome age predictors for humans (Horvath 2013) and mice (Stubbs et al. 2017) demonstrate similar characteristics. Median absolute error for both of these predictors is 5 percentages of the average length of human and murine lifespan (~85 years and ~2 years, respectively). In other words, the murine epigenetic clock ticks faster than the human clock, and the epigenetic clock-ticking rate appears to be linked to the lifespan expectancy. This underlines that the biological aging process includes evolutionarily conserved features. (Wagner 2017)

The epigenetic age acceleration may be defined as the residual fetched out from a linear regression model where the epigenetic age is regressed on chronological age. Another more robust and simple option is to calculate the absolute difference between chronological and epigenetic age ( $\Delta\text{-cAge-DNAmeAge}$ , in such way that positive values indicate for "younger epigenetic age" and negative "older epigenetic age") in years. (Chen et al. 2016) The estimated heritability of the Horvath's

epigenetic age acceleration is considerable ( $h^2$  approximately 0.4) (Horvath 2013; Levine, Lu et al. 2015; Marioni, Shah, McRae, Chen et al. 2015), and the acceleration rate varies between genders and different ethnic groups (Horvath, Gurven et al. 2016). The epigenetic clock-ticking rate has been shown to be highest during organismal growth (Horvath 2013; Simpkin et al. 2016). Analyses in monozygotic twins have implied that the heritability of the epigenetic age is more than 70% at the time of birth but decreases to the level of ~40% by late adulthood. From this perspective, genetic background and environment together define the epigenetic age of an individual. (Horvath 2013)

The biological significance of Horvath's DNAmAge (Horvath 2013) in the aging-research and its performance in prediction of phenotypic outcomes has been supported by many studies focusing on different tissues (e.g. brain, blood, liver). These studies indicate that older epigenetic age reflects older biological age, as the epigenetic age acceleration is more abundant in various detrimental health conditions or is even associated with mortality. Studies have highlighted that DNAmAge predicts all-cause and cause-specific (cancer, cardiovascular) mortality (Marioni et al. 2015; Perna et al. 2016; Christiansen et al. 2016; Cole et al. 2017). In a study with monozygotic Danish twins, a 35% higher mortality rate was associated with older epigenetic age, and the older twin had a more than two times higher mortality risk when compared to the epigenetically younger individual in a twin pair (Christiansen et al., 2016). A meta-analysis with 13 089 individuals highlighted that the mortality-risk prediction with DNAmAge estimate is independent of other risk factors (sociodemographic, lifestyle, comorbidity factors) (B. H. Chen et al. 2016). Individual CpG sites in the epigenetic clocks are also associated with lifespan expectancy. Specifically, CpG sites correlating with all-cause mortality from Weidner and Horvath clocks overlap in following regions: cg05294455 (*MYLA*), cg08598221 (*SNTB1*), cg09462576 (*MRPL55*), cg15804973 (*MAP3K5*), cg20654468 (*LPXN*), cg25268718 (*PSME1*), cg26581729 (*NPDC1*), and cg02867102 (no gene). Most of these sites are hypomethylated during aging. (Zhang, Hapala et al. 2017)

Interestingly, the extreme longevity of humans has been connected to younger epigenetic ages ( $\Delta$ -cAge-DNAmAge 8.6 years) in 82 Italian individuals who had survived to the age of 105-109 years (semi-supercentenarians) (Horvath, Pirazzini et al. 2015). More recent research has underlined that individuals aged 90 or more years display systematically younger epigenetic ages (Christiansen et al. 2016; Armstrong et al. 2017). Whether the young epigenetic age of the elderly is related to the overall survival or is the age predictor algorithm(s) providing underestimates remains to be shown. However, some details indicate that the first option is more likely to be true. For example, 63 progenies (aged, on average, 71.8 years) of the 82 Italian semi-supercentenarians and 47 age-matched controls were compared, and the offspring showed 5.1 years lower epigenetic age than the control group. Thus, the young epigenetic age appears to be family-linked. (Horvath et al. 2015) This link may be explained by mixture of shared genetics and environment.

In addition, DNAmAge has been connected to descriptive of cognitive and physical fitness (Marioni, Shah, McRae, Ritchie et al. 2015), neuro-pathology (Levine et al. 2015) and frailty (Breitling et al. 2016) in the elderly. Epigenetic age acceleration is also associated with Down's syndrome (Horvath, Garagnani et al. 2015), Huntington's disease (Horvath, Langfelder et al. 2016), Parkinson's disease (Horvath and Ritz 2015), HIV infection (Horvath and Levine 2015), menopause (Levine et al. 2016), and cellular senescence (Horvath 2013; Lowe et al. 2016). Furthermore, studies have suggested that behavioral and lifestyle factors accelerate or decelerate epigenetic aging. Specifically, higher exercise and educational level as well as intake of poultry and vegetables resulted in younger epigenetic ages in the study subjects (Quach et al. 2017). Moreover, symptoms of metabolic syndrome (i.e. higher BMI and blood pressure, and higher levels of cholesterol, insulin, glucose and triglycerides) and increased levels of CRP were observed in association with older epigenetic ages (Quach et al. 2017). Elsewhere, higher BMI has been linked to accelerated epigenetic aging in liver (Horvath et al. 2014) and in blood cells (Nevalainen et al. 2017). The epigenetic age by Horvath's calculator is proposed to be a predictor for the age of lung cancer onset by Levine et al. (2015a). Current smokers who developed lung cancer in the study also showed older epigenetic ages (adjusted for confounding factors) when compared to former or never smokers. (Levine, Hosgood et al. 2015)

Intriguingly, some findings regarding Horvath's DNAmAge estimate are against initial hypotheses and some of them are more difficult to interpret. Heroin users (Kozlenkov et al. 2017) and military personnel suffering from post-traumatic stress symptoms (Boks et al. 2015) have demonstrated younger epigenetic ages when compared to the controls. Moreover, a very recent study with 486 middle-aged Danish twins reported that epigenetic age measured from blood is not linked with cognitive abilities (Starnawska et al. 2017). McEwen et al. (2017) have also shown conflicting findings. In the study, the immune system aging in relation to epigenetic aging in longevity region in Costa Rica (age range of 60-100+ years) was investigated. The study reported that long-living individuals in Nicoyan have higher levels of naïve CD8<sup>+</sup>CD45RA<sup>+</sup>CCR7<sup>+</sup> cells in relation to memory CD8<sup>+</sup>CD28<sup>-</sup>CD45RA<sup>-</sup> cells when compared to the rest of the population in Costa Rica. The long-living Nicoyans showed less variation in their genome-wide methylation, i.e. smaller epigenetic drift, and thus exhibited younger epigenomes. However, the epigenetic age (determined with Horvath, Hannum and Weidner methods) was not accelerated in the general population in Costa Rica when compared to the Nicoyans. (McEwen et al. 2017)

It has been assumed that the DNAmAge estimates from a wide variety of different tissues of the same individual are substantially similar. Observations from a small group of middle-aged individuals pointed to that direction (Horvath 2013), but further studies with larger sample sizes are lacking. The few observations have highlighted that the epigenetic aging rate of the cerebellum is slower (Horvath et al. 2015b) and the epigenetic age of sperm is younger (Horvath 2013) than the age in other tissues of the human body. Although the available evidence is limited,

observations in non-proliferative and immortalized tissues have indicated that epigenetic age differs from mitotic age and that epigenetic age is more likely to be a measure describing the state of epigenetic maintenance (Horvath 2013; Horvath and Levine 2015).

### 3 AIMS OF THE THESIS

The hypotheses in this thesis comprise the following issues. The longitudinal behavior of the epigenetic age appears to be indefinite. Discrepancies in the lists of reported aging-associated DNA methylation sites have emerged, thus requiring further evaluation. A further question concerns methylomic mortality predictors: are there methylation sites that could be used as survival predictors in the elderly (90+ years of age) and could those predictors even overcome conventional biomarkers of aging in the survival analysis? Moreover, are biomarkers (such as CMV or blood cell composition) related to immune system aging associated with epigenetic aging?

Therefore, aging-associated DNA methylation level changes (I), and mortality predictors (IV) at single CpG site resolution are characterized. The longitudinal behavior of the DNA methylome age (III) and its role as mortality predictor (IV) is explored. Furthermore, the epigenetic age change associated with CMV infection is analyzed (II). In addition, in all sections of the thesis, the role of blood cell sample heterogeneity is considered (I-IV). Specifically, the relevance of controlling sample heterogeneity in EWASs is evaluated (I), the association between the DNA methylome age and blood cell subtype counts is explored (III), and the overall longitudinal behavior of the blood cell subtypes is investigated (III).

The specific aims of this thesis were:

- 1) To characterize aging-associated changes in DNA methylation levels in blood samples of clinically healthy middle-aged subjects.
- 2) To characterize the longitudinal behavior of the DNA methylome age.
- 3) To characterize mortality-associated DNA methylation pattern at very advanced age.
- 4) To characterize the longitudinal behavior of the blood cell subtype landscapes.
- 5) To analyze whether the DNA methylome age of the blood cells is associated with CMV infection in populations of young adults and nonagenarians.

## 4 SUBJECTS AND METHODS

### 4.1 Subjects

#### 4.1.1 The Young Finns Study (I, III)

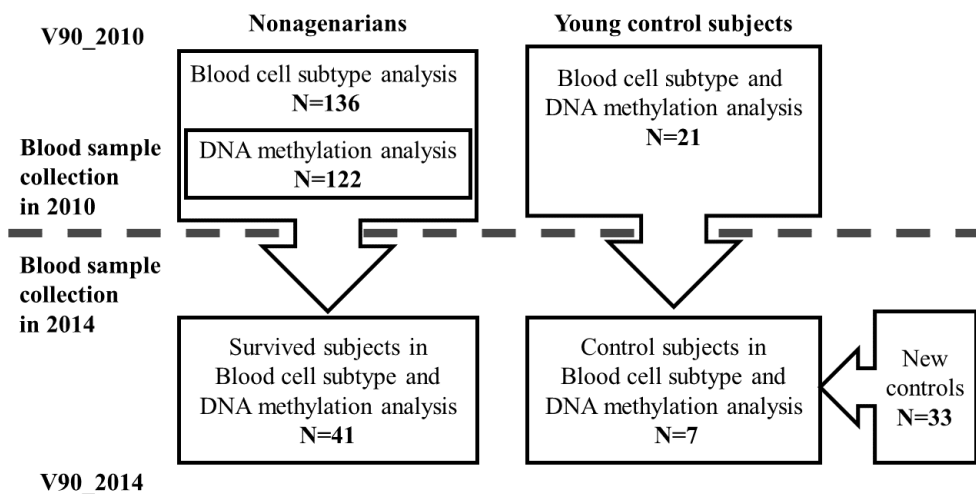
The study population in Studies I and III were selected from The Young Finns study (YFS) follow-up cohorts in 1986 and 2011 (denoted here as YFS\_1986 and YFS\_2011, respectively) (Nuotio et al. 2014). The YFS comprises a series of six cohorts, representing the general population, born in 1962, 1965, 1968, 1971, 1974 and 1977 from five cities with university hospitals in Finland (Helsinki, Kuopio, Oulu, Tampere and Turku) (Raitakari et al. 2008). Of these subjects, a subsample of 184 individuals was randomly assigned for a methylomic pilot study: the categories of chronological age in 1986 were 15, 18, 21 and 24 years, and the group sizes were 50, 44, 55 and 35 in which 58%, 68.2%, 56.4% and 60% were women, respectively. The cross-sectional sample from a follow-up in 2011 was used in Study I whereas samples in 1986 (baseline) and in 2011 (follow-up) were used in the Study III. All of the participants were of Western European descent and they provided informed consent. The study followed the guidelines of the Declaration of Helsinki and was approved by the Ethical Review Committee of Turku University Hospital.

#### 4.1.2 Vitality 90+ study (II, III, IV)

The Vitality 90+ study (denoted here as V90) is a prospective population-based study that includes subjects aged 90 years or over, who are both home-dwelling and institutionalized and live in the city of Tampere, Finland. The recruitment and characterization of the subjects was performed as reported for previous V90 cohorts (Goebeler et al. 2003). As shown in Figure 6, the population (denoted as V90\_2010 sample) consisted of 136 subjects born in 1920, and 21 control subjects aged 19-29 years. 122 nonagenarians and 21 young controls in V90\_2010 were in DNA methylation analysis (Figure 6).

Of the subjects in V90\_2010 sample, 48 (41 nonagenarians and 7 young controls) participated in the 4-year follow-up examination, where biological samples were also collected (Figure 6). In order to increase the sample size of young control group, 33 new control subjects aged 19-34 years were recruited together with the existing seven 4-year follow-up participants. Both V90\_2010 and V90\_2014 samples were used in Study III. Specifically, 48 subjects were used for DNA methylation analysis and cell subtype characterization at both time points, in 2010 and 2014 (Study III).

In Study IV, information about the all-cause mortality for V90\_2010 cohort after 2.55-year and 4-year follow-ups was obtained from the Population Register Center. Causes of death were unavailable. DNA methylation and the full covariate data were available for 111 nonagenarians. The mortality rate at the 2.55-year follow-up was 32.4% (36/111) and 47.7% (53/111) at the 4-year follow-up.



**Figure 6.** The Vitality 90+ populations in the analyses. Nonagenarians and controls in V90\_2010 cohort were used in Study II. In Study III, both V90\_2010 and V90\_2014 samples were used. A total of 111 nonagenarians of the V90\_2010 with DNA methylation data were used in Study IV.

All study subjects were of Western European descent and had not suffered from any infections or received any vaccinations in the 30 days prior to blood sample collection. The study participants provided their written informed consent. This study has been conducted according to the principles expressed in the Declaration of Helsinki, and the study protocol was approved by the ethics committee of the city of Tampere (1592/403/1996; 765/13.03.01/2008, PSHP 7/2014, ETL R14002).

## 4.2 Methods

### 4.2.1 DNA methylation analysis

#### 4.2.1.1 Blood sample collection and sample preparation (I-IV)

Whole blood DNA of the YFS\_1986 (Study III) and YFS\_2011 (Studies I and III) cohorts was obtained from blood samples stored in EDTA and -20°C using a Wizard® Genomic DNA Purification Kit (Promega Corporation, Madison, WI, USA) according to the manufacturer's protocol.

The blood samples of the subjects in the V90\_2010 (Studies II, III and IV) and V90\_2014 (Study III) were collected into EDTA-containing tubes during a home visit. Plasma was separated and stored at -70 °C. The samples were directly subjected to leucocyte separation on a Ficoll Paque density gradient (Ficoll-Paque™ Premium, cat. no. 17-5442-03, GE Healthcare Bio-Sciences AB, Uppsala, Sweden). The PBMC layer was collected and the cells that were directed to FACS analysis and DNA extraction, were suspended in 1 ml of a freezing solution (5/8 FBS, 2/8 RPMI-1640 medium, 1/8 DMSO; FBS: cat. no. F7524, Sigma-Aldrich, St. Louis, MO, USA; RPMI: cat. no. R0883, Sigma-Aldrich, St. Louis, MO, USA; DMSO: cat. no. 1.02931.0500, VWR, Espoo, Finland) and stored in liquid nitrogen. The PBMCs allocated for RNA extraction were suspended in 150 µl of RNeasy lysis solution (Qiagen, Crawley, UK) and stored in liquid nitrogen.

DNA was extracted from the PBMCs using the QIAamp DNA Mini Kit (Qiagen, CA, USA) according to the centrifugation protocol provided in the manufacturer's protocol. The DNA was eluted in 60 µl of AE elution buffer and stored at -20 °C. The concentration and the quality of the DNA samples were assessed using the Qubit dsDNA HS Assay (Invitrogen, Eugene, OR, USA).

#### 4.2.1.2 Illumina methylation assay (I-IV)

DNA methylation levels in whole blood (Study I and III) and PBMC (Study II, III, IV) samples were determined using Infinium HumanMethylation450 BeadChip technology (Illumina, San Diego, CA, USA) (Bibikova et al. 2006; Bibikova et al. 2009; Bibikova et al. 2011). The laboratory measurement of samples in V90\_2010, V90\_2014 and YFS\_2011 cohorts were processed at the Core Facility of the Institute of Molecular Medicine Finland (FIMM), University of Helsinki. The corresponding DNA methylation profiles of the samples in YFS\_1986 cohort were measured using the same methodology at Helmholtz Zentrum, München, Germany. All samples were analyzed in a randomized order.



According to the manufacturer's protocol, 1  $\mu\text{g}$  of DNA of each sample was subjected to bisulphite conversion using the EZ-96 DNA Methylation Kit (Zymo Research, Irvine, CA, USA). A 4- $\mu\text{l}$  aliquot of bisulphite-converted DNA was subjected to whole-genome amplification, and the sample was fragmented enzymatically and hybridized to an Infinium HumanMethylation450 BeadChip. The BeadChips were scanned using the iScan reader (Illumina) to obtain the U (unmethylated) and M (methylated) probe intensities.

Regions targeted by Illumina methylation 450Beadchip technology are the following. Of the 485 577 methylation probes, 150 254 (30.9%) are targeted to CGIs. In addition, 112 067 (23.1%) and 47 144 (9.7%) of the probes are harboring CGI shores (no more than two kilo bases from CGIs) and CGI shelves (from two to four kilo bases from CGIs). Consequently, there are 309 465 (63.7%) CGI-associated probes in the array. (Ma et al. 2013)

#### 4.2.1.3 Preprocessing of methylation data (I-IV)

The processing of DNA methylation data was performed using R software (R>=2.15.3). The quality of the methylation data was carefully explored using standard examinations (Studies I-IV). Several visualization styles, such as box plots and density plots as well as principal component analysis (PCA) with scatterplots were used to monitor raw signal intensities and processed methylation values. Gender prediction based on methylation levels at CpG sites located in X and Y chromosomes was performed as a quality control step.

The data of YFS\_2011 (Study I) and V90\_2010 (Study IV) cohorts were preprocessed using R software as follows: Of the 485 000 CpG sites, all unspecific or polymorphic sites ( $N = 76775$ ) were removed based on database information (Y. A. Chen et al. 2013). The methylation data were preprocessed with the `wateRmelon` methylation-array-specific package from Bioconductor (Pidsley et al. 2013); the data was set as a `methylumiset` object. Samples and CpG sites with a technically poor quality were filtered out by excluding CpG sites with a bead count of  $<3$  in 5% of the samples and sites for which 1% of the samples showed a detection  $p$ -value  $>0.05$ . The method termed `dasen` was used to conduct background correction and quantile normalization individually for the two applied chemistries (Infinium I and II) as well as for the intensities of methylation (M) and un-methylation (U). Finally, the U and M intensities were transformed to  $\beta$  values using a standard equation:  $M/(M + U + \alpha)$ , where  $\alpha$  is the constant offset, 100, M is the methylated and U is the unmethylated probe intensity. Thus,  $\beta$  is the ratio of the methylated probe intensities to the overall intensities, and the range of resulting  $\beta$  values for a heterogenic tissue sample is between 0 (completely unmethylated, 0%) and 1 (completely methylated, 100%).

#### 4.2.1.4 DNA methylome age calculation (II-IV)

The DNA methylome ages (DNAMAge) of subjects in cohorts of V90\_2010 (Study II, III, IV), V90\_2014 (Study III), YFS\_1986 (Study III) and YFS\_2011 (Study III) were determined using the DNA methylome age calculator (<https://dnamage.genetics.ucla.edu/home>) (Horvath 2013). For the calculator, the raw  $\beta$  values of the selected probes ( $N = 28587$ ) in the 450Beadchip, including the 353 clock-CpG sites were extracted from all datasets, and those values were used as the input. The methylation data were put into the calculator in “a blind sample” manner, i.e. without any preliminary background information concerning sample type, chronological age or gender. Batch effects were normalized using the BMIQ function implemented in the DNAMAge algorithm. The difference between chronological age and DNAMAge,  $\Delta$ -cAge-DNAMAge was extracted for each subject. Quality of epigenetic age data was verified with sample type and gender (average methylation in X chromosome) predictions manifested in the DNA methylome age calculator.

#### 4.2.1.5 Estimation of the blood cell counts (I, III)

The CD8<sup>+</sup> and CD4<sup>+</sup> T cell, monocyte, granulocyte, NK cell and B cell type distributions in the samples of YFS\_1986 (Study III) and YFS\_2011 (Study I and III) populations were estimated using the R software estimation algorithm termed estimateCellCounts which is implemented in the minfi Bioconductor package (Jaffe and Irizarry 2014). The algorithm utilizes a subset of 600 control probes, representing specific DNA methylation signatures of different blood cell types. The reference data used in the estimation algorithm are available in the FlowSorted.Blood.450K Bioconductor package (Jaffe and Irizarry 2014). The data include methylomic data of purified cell populations (CD4<sup>+</sup> T cells, CD8<sup>+</sup> T cells, CD56<sup>+</sup> NK cells, CD19<sup>+</sup> B cells, CD14<sup>+</sup> monocytes, neutrophils, and eosinophils) from six healthy males aged  $38 \pm 13.6$  (Reinius et al. 2012).

#### 4.2.1.6 Predicting functional roles for regions of interest (I, IV)

In Studies I, III and IV, the annotation information for each CpG site was fetched from the GRCh36/hg18 genome assembly, dated March 2006. In Study I, the enriched gene ontology (GO) terms of the genes of interest were explored using an enrichment analysis tool, GOrilla (Eden et al. 2009), and the discovered significant terms were further clustered using the REVIGO tool (Supek et al. 2011). The GOrilla analysis was performed for the categories of process, function and component using two unranked lists: the first consisted of genes of interest, and the second of genes in the background ( $N = 20902$ ; analysis date, 9.3.2015). In addition, prediction

analyses of transcription factors for the groups of target genes were conducted using Pscan (JASPAR database; analysis date 10.3.2015) (Zambelli et al. 2009). The nominal p-value was set to the Bonferroni-corrected value of 0.05 in each analysis.

In Study IV, the pathway analysis tool IPA (QIAGEN Ingenuity Pathway Analysis (IPA®, QIAGEN Redwood City, [www.qiagen.com/ingenuity](http://www.qiagen.com/ingenuity)) was used in order to identify canonical pathways and networks for the mortality-associated genes harboring the CpG sites. If a CpG site was annotated to more than one gene, each of the genes were included in the analysis. Benjamini-Hochberg corrected p-value threshold was set to value 0.05.

## 4.2.2 Gene expression data (IV)

### 4.2.2.1 Origin of RNA

In Study IV, total RNA of V90\_2010 sample were extracted from the same PBMCs, which were subjected to DNA methylation and FACS analysis. Leukocyte separation was performed using a Ficoll-Paque density gradient (Ficoll-Paque™ Premium, cat. no. 17-5442-03, GE Healthcare Bio-Sciences AB, Uppsala, Sweden). The PBMC layer was collected, and the cells allocated for RNA extraction were suspended in 150 µl of RNeasy lysis solution (Qiagen, Crawley, UK). Following this, RNA extraction was performed using miRNeasy Mini Kit (Qiagen, Hilden, Germany).

### 4.2.2.2 Illumina gene expression assay

The gene expression data of V90\_2010 sample, which was used in Study IV were obtained from Illumina gene expression array, HumanHT-12 v4 Expression BeadChip (Cat no. BD-103-0204, Illumina Inc., CA, USA). The measurement was performed at the Core Facility of the Department of Biotechnology at the University of Tartu according to manufacturer's instructions. Preprocessing of the data was performed as previously described (Jylhava et al. 2014).

Briefly, the lumi pipeline was used; the background was corrected with the `bgAdjust.affy` package; further, the data were log<sub>2</sub>-transformed and quantile-normalized. Data with poor quality and background noise were filtered out as follows: probes exhibiting expression levels of < 5 or > 100 in more than five (3.3%) samples per transcript were excluded.

### 4.2.3 Fluorescence-activated cell sorting analysis (II, III, IV)

The PBMC samples of the V90 population were stored in the freezing solution and in liquid nitrogen prior to analysis (as described in 4.2.1.1). The proportions of different leukocyte types in these samples were determined using fluorescence-activated cell sorting analysis (FACS; BD FACSCanto II). With V90\_2010 samples, the results were analyzed with BD FACSDiva, version 6.1.3 (BD Biosciences, Franklin Lakes, NJ, USA), and with V90\_2014 samples, the results were analyzed with FlowJo software (Tree Star Inc., Ashland, OR, USA). The antibodies used were FITC-CD14 (cat. no. 11-0149), PerCP-Cy5.5-CD3 (45-0037), APC-CD28 (17-0289), PECD19 (12-0199) (eBioscience, San Diego, CA, USA), PECy<sup>TM</sup>7-CD4 (cat. no. 557852) and APC-Cy<sup>TM</sup>7-CD8 (557834) (BD Biosciences). In the analysis of V90\_2010 samples, the CD19 antibody was not used. To minimize nonspecific staining of the cells, staining was performed in phosphate-buffered saline (PBS) containing 1% fetal bovine serum (FBS) after an incubation step with Fc Receptor Binding Inhibitor (cat. no 16-9161, eBioscience). All samples were analyzed in a randomized order.

### 4.2.4 Cytomegalovirus titer (II)

In Study II, the anti-CMV titer was measured from the plasma samples of V90 population using an enzyme-linked immunosorbent assay (Enzygnost Anti-CMV/IgG, Siemens Healthcare, Marburg, Germany). According to manufacturer's protocol, sample was defined as CMV seropositive with serum anti-CMV immunoglobulin (Ig)G titer > 230. The measurement was performed in Fimlab Laboratories, Tampere, Finland.

### 4.2.5 Other covariate data (IV)

In Study IV, a set of anthropometric measures, functional performance and plasma biomarkers were utilized in the analysis (shown in Table 2). These data were available for 111 participants in V90\_2010. The collection of these variables, and the references therein, have been previously described elsewhere (Jylhava et al. 2014).

**Table 2.** The population (N=111) characteristics in Study IV. Distributions of the variables are shown according to the data at the 2.55-years mortality follow-up (table is modified from (Jylhava et al. 2016). The type of descriptive depends on the type or distribution of a variable. Symbols: \* = mean±sem is not valid, and thus median (Med) and IQR (interquartile range) values are used; <sup>a</sup> = percentage of live-gated cells; <sup>b</sup> = percentage of total T lymphocytes i.e. CD3<sup>+</sup>cells; <sup>c</sup> = percentage of CD4<sup>+</sup> cells; <sup>d</sup> = percentage of CD8<sup>+</sup> cells; SEM = standard error of mean

Variable	Non-survivors		Survivors	
	Mean/Med*	SEM/IQR/%	Mean/Med*	SEM/IQR/%
Women (n/%)	27	75	54	72
Age (months)	1079.5	0.61	1080.2	0.37
Systolic blood pressure (mmHg)	145	4.6	149	3.4
Diastolic blood pressure (mmHg)*	71.5	13.5	74	19
Weight (kg)	61.9	2.2	70.6	1.6
BMI (kg/m <sup>2</sup> )	24.3	0.75	27.5	0.54
Waist circumference (cm)	89.6	2.1	95.5	1.4
Hip circumference (cm)*	98	10	102	12
MMSE*	23.5	8	26	4
Barthel index*	95	20	95	5
Handgrip (kg)*	18	11	20	7
Able to perform chair-rise test (n=yes/%)	19	57.6	59	78.7
Able to perform chair-stand test (n=yes/%)	22	71	62	82.7
Frailty index (n/%)				
Non-frail	3	8.3	26	34.7
Pre-frail	22	61.1	37	49.3
Frail	11	30.6	12	16
CRP level (ng/ml)*	1.8	3.3	1.9	3.5
IL-1 $\beta$ level (pg/ml)*	14.2	27.6	19	34
IL-6 level (pg/ml)*	4.5	3.3	3.8	3.8
IL-7 level (pg/ml)*	7.8	5.3	6.4	5.2
IL-10 level (pg/ml)*	1.8	1.5	1.5	2.6
cf-DNA level ( $\mu$ g/ml)*	0.93	0.19	0.87	0.16
Unmethylated cf-DNA level ( $\mu$ g/ml)*	0.75	0.2	0.67	0.15
Plasma mitochondrial DNA (copy number)*	4.3E+08	2.4E+08	3.8E+08	2.1E+08
<i>Alu</i> repeat cf-DNA (GE)*	74.4	50.4	66.8	38.3
Dehydroepiandrosterone sulfate ( $\mu$ g/ml)*	0.25	0.48	0.25	0.31
Cortisol (ng/ml)*	133	54.3	117	68
Indoleamine 23-dioxygenase activity*	44.3	25.5	51.8	25.3
Anti-CMV antibody titer*	19	8	19	9000
Anti-Epstein-barr virus antibody titer*	405	315	410	410
Epigenetic age in years	76.1	1.04	76.1	0.64
CD3 <sup>+</sup> cells (%) <sup>a</sup>	62	15.8	57.9	12
CD4 <sup>+</sup> cells (%) <sup>b</sup>	62.9	2.5	63.8	1.6
CD8 <sup>+</sup> cells (%) <sup>b</sup>	30.6	2.3	28.9	1.5
CD4 <sup>+</sup> /CD8 <sup>+</sup> cells (ratio)*	2.4	2.3	2.3	2.4
CD4 <sup>+</sup> CD28 <sup>-</sup> cells (%) <sup>c</sup>	9.2	16.2	9.2	13

#### 4.2.6 Availability of raw data (I-IV)

The genome-wide methylation data of YFS\_2011 (Study I) and V90\_2010 (Study IV) are available in the GEO database (<http://www.ncbi.nlm.nih.gov/geo/>) under accession number GSE69270 and GSE68194, respectively. Analogously, the transcriptomic data of V90\_2010 (Study IV) are under accession number GSE65218. The data used in Study III are available as electronic Supplementary Material in the article (Kananen et al. 2016).

#### 4.2.7 Testing hypotheses (I-IV)

In Study I, to explore aging-associated CpG sites using EWAS approach, the association between chronological age and DNA methylation level in each CpG site was analyzed using a generalized linear regression termed beta regression. Gender and estimated cell type proportions were adjusted for (Ferrari 2004; Francisco Cribari-Neto 2010). In Study IV, in order to explore methylomic predictors for mortality, three analysis stages took place: (1) Gender and the proportions of blood cell types were adjusted for, after which the standardized weighted residuals were extracted and used in the following stages. (2) Cox univariate assessment was used to pre-select the most promising methylomic predictors. (3) The predictive capacity of large group biomarkers were analyzed using penalized Ridge regression (Hoerl AE 2000). The final mortality-predicting signature shown in Table 10 was assessed using a Cox regression model. The models were evaluated with the goodness of fit statistics, Akaike Information criterion (AIC), and Harrel's C for the discriminative power. To correct for multiple testing of genome-wide data (Studies I and IV), a Benjamini-Hochberg method was used. In Studies II and III, as appropriate, the analyses of the DNA methylome ages included non-parametric between-group comparisons (Mann-Whitney U tests), Spearman's rank sum tests and parametric multivariate linear regression analyses. In Studies I-IV, statistical analyses and related visualizations were performed using the R software (R >= 2.15.3), Stata software (v13.0 for Windows, StataCorp LP, TX, USA) and IBM SPSS Statistics v.22 (IBM Corporation, Armonk, NY, USA); and the nominal P-value threshold was set to 0.05.

## 5 RESULTS

### 5.1 The Studies

In the thesis, aging-associated DNA methylation level changes (I), and mortality predictors (IV) at single CpG site resolution were characterized. The longitudinal behavior of the DNA methylome age (III), and its role as a mortality predictor (IV) was explored. Furthermore, the epigenetic age change associated with CMV infection was analyzed (II). In all sections in the thesis, the role of blood cell sample heterogeneity was considered (I-IV). Specifically, the relevance of controlling sample heterogeneity in EWASs was evaluated (I), the association between the DNA methylome age and blood cell subtype counts was explored (III), and the overall longitudinal behavior of the blood cell subtypes was investigated (III).

### 5.2 Aging-associated DNA methylation changes in middle-aged individuals (I)

In Study I, in the cross-sectional EWAS, 1202 aging-associated CpG sites were identified from whole blood samples of YFS\_2011 population where the age range was between 40-49 years. The hypermethylated and hypomethylated CpG sites were annotated on 437 and on 440 genes, respectively. The top 20 lists of the most significant hypermethylated and hypomethylated CpG sites associated with aging are shown in Table 3 and 4. Moreover, the CpG sites, which are located in genes with a greater number of aging-associated CpG sites, were more often hypermethylated than hypomethylated with increasing chronological age (Figure 7).

As a remark on quality control, well-known CpG sites which are hypermethylated with advanced age were among sites identified in Study I: cg16867657, cg24724428 and cg21572722 located in gene *ELOVL2*, cg06639320, cg22454769 and cg24079702 located in gene *FHL2*, cg16219603, cg16419235 located in gene *PENK*, and cg08097417, cg09499629 and cg07955995 located in gene *KLF14* (Heyn et al. 2012; Garagnani et al. 2012; Hannum et al. 2013; Florath et al. 2014; Steegenga et al. 2014). Moreover, acknowledged smoking-associated hypomethylation of cg03636183 located in gene *F2RL3* (Breitling et al. 2011; Zhang et al. 2014) was seen in the quality control exploration of Study I.

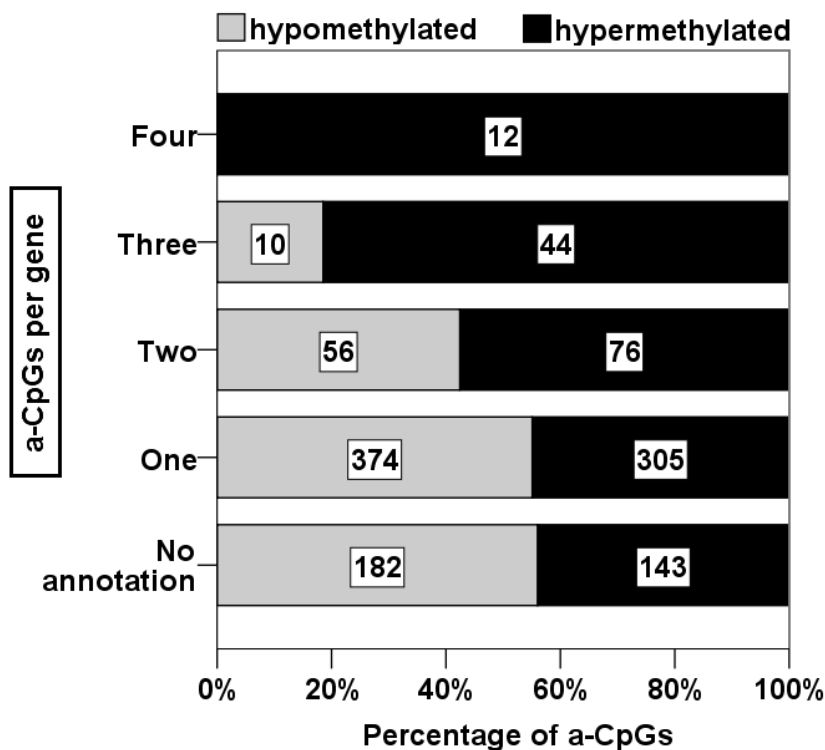
**Table 3.** The top 20 hypermethylated aging-associated CpG sites in middle-aged individuals. The top-ranking hypermethylated CpG sites were selected with the following criteria: 1) positive direction of the association based on the value of beta regression (denoted as 'betareg') estimate of age; 2) more than one hit identified per gene (q-value<0.05 which corresponds to false discovery rate <5%) and 3) the top-ranking q-values. The q-value denotes the Benjamini-Hochberg-corrected p-value. (Study I)

ProbeID	Gene	Chr	Coordinate36	Betareg estimate of age	q-value
cg16867657	<i>ELOVL2</i>	6	11152863	0.022	0.00E+00
cg24724428	<i>ELOVL2</i>	6	11152874	0.021	4.80E-07
cg21572722	<i>ELOVL2</i>	6	11152880	0.013	3.46E-06
cg06639320	<i>FHL2</i>	2	105382171	0.018	3.46E-06
cg00059225	<i>GLRA1</i>	5	151284550	0.013	5.13E-06
cg08097417	<i>KLF14</i>	7	130069673	0.020	1.87E-05
cg22454769	<i>FHL2</i>	2	105382199	0.021	5.03E-05
cg07553761	<i>TRIM59</i>	3	161650671	0.016	6.12E-05
cg01588592	<i>ETV3L</i>	1	155335949	0.011	1.14E-04
cg11176990	<i>LOC375196</i>	2	39041037	0.014	1.54E-03
cg09499629	<i>KLF14</i>	7	130069676	0.018	1.54E-03
cg22158769	<i>LOC375196</i>	2	39041043	0.020	2.43E-03
cg18898125	<i>NEFM</i>	8	24826286	0.012	2.49E-03
cg21911021	<i>ZIK1</i>	19	62786823	0.020	3.07E-03
cg27217742	<i>RGS12</i>	4	3335078	0.013	3.07E-03
cg17737681	<i>DLX1</i>	2	172660382	0.015	3.29E-03
cg24079702	<i>FHL2</i>	2	105382203	0.015	5.99E-03
cg16219603	<i>PENK</i>	8	57523140	0.013	7.00E-03
cg23930856	<i>TFAP2B</i>	6	50919683	0.013	7.22E-03
cg11152943	<i>TRAPPC9</i>	8	141318170	0.013	7.57E-03



**Table 4.** The top 20 hypomethylated aging-associated CpG site in middle-aged individuals. The top-ranking hypomethylated sites were selected with following criteria: 1) negative direction of the association based on the value of beta regression (denoted as 'betareg') estimate of age; 2) more than one hit identified per gene (q-value<0.05 which corresponds to false discovery rate<5%) and 3) the top-ranking q-values. The q-value denotes the Benjamini-Hochberg-corrected p-value. (Study I)

ProbeID	Gene	Chr	Coordinate36	Betareg estimate of age	q-value
cg00791074	<i>MTHFD1L</i>	6	151227862	-0.018	7.51E-04
cg18618815	<i>COL1A1</i>	17	45630323	-0.018	5.99E-03
cg14169886	<i>PRDM16</i>	1	3101709	-0.014	5.99E-03
cg01820374	<i>LAG3</i>	12	6752344	-0.014	9.24E-03
cg19421125	<i>LAG3</i>	12	6753117	-0.022	1.02E-02
cg14829066	<i>NTRK3</i>	15	86360145	-0.013	1.49E-02
cg03290281	<i>C6orf195</i>	6	2577606	-0.021	1.49E-02
cg05561193	<i>DCLK2</i>	4	151218492	-0.017	1.96E-02
cg20249566	<i>NWD1</i>	19	16691739	-0.024	1.97E-02
cg23928726	<i>PEX10</i>	1	2334858	-0.014	1.97E-02
cg20007894	<i>SCAND3</i>	6	28648421	-0.019	2.08E-02
cg16355231	<i>PEX10</i>	1	2334839	-0.019	2.14E-02
cg15058210	<i>HDAC4</i>	2	239861814	-0.018	2.16E-02
cg06030846	<i>TMEM108</i>	3	134581182	-0.011	2.16E-02
cg25994988	<i>UBASH3B</i>	11	122157592	-0.011	2.16E-02
cg18345924	<i>NCAM2</i>	21	21294102	-0.016	2.18E-02
cg00638021	<i>COL1A1</i>	17	45622061	-0.013	2.26E-02
cg19344626	<i>NWD1</i>	19	16691749	-0.024	2.36E-02
cg01288258	<i>ITFG2</i>	12	2792128	-0.011	2.41E-02
cg05221385	<i>TAF10</i>	11	6590080	-0.010	2.43E-02



**Figure 7.** Numbers of aging-associated CpG sites (a-CpG hits) per gene in regard to hypermethylation and hypomethylation are visualized as bars. Aging-associated hypermethylation was more frequent within genes with more association hits. First, the genes were categorized into groups based on the number of hypermethylated or hypomethylated a-CpG hits per gene. Next, the frequencies of hypermethylated and hypomethylated a-CpGs within the groups were calculated. The number of a-CpGs hits for each group is shown inside each bar. (Study I)

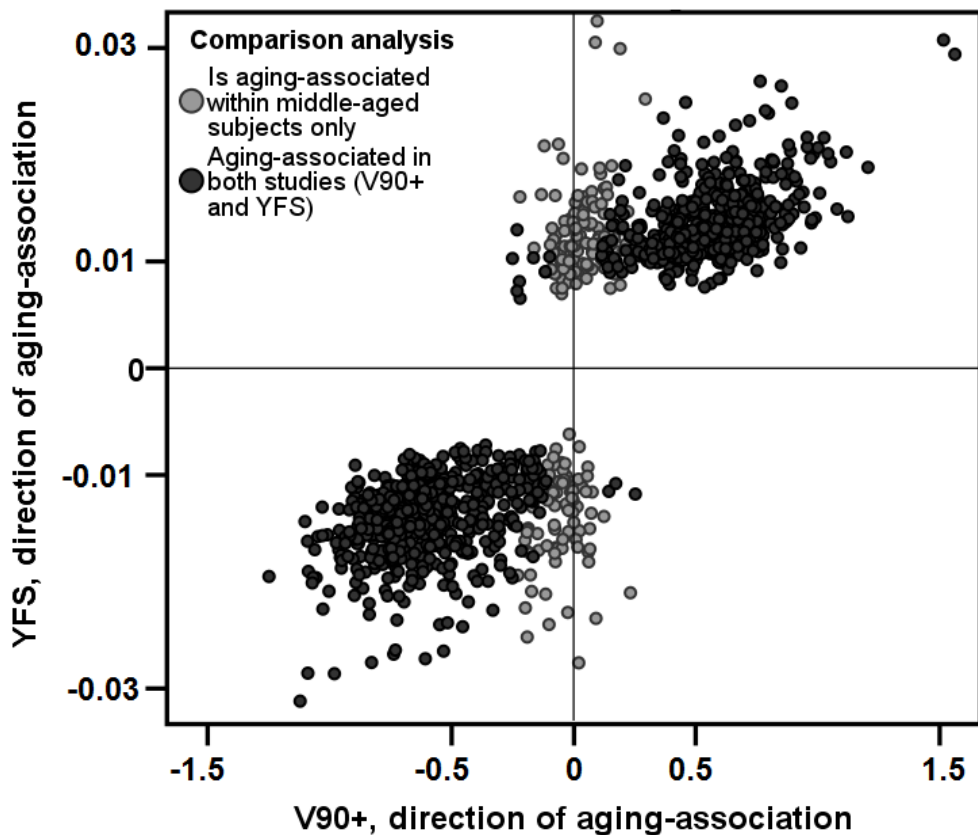
The functional roles of the genes with aging-associated CpG sites were explored using the gene ontology (GO) term enrichment analysis (Eden et al. 2009). The results demonstrated that regions with aging-associated hypermethylation were enriched to 73 GO process terms and to eight GO function terms whereas hypomethylated sites showed no enrichment. The most significant GO processes (anatomical structure development, GO:0048856,  $P=1.02 \times 10^{-11}$ ; morphogenesis GO:0009653,  $P = 5.02 \times 10^{-10}$ ) clustered under the term 'developmental process'. The GO function terms were primarily centered on DNA binding. In addition, a Pscan analysis tool (Zambelli et al. 2009) was used to predict common regulators for genes with the aging-associated methylation sites. Genes with hypermethylation-associated CpG sites were predicted to be regulated by 11 common transcription factors in categories of Zinc-coordinating, Winged Helix-Turn-Helix and Zipper-Type. Conversely, for genes with hypomethylation-associated CpG sites, no common transcription factors were found.

## 5.2.1 Comparison analysis of different studies

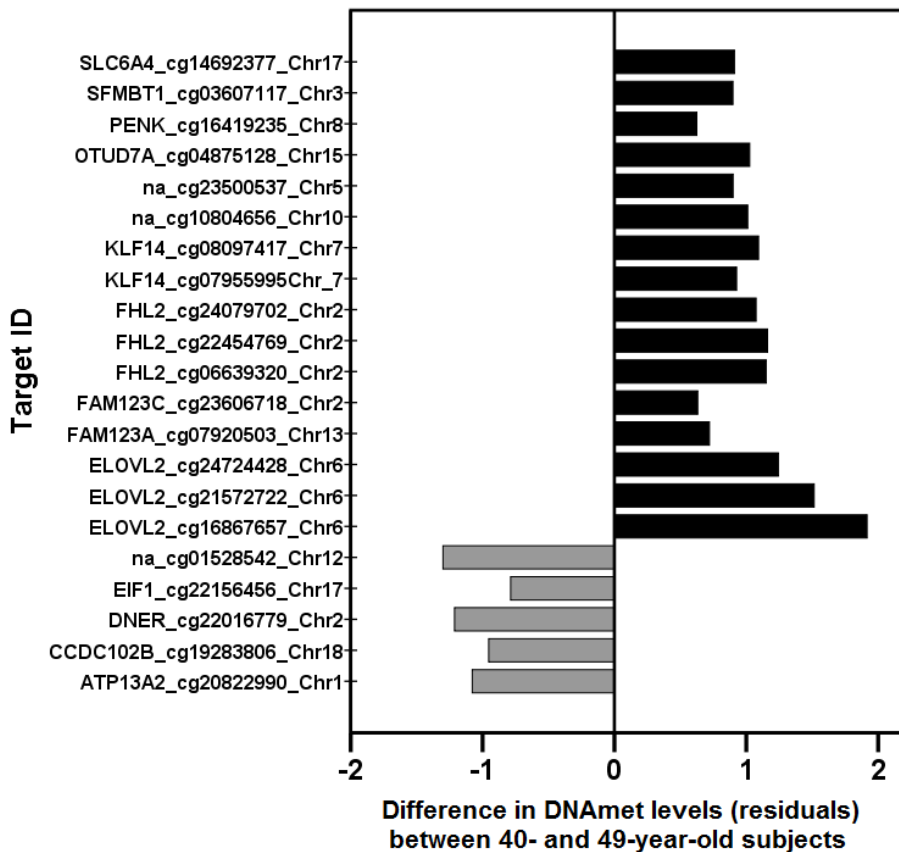
The list of aging-associated CpG sites identified among the middle-aged population in Study I was compared to other lists of previously reported aging-associated CpG sites. Specifically, the repeatability of these aging-associations was explored in two ways: (option 1) using studies with practically identical bioinformatics analysis pipeline with blood cell subtype count considerations, and (option 2) using studies with varying analysis pipelines without cell subtype count considerations.

In practice, first comparison (option 1) was performed with a list of CpG sites where the DNA methylation differences by age were examined between young adults and nonagenarians (Marttila et al. 2015). The age difference between the age groups in the differential analysis by Marttila et al. (2015a) was 60 years. As shown in Figure 8, the comparison analysis revealed that of the 1202 aging-associated CpG sites identified in the Study I, Marttila et al. (2015) also identified 987 as differentially methylated with similar directions of changes.

A second comparison analysis (option 2) was performed with other previously reported aging-associated CpG site lists by Hannum et al. (number of aging-associated CpG sites was 89) (Hannum et al. 2013), Garagnani et al. (number of hits, 9) (Garagnani et al. 2012) and Florath et al. (number of hits, 162) (Florath et al. 2014) (summarized by Steegenga et al. (Steegenga et al. 2014)). In these analyses, the array technology and sample origin was the same as in Study I, but the heterogeneity of the blood samples was not considered. The corresponding age ranges of these samples were between 19 and 101, 9 and 83, and 50 and 75 years, respectively. When two or more of these studies in addition to Study I and the one of Marttila et al. (2015) were considered, the comparison analysis demonstrated that there were only 21 aging-associated CpG sites in common (Figure 9).



**Figure 8.** The majority of aging-associated methylation changes in the CpG sites in the comparison analysis (option 1) behaved concordantly. The data is visualized as a scatterplot where each dot corresponds to a single CpG site; directions of associations correspond to estimates of age, which are fetched from the regression models, and negative values point to hypomethylation and positive values to hypermethylation. In the first comparison analysis (option 1), Study I with YFS\_2011 and earlier aging-association study with V90\_2010 (Marttila et al. 2015) were compared, and of the 1202 sites found in Study I, 987 CpG sites were analogously associated with aging in both of these studies. The regression analyses in the studies were adjusted for blood cell subtype variation, and the studies consisted of samples with distinct age ranges: there were 40- to 49-year-old subjects in the YFS\_2011 whereas in the V90\_2010, there were 19-30- and 90-year-old subjects (Marttila et al. 2015). (Study I)



**Figure 9.** The top 21 most frequently reported aging-associated CpG sites and their direction of association with aging. In the second comparison analysis (option 2), the top 21 sites were selected with the following criteria. (1) CpG site was identified in Study I and in the study by Marttila et al. (2015a) with V90\_2010 population, and also in two or more other studies. (2) The sites were reported as aging-associated in blood samples and the data were obtained using 450BeadChip technology. Here, methylation level differences in YFS population between the highest and the lowest age groups i.e. between 40- and 49-year-old individuals; (calculated from the medians of residuals after adjusting for effects of gender and cell subtype proportions), are illustrated as bars. The bars are colored according to the hypomethylation or hypermethylation status (grey = hypomethylated, black = hypermethylated). Gene annotation is shown for each bar, where applicable (na = no gene annotation). (Study I)

### 5.3 DNA methylome ages (II, III, IV)

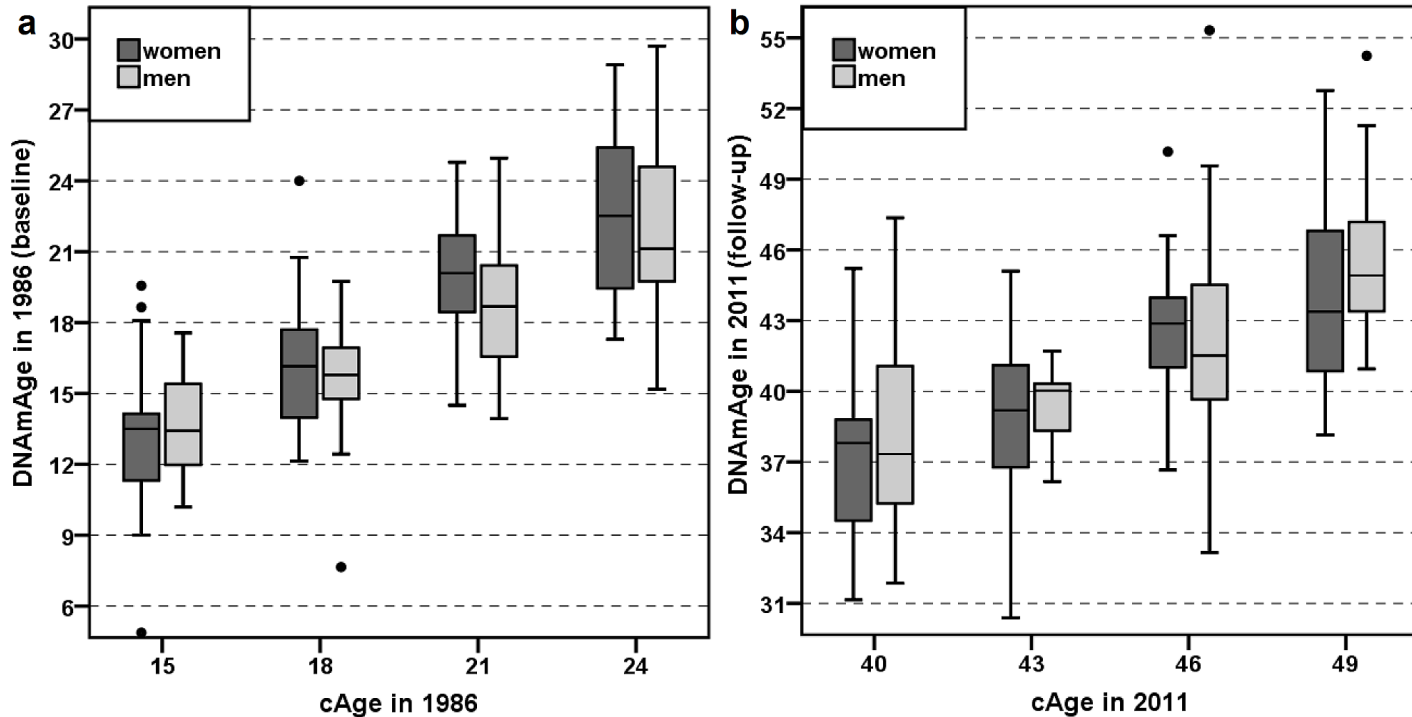
The epigenetic age (DNAmAge) was quantified for each subject in V90 and YFS populations at baseline and follow-up time points using a multi-tissue age predictor termed Horvath's DNA methylome age calculator (Horvath 2013). At first stage in the analysis of Study III, the DNAmAge values were compared to subjects' chronological ages. Overall, as shown in Table 5 and Figure 10, the median values of DNAmAge were less than the corresponding chronological ages within most, but not all, subjects: the median DNAmAges of the subjects aged 19-34 years in the V90\_2014 cohort were higher than their chronological ages.

The relationships between chronological age and DNAmAge values were explored cross-sectionally with Spearman's rank sum correlation analysis. In the YFS\_1986 population, the coefficients were 0.785 ( $P=8.10 \times 10^{-25}$ ) and 0.702 ( $P=1.25 \times 10^{-11}$ ) among women and men, respectively. The corresponding values in the YFS\_2011 population were 0.635 ( $P=4.24 \times 10^{-14}$ ) and 0.591 ( $P=7.36 \times 10^{-8}$ ). In the V90\_2010 population, the coefficients were among women and men 0.594 ( $P=3.6 \times 10^{-11}$ ) and 0.661 ( $P=3.4 \times 10^{-6}$ ), and in V90\_2014 population the coefficients were 0.925 ( $P=5.4 \times 10^{-24}$ ) and 0.903 ( $P=1.2 \times 10^{-10}$ ), respectively.

#### 5.3.1 Follow-up of DNA methylome age values (III)

In Study III, the quantities of DNA methylome aging during follow-up periods were calculated (DNAmAge change = DNAmAge at follow-up time point – DNAmAge at baseline time point) for YFS and V90 populations. When the subjects in V90 population were categorized as nonagenarians ( $N=41$ ) and young controls ( $N=7$ ), the median  $\pm$ MAD ( $\pm$ median absolute difference) values of DNAmAge change from the year 2010 to 2014 were  $4.53 \pm 3.52$  and  $7.39 \pm 2.11$  years, respectively. The median value of DNAmAge changes from 1986 to 2011 in YFS population ( $N=183$ ) was  $23.26 \pm 2.97$ . In order to compare the results of YFS to V90, this change corresponds to 3.7 years of methylome aging in four chronological years.

In addition, the difference between chronological age and DNA methylome age ( $\Delta$ -cAge-DNAmAge at baseline and follow-up time points for each subject) was calculated, and the relationship between these two values in both populations is presented in Figure 11. In the YFS population, the Spearman's rank sum correlation coefficient between the  $\Delta$ -cAge-DNAmAge values in 1986 and 2011 was 0.535 ( $P=6.1 \times 10^{-15}$ ) for all subjects ( $N=183$ ). The corresponding Spearman's rank sum correlation coefficient value of V90 population ( $N=48$ ) between the  $\Delta$ -cAge-DNAmAge values in 2010 and 2014 was 0.895 ( $P=9.2 \times 10^{-18}$ ).

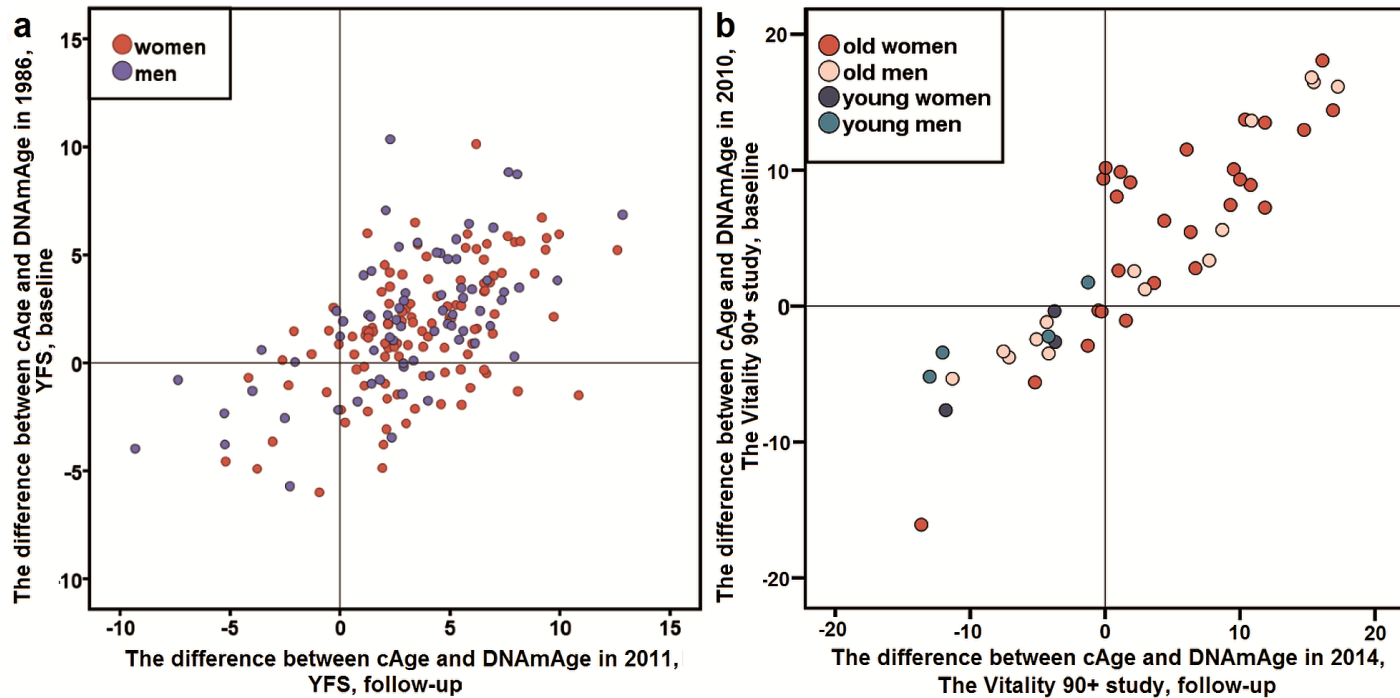


**Figure 10.** The DNAmAge values in the YFS samples at baseline and after 25 years are visualized as boxplots. The YFS\_1986 sample (N=183) is shown in panel a, and the YFS\_2011 sample (N=183) is shown in panel b. The subjects were organized into separate groups according to chronological age (cAge) and gender. The differences in the median DNAmAge values between men and women at each cAge group were not significant (Mann-Whitney U-test,  $P > 0.05$ ). (Study III)

**Table 5.** Summary of DNAmAge values and PBMC proportions for the V90 sample collected in 2010 and 2014. The values were categorized by chronological age group and gender. Table shows the median (median absolute deviation) values. Data not available are denoted as 'na'. (Study III)

Sample collection year	Group	N	Chronological age	DNAmAge of PBMCs	CD3+ of live-gated cells (%)	CD4+ of CD3+ (%)	CD8+ of CD3+ (%)	CD4+: CD8+ ratio	CD28- of CD4+ (%)	CD28- of CD8+ (%)	CD19 of live-gated cells (%)	CD14+ of live-gated cells (%)
2010	19-28 years, women	14	22.5 (3.0)	22.3 (2.5)	71.6 (3.3)	59.0 (4.5)	30.1 (6.0)	2.1 (0.5)	0.2 (0.1)	14.7 (6.8)	na	2.0 (1.11)
	20-29 years, men	7	24.0 (4.5)	22.5 (4.4)	67.0 (3.0)	60.2 (2.7)	31.8 (4.7)	1.9 (0.5)	0.3 (0.4)	18.1 (13.3)	na	2.4 (0.44)
	90 years, women	84	90.0 (0)	81.7 (7.3)	60.5 (9.6)	65.6 (15.6)	27.1 (13.0)	2.4 (1.7)	9.3 (9.7)	65.3 (17.5)	na	8.5 (4.52)
	90 years, men	31	90.0 (0)	87.2 (5.3)	54.3 (10.2)	57.2 (12.5)	28.7 (13.0)	2.0 (1.1)	9.2 (8.5)	69.0 (16.0)	na	7.5 (4.30)
2014	19-34 years, women	27	26.0 (4.4)	29.4 (6.8)	67.0 (7.0)	58.9 (5.6)	30.2 (6.8)	1.9 (0.5)	2.3 (2.4)	30.0 (10.8)	5.8 (2.5)	4.4 (4.6)
	22-32 years, men	13	28.0 (4.4)	35.1 (9.2)	59.0 (6.2)	55.1 (4.2)	32.8 (4.7)	1.6 (0.3)	1.4 (1.9)	39.2 (17.0)	8.0 (4.4)	8.5 (9.9)
	94 years, women	27	94 (0)	89.6 (7.2)	58.8 (11.7)	49.1 (18.2)	35.6 (19.1)	1.3 (1.1)	14.3 (12.2)	74.4 (16.5)	2.8 (2.3)	8.4 (8.7)
	94 years, men	14	94 (0)	91.5 (11.8)	62.5 (8.9)	58.3 (19.3)	28.1 (14.6)	2.1 (1.5)	10.0 (11.6)	75.0 (16.1)	2.2 (0.7)	10.8 (11.6)

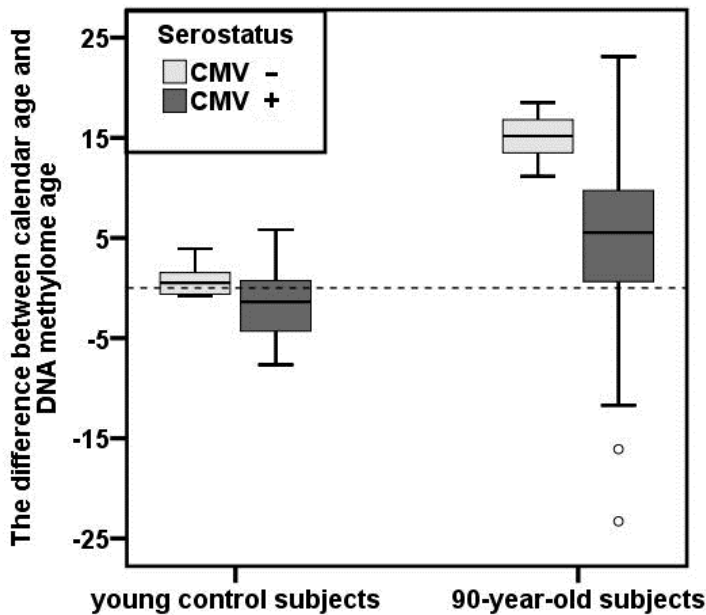




**Figure 11.** The correlations between the  $\Delta$ -cAge-DNAge values at baseline and follow-up time points are shown in scatterplots. Panel a: Subjects participating in the YFS in 1986 and 2011. The Spearman's rank sum correlation coefficient between the  $\Delta$ -cAge-DNAge values in 1986 and 2011 was 0.535 ( $P=6.1 \times 10^{-15}$ ) for all subjects ( $N=183$ ). Panel b: Subjects participating in the V90+ study in 2010 and 2014. The Spearman's rank sum correlation coefficient value of V90+ population ( $N=48$ ) between the  $\Delta$ -cAge-DNAge values in 2010 and 2014 was 0.895 ( $P=9.2 \times 10^{-18}$ ). The colors of the dots for each subject are explained in the graph legend. (cAge=chronological age) (Study III)

### 5.3.2 Cytomegalovirus-association with epigenetic aging (II)

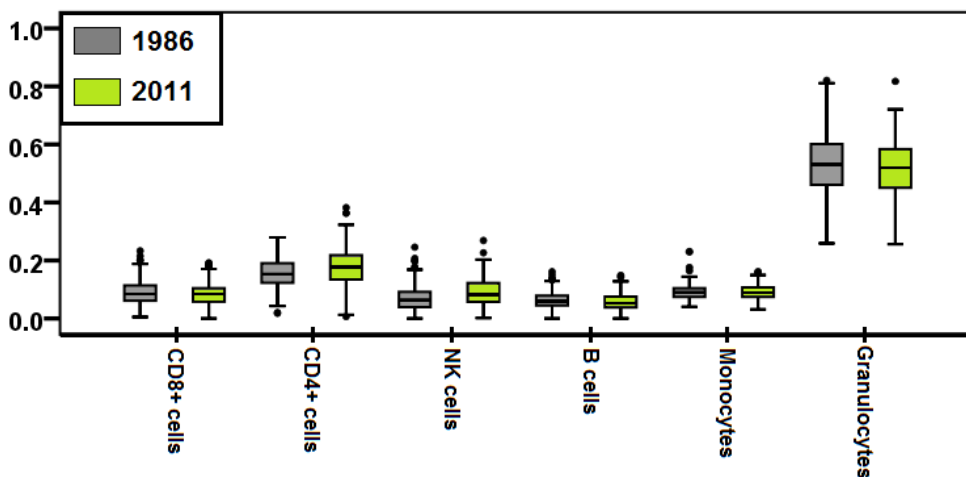
In Study II, the CMV seropositivity (CMV<sup>+/−</sup>) of subjects in V90 population was measured. The V90\_2010 population was analyzed in separate chronological age categories of young (N=21) and 90-year-old (N=122) subjects. Of the nonagenarians, there were 6 CMV<sup>−</sup> and 116 CMV<sup>+</sup> subjects, and of the young controls, there were 9 CMV<sup>−</sup> and 12 CMV<sup>+</sup> subjects. The  $\Delta$ -cAge-DNA<sub>m</sub>Age values of CMV serostatus positive and negative subjects were compared, as shown in Figure 12. The analysis revealed that CMV<sup>+</sup> subjects demonstrated an increase in epigenetic aging when compared to CMV<sup>−</sup> subjects in both chronological age categories (P<0.05). In V90\_2014, there were an insufficient number of survived nonagenarian CMV<sup>−</sup> subjects for a statistical analysis. However, CMV<sup>+</sup> nonagenarians and young controls in V90\_2014 showed a trend toward increased epigenetic aging when compared to CMV<sup>−</sup> subjects (young controls: Mann–Whitney U-test, P=0.191, nonagenarians P=na).



**Figure 12.** The epigenetic aging of CMV seropositive subjects was increased in a statistically significant level in both chronological age categories (21 young controls and 122 nonagenarians; within both categories, CMV<sup>+</sup> and CMV<sup>−</sup> were compared; Mann–Whitney U-test, P<0.05). (Study II)

## 5.4 The effect of blood cell subtypes (I-IV)

The blood cell counts in Studies I-IV were determined either using genome-wide DNA methylation profiles together with a specific cell count estimation algorithm (YFS) or FACS analysis (V90). The cell compositions of V90 and YFS samples are summarized in Table 5 and Figure 13, respectively. Specifically, the aging-associated shift in blood cell subtype proportions is demonstrated in the results shown in Table 5. The difference between CD28-CD4<sup>+</sup>, CD28-CD8<sup>+</sup> and CD14<sup>+</sup> of live-gated cells between young controls and nonagenarians stands out the most. This is seen in both time points, in 2010 and in 2014.



**Figure 13.** Estimated cell counts of CD8T, CD4T, NK, B cell, monocytes and granulocytes in samples of YFS populations (N=183) at different time points (in 1986 and 2011, grey and green, respectively) are visualized as boxplots. (Studies I, III)

Study I highlighted that a large proportion (24%) of the variance in the genome-wide DNA methylation data is explained by the variability in blood cell subtypes. This was explored in a correlation analysis where at first, principal components (PCs) of the genome-wide DNA methylation data were defined with principal component analysis (PCA), and then the PCs were correlated against the cell count estimates of monocytes, granulocytes, CD8, CD4, NK and B cells. The results showed that the main principal components (PCs) correlated highly ( $-0.5 > \text{correlation coefficients} > 0.5$ ) with the blood cell subtype counts. Similar correlation analysis with DNA methylation data of the 1202 aging-associated CpG sites alone revealed that the cell counts were important determinants of variation also in that data. In the end, the cell count estimates were adjusted for in the final regression analyses where the aging-association and mortality-association hypotheses were tested one by one (Studies I and IV).

**Table 6.** The associations between PBMC subtype proportions and  $\Delta$ -cAge-DNAAge values in different age groups in V90 sample at baseline (in 2010) and follow-up (in 2014). Additional young control subjects (N = 33) were recruited at follow-up in 2014. The cell counts were determined using FACS, and the associations were determined using Spearman rank-sum correlation analysis; for the purpose of interpretation, as a clarification: larger and smaller  $\Delta$ -cAge-DNAAge values are referred to as 'younger' and 'older' DNA methylome age, respectively. Data not available are denoted as 'na'. Associations with  $P < 0.05$  are denoted by '\*'. (Study III)

	Nonagenarians				Young controls			
	$\Delta$ -cAge-DNAAge at baseline in 2010, N=115		$\Delta$ -cAge-DNAAge at follow-up in 2014, N=41		$\Delta$ -cAge-DNAAge at baseline in 2010, N=21		$\Delta$ -cAge-DNAAge at follow-up in 2014, N=40	
Cell type proportion (%)	r	p	r	p	r	p	r	p
CD3+ of live-gated cells	0.106	0.261	-0.251	0.114	-0.065	0.780	0.404*	<b>0.010</b>
CD4+ of CD3+	0.316*	<b>0.001</b>	0.311*	<b>0.048</b>	0.505*	<b>0.019</b>	0.382*	<b>0.015</b>
CD8+ of CD3+	-0.310*	<b>0.001</b>	-0.326*	<b>0.038</b>	-0.233	0.309	-0.232	0.150
CD4+:CD8+ ratio	0.309*	<b>0.001</b>	0.346*	<b>0.027</b>	0.396	0.075	0.332*	<b>0.036</b>
CD4+ CD28-	-0.540*	<b>4.75e-10</b>	-0.626*	<b>1.20e-5</b>	-0.222	0.333	-0.344*	<b>0.030</b>
CD8+ CD28-	-0.269*	<b>0.004</b>	-0.327*	<b>0.037</b>	-0.369	0.100	-0.348*	<b>0.028</b>
CD14+ of live-gated cells	-0.112	0.235	0.033	0.836	-0.139	0.549	0.077	0.635
CD19+ of live-gated cells	na	na	0.246	0.122	na	na	-0.025	0.879

**Table 7.** The associations of whole blood cell subtype counts with  $\Delta$ -cAge-DNAAge values at baseline in 1986 (cAge 15-24 years) and at follow-up in 2011 (cAge 40-49 years) of the subjects in the YFS. The cell counts were determined using genome-wide DNA methylation profiles and a specific cell count estimation algorithm (Jaffe and Irizarry 2014). Associations were determined using Spearman rank-sum correlation analysis; larger and smaller  $\Delta$ -cAge-DNAAge values are referred to as 'younger' and 'older' DNA methylome age, respectively. Associations with a p-value less than 0.05 are denoted by '\*'. (Study III)

Cell type proportion (%)	YFS, $\Delta$ -cAge-DNAAge in 1986, N=183		YFS, $\Delta$ -cAge-DNAAge in 2011, N=183	
	R	p	r	p
CD8+ T cells	-0.085	0.255	-0.089	0.231
CD4+ T cells	0.150*	<b>0.044</b>	0.093	0.211
CD4+:CD8+ ratio	0.106	0.16	0.224*	<b>0.003</b>
NK cells	-0.024	0.747	-0.192*	<b>0.009</b>
B cells	0.106	0.152	0.115	0.12
Monocytes	-0.028	0.711	-0.015	0.84
Granulocytes	0.029	0.7	0.054	0.465

**Table 8.** Correlations of the PBMC subtype proportions of subjects (N=63) in the V90 between 2010 and 2014. The cell counts of the samples were determined using FACS. Associations were determined using Spearman rank-sum correlation analysis. Associations with a p-value less than 0.05 are denoted by '\*'. Correlation analysis was performed on the total sample without further categorization due to a small sample size. (Study III)

Cell type proportion (%)	r	P
CD3+ of live-gated cells	0.628*	2.26E-06
CD4+ of CD3+	0.830*	5.34E-13
CD8+ of CD3+	0.855*	2.13E-14
CD4+:CD8+ ratio	0.811*	7.93E-16
CD4+ CD28-	0.901*	9.48E-24
CD8+ CD28-	0.838*	2.06E-13
CD14+ of live-gated cells	0.317*	3.01E-02

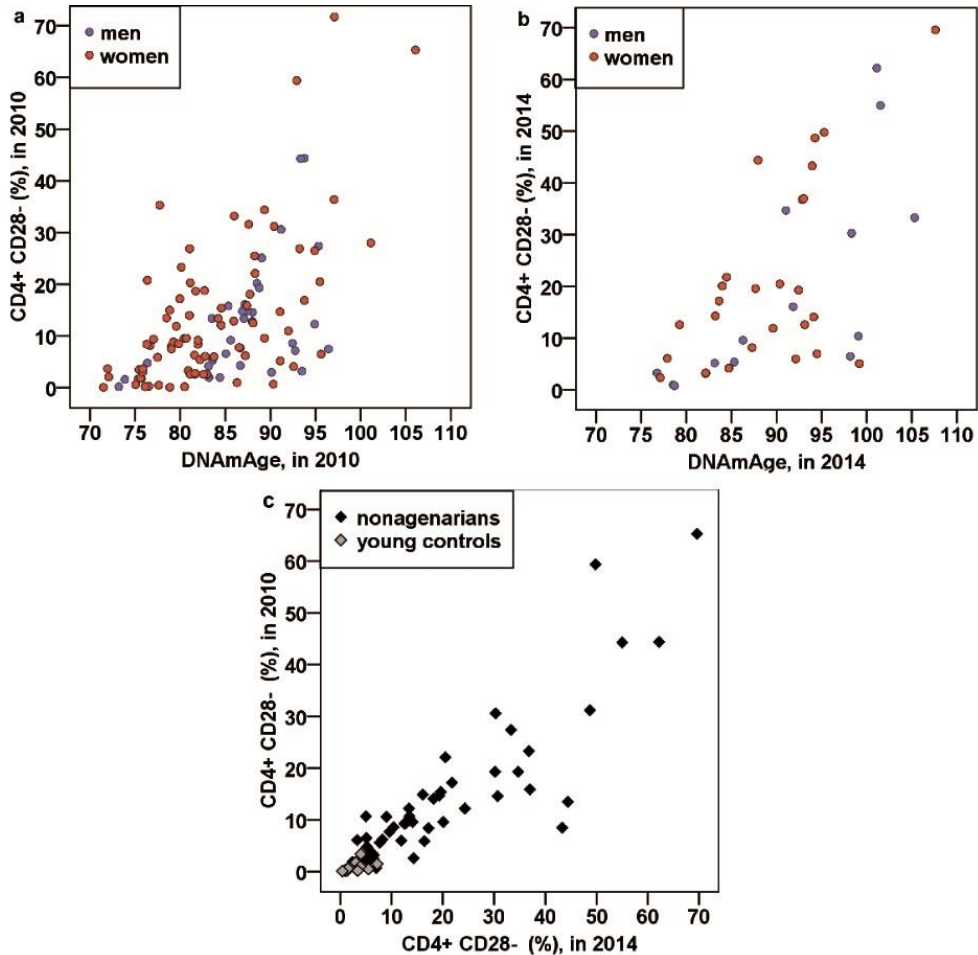
**Table 9.** Correlations of blood cell subtype counts for subjects in the YFS population between 1986 and 2011. The cell counts were determined using genome-wide DNA methylation profiles and a specific cell count estimation algorithm (Jaffe and Irizarry 2014). Associations were determined using Spearman rank-sum correlation analysis. Associations with a p-value less than 0.05 are denoted by '\*'. (Study III)

Cell type proportion (%)	YFS, all N=183		YFS, women N=113		YFS, men N=70	
	R	p	r	p	r	p
CD8+ T cells	0.317*	1.20E-05	0.356*	1.35E-04	0.261*	2.90E-02
CD4+ T cells	0.312*	1.57E-05	0.349*	1.86E-04	0.360*	2.19E-03
CD4+:CD8+ ratio	0.453*	1.97E-10	0.464*	4.12E-07	0.511*	1.00E-05
NK cells	0.398*	2.18E-08	0.430*	2.79E-06	0.312*	8.55E-03
B cells	0.545*	1.33E-15	0.484*	8.69E-08	0.589*	8.39E-08
Monocytes	0.547*	9.73E-16	0.561*	1.79E-10	0.521*	3.71E-06
Granulocytes	0.206*	4.92E-03	0.268*	4.60E-03	0.08	0.509

In Study II, the effect of PBMC subtype proportions on DNAmAge in V90\_2010 samples was evaluated using the linear multivariate regression model; the DNA methylome age was explained with the cell subtypes and CMV serostatus. The analysis showed that the proportion of CD4<sup>+</sup>CD28<sup>-</sup> determines most of the variation in the DNAmAge, and is a more important covariate than the CMV serostatus or all the other blood cell subtypes.

In Study III, the associations between the blood cell subtype counts and  $\Delta$ -cAge-DNAmAge values in V90 and YFS samples were determined using Spearman rank sum correlation analysis at each time point (shown in Tables 6 and 7). A large proportion of the cell subtypes correlated significantly with the  $\Delta$ -cAge-DNAmAge in separate age groups and time points.

In nonagenarians of V90\_2010 and V90\_2014 populations, CD4<sup>+</sup>CD28<sup>-</sup> cells were the most significant PBMC subtype correlate with the  $\Delta$ -cAge-DNA<sub>m</sub>Age (Figure 14). Interestingly, as shown in Figure 14C, the proportion of CD4<sup>+</sup>CD28<sup>-</sup> cells demonstrated remarkable stability during the follow-up period, and this phenomenon was also observed consistently among other cell subtypes (Tables 8 and 9).



**Figure 14.** CD28-CD4<sup>+</sup> cell counts of the subjects in V90\_2010 and V90\_2014 correlate with their corresponding DNAmAge values and with one another (Study III). These associations are visualized as scatterplots. Each dot corresponds to one subject; the colors of the dots are explained in the graph legends. a) The CD28-CD4<sup>+</sup> proportions of 90-year-old subjects in V90\_2010 correlated with their DNAmAge in 2010 ( $r = 0.540$ ,  $P = 4.75 \times 10^{-10}$ ,  $N = 115$ ). b) The CD28-CD4<sup>+</sup> proportions of 94-year-old subjects in V90\_2014 correlated with their DNAmAge in 2014 ( $r = 0.626$ ,  $P = 1.20 \times 10^{-5}$ ,  $N = 41$ ). c) The CD28-CD4<sup>+</sup> proportions of all participants in V90\_2010 correlated with their CD28-CD4<sup>+</sup> proportions in V90\_2014 (among all:  $r = 0.901$ ,  $P = 9.48 \times 10^{-24}$ ,  $N = 63$ ; among elderly:  $r = 0.864$ ,  $P = 3.33 \times 10^{-16}$ ,  $N = 51$ ; among young controls:  $r = 0.684$ ,  $P = 0.014$ ,  $N = 12$ ). (Study III)

In addition to analysis using simple  $\Delta$ -cAge-DNA<sub>m</sub>Age, the associations between cell subtypes and epigenetic age in Study III were confirmed to be highly similar when using DNA<sub>m</sub>Age residuals adjusted for chronological age instead of plain  $\Delta$ -cAge-DNA<sub>m</sub>Age (data not shown). In the analysis, the residuals were fetched from the regression model, where the Horvath DNA<sub>m</sub>Age was regressed on chronological age.

## 5.5 Methyloomic predictors of mortality (IV)

The Cox univariate assessment highlighted 19 621 and 15 505 CpG sites associated with mortality ( $P < 0.05$ ) in the 2.55-year and 4-year follow-up data of nonagenarians ( $N = 111$ ), respectively. Of those, 19 CpG sites for the 2.55-year follow-up and seven CpG sites for the 4-year follow-up data showed significance at the level of  $FDR < 0.5$ . The Ingenuity Pathway Analysis (IPA) gene networks were generated for the 16 known genes associated with the 19 top-ranking CpG sites at the 2.55-year follow-up (Figure 15). Analogous analysis was also performed with genes harboring the 250 top-ranking CpG sites. In both cases, NF- $\kappa$ B was displayed as the central node, and were enriched for the common term 'Hematological System Development and Function'. The canonical pathway analysis demonstrated a wide spectrum of cellular signaling functions associated with methyloomic predictors. Of those, several were inflammation and immunity-related processes. The IPA analysis with seven or 250 top-ranking predictor CpG sites of the 4-year follow-up data resulted as non-significant. Due to low number mortality-associated CpG sites with  $FDR < 0.5$  in the 4-year follow-up data, further analyzes were not performed with that data.

Correlation analysis between the DNA methylation levels in the 19 mortality-associated CpG site methylation level (the standardized weighted residuals) and the corresponding (i.e. transcript and CpG site in overlapping genomic location) gene expression levels were analyzed using Spearman's rho. The transcripts were selected in the analysis if the gene expression level was above threshold of 5. Thus, the list of genes in the analysis was *ATP5SL*, *FOXP1*, *HIVEP3*, *IQSEC1*, *ITPR3*, *MAP3K14*, *METAP1*, *RGS10*, *RIOK1* and *VOPPI*. This analysis resulted three significant CpG site/transcript pairs. Inverse correlations were observed between the cg03348466 (*CRTC3*) and *CRTC3* mRNA level and between cg04182483 (*RGS10*) and the *RGS10* mRNA level. A positive correlation was observed between cg22794214 (*HIVEP3*) and *HIVEP3* mRNA level.

The Ridge regression selection-procedure revealed that the model with methyloomic markers performed better than models with conventional aging biomarkers (listed in Table 2). In the selection procedure, mortality-predictive accuracy of separate Ridge regression models was compared between: (1) a model containing only the conventional biomarkers, (2) a model containing both the conventional biomarkers, and the 19 mortality-associated CpG sites, and (3) a model containing only the 19 mortality-associated CpG sites. A DNA<sub>m</sub>Age estimate was

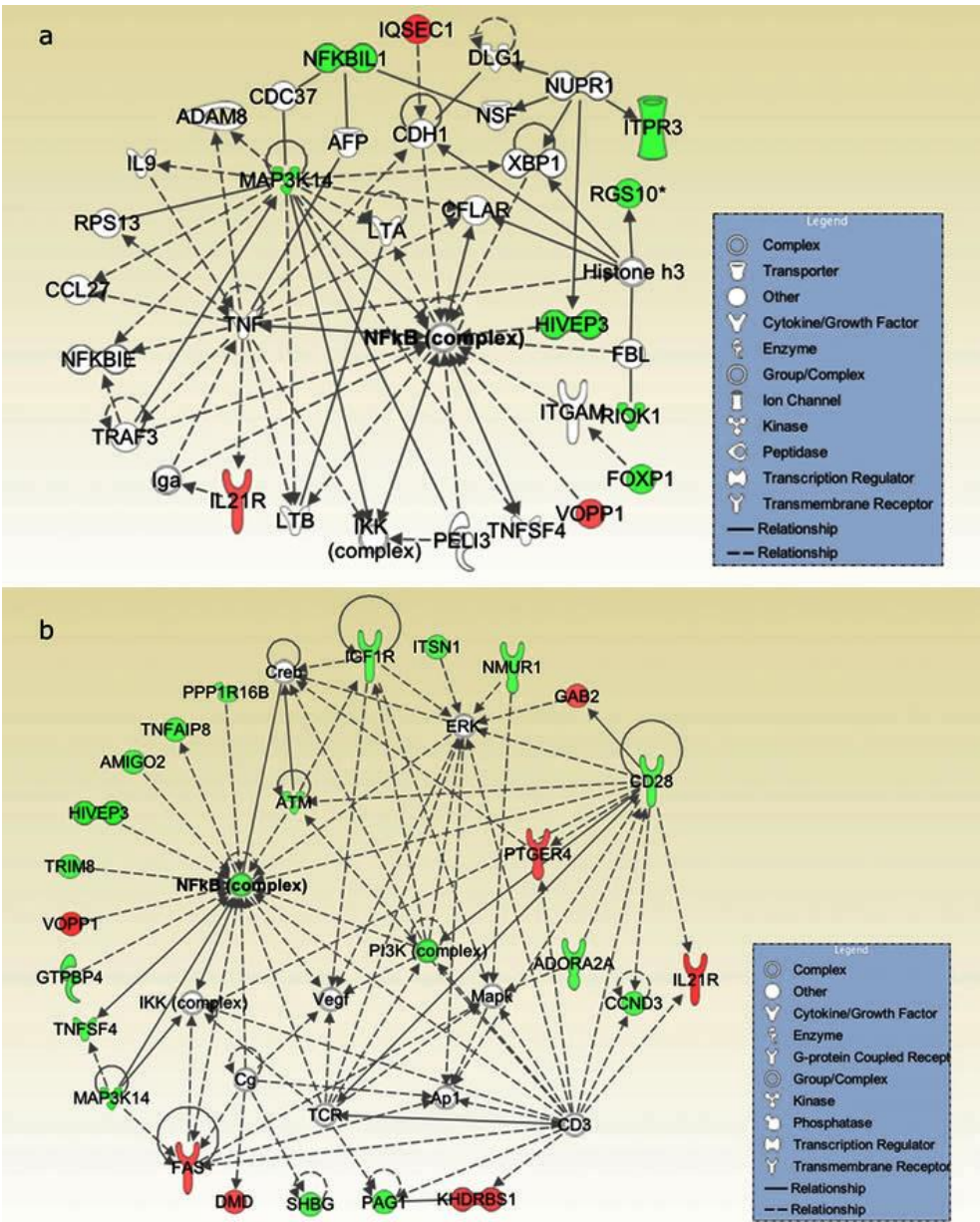
included in the regression models as part of a group of conventional biomarkers. The best prediction model was selected based on goodness of fit descriptive and Akaike Information criterion. The most important methylomic predictors were selected from the 19 CpG sites. The final mortality-predicting signature shown in Table 10 was assessed using a Cox regression model, and the discriminative power (Harrell's C) for this model was 89.9%.

**Table 10.** The final mortality-predicting methylomic signature from the 2.55-year follow-up data (V90\_2010, N=111 with full pattern of variables shown in Table 2) assessed from the Ridge regression-selection procedure and Cox regression analysis. These seven CpG sites of the 19 top-ranking sites provided the best Cox regression model. Symbols: HR=hazard ratio, CI=confidence level interval, S.E. =standard error, Z=Wald statistics. (Study IV)

	HR (95% CI)	S.E.	Z	P
<b>cg08421934 (NA)</b>	0.41 (0.26-0.64)	0.1	-3.84	<b>&lt;0.001</b>
<b>cg15770702 (MAP3K14)</b>	0.40 (0.27-0.61)	0.08	-4.38	<b>&lt;0.001</b>
<b>cg08596308 (ATP6V1G2, NFKB1L1)</b>	0.50 (0.34-0.73)	0.1	-3.6	<b>&lt;0.001</b>
<b>cg23282964 (RIOK1)</b>	0.56 (0.37-0.84)	0.12	-2.82	<b>0.005</b>
<b>cg16720947 (PLEC1)</b>	0.52 (0.34-0.80)	0.13	-2.94	<b>0.003</b>
<b>cg27027151 (IL21R)</b>	2.09 (1.44-3.02)	0.39	3.9	<b>&lt;0.001</b>
<b>cg26843567 (NA)</b>	0.68 (0.46-0.99)	0.13	-2.01	<b>0.045</b>

Finally, it was examined whether the methylomic mortality-predictors are central contributors to the aging-associated epigenetic drift. Therefore, the top-ranking 250 mortality-associated CpG sites were compared to the lists of previously reported aging-associated methylation sites (Marttila et al. 2015a; summary by Steegenga et al. 2014). The examination showed that there were no overlapping sites. Furthermore, the epigenetic age (DNAmAge) showed no association with mortality in Study IV. The epigenetic ages of survivals and non-survivals for 2.55-year follow-up analysis are shown in Table 2. There were no significant differences between 2.55- or 4-year-follow-up survivals and non-survivals in their baseline DNA methylome ages.





© 2000-2015 QIAGEN. All rights reserved.

**Figure 15.** The highest-ranking networks from the genes with (panel a) the top 19 and (panel b) the top 250 mortality-associated CpG sites. These sites were identified in the 2.55-year follow-up (V90\_2010, N=111 with full pattern of variables). The green color of the molecules indicates that hypomethylation of a CpG site in the gene was associated with increased mortality, and the red indicates that hypermethylation of a CpG site in the gene was associated with increased mortality. (Study IV)

## 6 DISCUSSION

### 6.1 The Studies

In this thesis, the objective was to investigate aging-associated changes in DNA methylation landscape (I) and the behavior of the epigenetic clock that is manifested in the human immune cells (III). Methyloomic markers predicting survival in the nonagenarians were explored (IV) and the effect of immune system aging-accelerating marker CMV in epigenetic aging was analyzed (II). Special interest was given to the epigenetic age that is demonstrated with DNA methylome age estimation algorithm (Horvath 2013).

The specific hypotheses for this thesis comprised the following issues. The longitudinal behavior of the epigenetic age (i.e. Horvath's DNAmAge) as well as the blood cell subtype distributions appeared to be indefinite. Discrepancies in the lists of reported aging-associated methylation sites had emerged, thus requiring further evaluations. Another question concerned methyloomic mortality predictors: were there methylation sites that might be used as survival predictors in the elderly (90+ years of age) and might those predictors even overcome conventional biomarkers of aging in the survival analysis? Moreover, the associations between epigenetic aging and biomarkers related to immune system aging (such as CMV or blood cell distribution) was unclear.

### 6.2 Aging-associated DNA methylation changes at single CpG sites

Overall, observations in Study I support the existence of epigenetic drift and the epigenetic clock. The study showed that clocklike-behaving CpG site methylation may be reliably detected from a cross-sectional sample with an age range less than 10 years and relatively small sample size. In the study, aging-associated CpG sites were explored using EWAS approach on YFS\_2011 sample. The blood cell subtype counts in the samples were characterized using methodology based on methylation signatures (Jaffe and Irizarry 2014), and the cell subtype influence on DNA methylation was shown to be significant. The resulting top-rank aging-associated CpG sites are shown in Tables 3 and 4.

The different characteristics of hypomethylation and hypermethylation during aging have been reported (Zampieri et al. 2015), and in Study I, this difference was seen in GO term enrichment analysis, in Pscan transcription binding partner analysis (Zambelli et al. 2009), and in numbers of association hits per gene. Even though active demethylation pathways exist (Tahiliani et al. 2009; Kohli and Zhang 2013; Rasmussen and Helin 2016), aging-accelerated global hypomethylation seems to largely resemble random erosion (Zampieri et al. 2015), and appears to be influenced

by many factors including aging-associated deficiency of methyl group donors in the one-carbon cycle (Guéant 2016). Hypermethylation is a more actively guided process in nature, as the addition of methylation group itself is always an enzymatic reaction (Riggs 2002; Rasmussen and Helin 2016). In Study I, there was a clear enrichment bias of the GO function terms and process terms within genes with aging-associated CpG sites. The terms were enriched to genes with hypermethylated CpG sites, exclusively, while hypomethylated genes showed no enrichment. The most statistically significant processes enriched to genes with hypermethylated CpG sites were ‘anatomical structure development’ and ‘morphogenesis’, both of which cluster under the term ‘developmental process’. The enrichment of hypermethylated aging-associated CpG sites to these processes has been previously reported (Rakyan et al. 2010; Hernandez et al. 2011; Johansson et al. 2013; Florath et al. 2014; Marttila et al. 2015). In addition, other studies have demonstrated (Reynolds et al. 2014; Yuan et al. 2015) that the CpG sites hypermethylated during aging are enriched to common processes and exhibit shared features, whereas hypomethylated CpG sites are a less homogenous group. Furthermore, aging-associated hypermethylation interactome hotspots have been reported (West et al. 2013).

Regarding the numbers of aging-associated CpG sites per one gene, aging-associated hypermethylation was more frequent within genes with more hits of aging-association (Figure 7). Moreover, the Pscan analysis demonstrated that genes with hypermethylation-associated CpG sites were predicted to be regulated by several common transcription factors, whereas there were no common transcription factors found for genes with hypomethylated CpG sites. Thus, with these findings in Study I, the differing characteristics of aging-associated hypermethylation and hypomethylation in sites covered by 450Beadchip array are underlined even further.

## 6.2.1 Comparison analysis of different studies

The most commonly replicated aging-associated DNA methylation sites were explored in Study I. In the exploration, the resulting lists of aging-associated CpG sites in Study I and other studies (Garagnani et al. 2012; Hannum et al. 2013; Florath et al. 2014; Marttila et al. 2015a) were compared. All of the studies in the comparison had been using similar array technology (Illumina 450Beadchip) and heterogeneous blood specimens. Specifically, the reproducibility of these aging-associations was explored in two ways: (option 1) using studies with virtually identical bioinformatics analysis pipelines with blood cell subtype count considerations, and (option 2) using studies with varying analysis pipelines without cell subtype count considerations.

Comparison analysis of option 2 revealed that the number of the same aging-associated CpG sites reported by multiple studies was substantially low (Figure 9). This discrepancy in results is an acknowledged paradigm. For instance, a study by Dozmorov (2015), where comparison analysis highly similar to Study I was performed, reached the same conclusion. The author reported that of the seven

studies in comparison, only one aging-associated methylation region was identified uniformly, and 335 overlapping regions out of 16 854 regions were reported by three studies (Dozmorov 2015).

Various environmental and genetic factors modify the aging-associated epigenetic drift (van Dongen et al. 2016). Factors that are uncontrolled in the methylomic analyses may explain the differences between results from separate EWASs focusing on aging. In particular, the heterogeneity of the blood samples has shown its importance. A study by Gervin et al. (2016) supports this conclusion in an analysis using longitudinal setting with 2-, 10- and 16-year-old subjects and reduced representation bisulfite sequencing. The sequencing method captured 635899 CpG sites. The authors concluded that the effects of aging seen in methylomic landscape in blood cells are largely mediated through variation in cell subtypes. Less than 10% of the 346 aging-associated CpG sites demonstrated association with age directly. (Gervin et al. 2016) Earlier findings have also demonstrated that caution is required when interpreting results from whole blood samples without cell proportion considerations (Reinius et al. 2012; Jaffe and Irizarry 2014).

Nevertheless, in spite of the discrepancies at single CpG site resolution, aging-associated methylation changes reported by different studies demonstrate many universal characteristics (e.g. Polycomb Repressive Complex 2 signature (Dozmorov 2015) and enrichment bias of hypermethylation and hypometylation (Zampieri et al. 2015). The most acknowledged and repeatedly aging-associated single CpG site is cg16867657, which is located in promoter region in gene *ELOVL2*. Hypermethylation of cg16867657 has been reported as aging-associated in multiple tissues and populations with different surrounding environments (populations comprising, for example, European, Hispanic, Arab, and African hunter-gatherers). (Garagnani et al. 2012; Gopalan et al. 2017) cg16867657 association with age was also seen in Study I (Table 3). Furthermore, the DNA methylation levels of *Elovl2* have been shown to change in several tissues of mice in a directionally consistent manner when compared to humans (Spiers et al. 2016). In spite of the solid evidence of the aging-association, the functional role of *ELOVL2* gene in the aging process is unclear. A study that focused in *ELOVL2* methylation in details (Bacalini et al. 2017), has suggested that hypermethylation level of *ELOVL2* is a descriptive of number of replication events in whole-blood cells. They observed using EpiTYPER assay that CMV-serostatuspositive nonagenarians exhibit higher methylation levels in a specific locus in the gene. However, currently, the link between fundamental aging-related processes, *ELOVL2* and its methylation is missing. *ELOVL2* is a gene that encodes polyunsaturated fatty acid elongase. The enzyme operates in condensation reaction during the very long fatty acid elongation cycle in the endoplasmic reticulum (Leonard et al. 2002; Jakobsson et al. 2006). Noteworthy, while DNA methylation of *ELOVL2* clearly increases during aging, studies analyzing aging-associated transcriptomic changes in human populations have not observed differential expression of the gene (Steegenga et al. 2014; Marttila et al. 2015).

In summary, regarding comparison analysis, Study I and study by Marttila et al. (2015a) were compared in option 1, and in both of these, the cell subtype heterogeneity was controlled and similar analysis pipelines were used. When *option 1* (i.e. results from studies with virtually identical bioinformatics analysis pipelines with blood cell subtype count considerations) was compared to *option 2* (i.e. results from studies with varying analysis pipelines without blood cell subtype count considerations), there were multiple times higher number of overlapping aging-associated CpG sites in option 1 (as shown in Figure 8 and Figure 9). Moreover, 987 CpG sites shown in Figure 8 demonstrated clocklike behavior through entire adulthood. These sites appear to change their methylation levels linearly as a function of growing chronological age. This is assumed to be because the sites were observed as aging-associated in a cross-sectional sample of middle-aged individuals with narrow chronological age range (9 years; Study I) as well as in a sample of two age groups having substantially wider chronological age range (~60 years; Marttila et al. 2015a). Thus, these 987 CpG sites exemplify the paradigm of epigenetic clock.

The results in Study I underline that the blood cell subtype consideration is a mandatory practice for replicative EWASs in blood. Results in this thesis demonstrated that the number of overlapping association hits increases when the analysis is adjusted for cell heterogeneity. In other words, in order to detect association between aging and DNA methylation only, cell heterogeneity may interfere this analysis. This is because cell composition is remodeled during aging, and specific DNA methylation signatures are so tightly connected to different cell subpopulation types.

As another conclusion, Study I suggests that the aging-associated 987 CpG sites might be discovered as universally clocklike-behaving sites in the blood in almost any human adult cohort with various age ranges when using appropriate analysis techniques. Future replication studies may determine whether this is true.

### 6.3 Methyloomic mortality predictors

Analogous to Study I, Study IV was performed at single CpG site resolution meaning that associations between mortality and DNA methylation levels were analyzed in each CpG site covered by 450Beadchip array. The methyloomic all-cause mortality-prediction analysis was performed genome-wide in nonagenarians (V90\_2010) with the blood sample heterogeneity consideration. As a result, the top-ranking mortality-associated CpG sites were identified. These sites were different from the established aging-associated DNA methylation sites, and performed better in mortality prediction than conventional aging biomarkers. The addition of epigenetic age to the mortality analysis did not enhance the predictive capacity of the regression model. Moreover, the methyloomic predictor signature established in this thesis supports the genomic-level role of NF- $\kappa$ B at the very end of the human lifespan.

The methylomic mortality prediction signature of blood specimens has the potential to serve as a molecular biomarker for the detection of health risk factors and increased mortality risk. Because the epigenetic landscape is reversible in nature and changes due to environmental input, population-based intervention studies using methylomic mortality risk scores might provide new information for the basis of health guidance.

In Study IV, the performance of conventional aging biomarkers in mortality prediction was compared to the methylomic ones. The final mortality-predicting model in Study IV built up from the top-ranking 19 mortality-associated CpG sites at the 2.55-year follow-up alone showed the best predictive accuracy when compared to other models containing either the conventional aging biomarkers (listed in Table 2) only, or conventional aging biomarkers (Table 2) in combination with the CpG sites. Even though the biomarkers presented in Table 2 are acknowledged measures of aging rate and commonly associated with altered mortality risk, the results here imply that the methylomic signature may be an even more sensitive marker of increased risk of death than the conventional biomarkers. However, the magnitude of change in these mortality-predicting CpG sites is unknown. Without DNA methylation levels measured from multiple time points before deceasing, it remains as being only a hypothetical suggestion that changes in DNA methylation levels may indicate imminent death. Furthermore, the effect on gene expression mediated by these methylomic mortality predictors appears to be complicated and warrant further studies. In Study IV, the mRNA levels of transcripts that are situated in overlapping locations with mortality-predicting CpG sites showed only few moderate signs of being associated with DNA methylation levels.

The top-ranking mortality predicting DNA methylation signatures established in nonagenarians in Study IV turned out to be different from other studies (Moore et al. 2016; Zhang, Wilson et al. 2017) that have been performing corresponding methylomic mortality analyses genome-wide in population-based samples. However, some of the findings in Study IV are consistent with a study by Zhang et al. (2017b) in which discovery, validation and replication cohorts spanning the age range of 31-82 years were utilized. The authors found 58 CpG sites that are all-cause mortality-associated, and these findings were also independent from the DNA methylome age estimate and independent from CpG sites on the basis in the DNA methylome age calculation. In a similar manner to Study IV, Zhang et al. (2017b) comprised a top-hit panel of the best mortality-predicting CpG sites. Furthermore, the authors used the panel in mortality risk-score calculation. Risk-score approach was verified in an independent replication cohort, and the score also predicted cancer, cardiovascular disease and all-cause-mortalities in those samples.

The top-ranking 58 sites by Zhang et al. (2017b) mapped to 38 genes that are linked to various aging-associated diseases including cancers and cardiovascular diseases. Tobacco smoking was an important covariate as it was associated with 48 of the 58 mortality-CpGs. The well-known smoking-site (cg03636183) located in gene *F2RL3* (Breitling et al. 2011; Zhang et al. 2014) was also among these mortality-

predicting CpGs. Regarding this finding, it is interesting that tobacco smoking is an important mortality risk increasing lifestyle factor. It is also notable that previous research has indicated that the hypomethylation of the cg03636183-smoking-site might be reversed in the following decades after smoking has ceased (Zhang et al. 2014), and this underlines the importance of findings by Zhang et al. (2017b). The data in this thesis have pointed to smoking-associated changes in cg03636183 located in gene *F2RL3*: the quality control exploration of Study I showed that tobacco smokers in the YFS\_2011 sample demonstrated hypomethylation of cg03636183.

The genes with mortality-associated CpGs clustered functionally around the nuclear factor kappa B (NF- $\kappa$ B) complex (Figure 15). The genes, nuclear factor of kappa light polypeptide gene enhancer in B-cells 1 (*NFKB1*) and ataxia telangiectasia mutated (*ATM*) were also identified in the network. Consequently, these results demonstrate the role of NF- $\kappa$ B complex in human longevity and support previous findings of its lifespan-regulating role. Specifically, the NF- $\kappa$ B complex is associated with accelerated aging and cellular senescence in studies with mouse models (Osorio et al. 2012; Tilstra et al. 2012; Jurk et al. 2014). The studies implied that disrupted NF- $\kappa$ B activation influenced by *NFKB1* and *ATM* regulation and the following chronic systemic inflammatory state is driving the senescence and aging-associated decline.

Moore et al. (2016) also performed genome-wide methylomic mortality-association analysis using a population-based sample with the age range of 30-100 years. The participants were followed for, on average, 4.4 years. The study reported 88 mortality-associated CpG sites and 76 CpG sites where the rate of methylation level change was associated with mortality. The individual mortality-associated CpG sites in the study did not overlap with those of Study IV, but the study was in accordance with Study IV in other aspects. The mortality-associated CpG sites were largely different from the most acknowledged aging-associated sites. Interestingly, mortality-associated CpG sites by Moore et al. (2016) also comprised genes with immunoinflammatory functions and a link to NF- $\kappa$ B regulation. (Moore et al. 2016)

In conclusion, the results (Study IV, Moore et al. 2016) suggest that the genomic factors controlling mortality operate through mechanism(s) that involve(s) an inflammatory component. However, it should be underlined that, instead of complete mechanistic link(s), Study IV and other such analyses provide only associations between mortality and DNA methylation changes in the predictor CpG sites.

The age range appears to be the main difference in the study design in Study IV (nonagenarians) and data by Zhang et al. (2017b; all participants aged 32-82 years) and Moore et al. (2016; participants aged 30-100 years). The individual top-ranking mortality-predicting CpG sites in the Study IV were different from those reported by Zhang et al. (2017b) and Moore et al. (2016). It may be speculated that the outcome differences between studies using participants with differing ages are equally plausible, as genetics appears to have a greater impact on the survival rate at ages above 60 years (Jylhava et al. 2010). In general, the biological factors influencing

mortality risk at a younger age (below 80 years) appear to be different from those in the oldest group. A previous population-based epidemiological study with 19 430 participants has demonstrated that chronic diseases in the eldest group (aged 90-99 years) are less powerful all-cause mortality predictors when compared to younger adults (aged 50-59 years). (Lee et al. 2008) In this light, the mortality predictors are likely to differ between age categories, and the methylomic predictor model presented in Study IV (Table 10), may be an old-age mortality-predicting signature that is unsuitable for populations with other age ranges.

On the other hand, analysis protocols used in many studies (e.g. Study IV and Zhang et al. (2017b)) have been utilizing Lasso or Ridge regression that are designed to capture the most predictive variables among a large set of variables demonstrating high dimensionality and multicollinearity. For example, principal component analysis similarly aims at the dimensionality shrinkage. Consequently, it is less likely to capture precisely identical CpG sites to the final and statistically fine-tuned prediction signatures. Preferably, the overlapping sites might be found through list comparison with the preliminary lists of mortality-associated sites. Often, the preliminary association lists are unpublished.

## 6.4 Longitudinal analysis of the epigenetic age

In Study III, the behavior of human epigenetic clock, assessed as the DNAmAge estimate and DNAmAge acceleration (the difference between chronological age and DNAmAge estimate;  $\Delta$ -cAge-DNAmAge), was characterized longitudinally using two independent populations representing early adulthood, middle-age and advanced ages. As a main finding, longitudinal data provided evidence that the difference between chronological and epigenetic age is surprisingly stable over several years or even decades of human life, and when accompanied with previous reports, it may be concluded that the main trajectory of the blood DNA methylome aging rate is largely set before adulthood.

In line with previous research (Christiansen et al. 2016; Armstrong et al. 2017), the estimated epigenetic ages in Study III were at lower level than the actual chronological ages of the nonagenarians ( $90 \pm 0$  and  $94 \pm 0$  years of age, Table 5). Whether this issue is a DNAmAge algorithm-based underestimation problem or a true phenomenon remains to be shown. However, it may be hypothesized that this observation is an indication of a higher survival rate at earlier adulthood due to overall younger epigenetic age. DNAmAge algorithm-based underestimation may be another explanation, because there were considerably low number of individuals with  $>90$  years of age in the original epigenetic age-predictor training analysis (Horvath 2013). Due to insufficient algorithm training, the epigenetic age estimates of nonagenarians may rely on more unreliable assumptions than other age groups with better coverage in the training analysis.

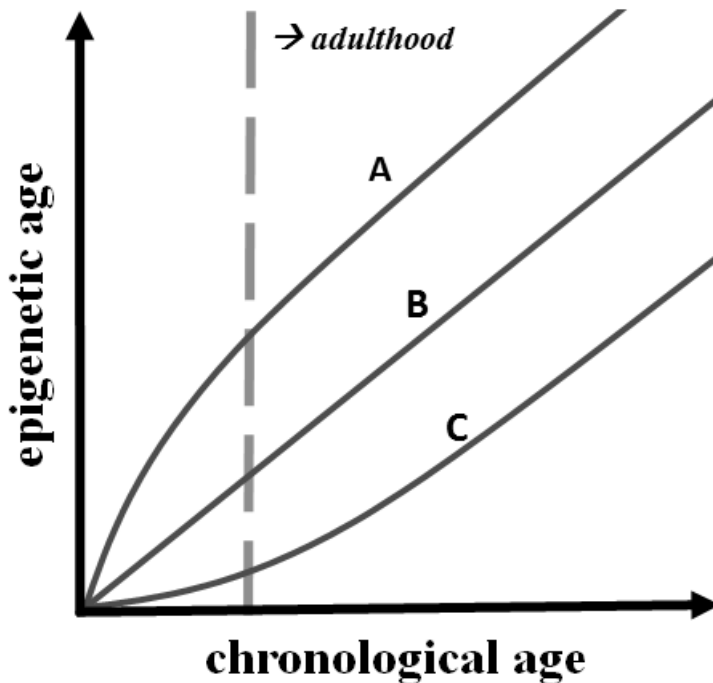


However, the results in Study III demonstrated, in both samples (V90 and YFS), with varying age ranges and follow-up periods, that the difference between epigenetic and chronological ages was largely unchanged during the follow-ups (Figures 11A and 11B). The analysis in Study III pointed that, at the age of 15-24 years, the level of epigenetic aging was already on an individual-specific level, in such way that over the 25-year follow-up period, very little deviation from this line was observed. A clear association in the scatterplot in Figure 11A demonstrates this finding. In the elderly and in the young adults, the DNA methylome ages were similarly fixed for four years (Figure 11B).

Several environmental factors, such as diet, physical characteristics and toxin exposure, may influence the ticking rate of the epigenetic clock (Zampieri et al. 2015; Simpkin et al. 2016; Nevalainen et al. 2017; Quach et al. 2017). Consequently, it is possible that the shifts from the epigenetic age trajectories shown in Figure 11 are caused by environmental factors. The shift is demonstrated in the observation that some of the  $\Delta$ -cAge-DNAme values diverge from the diagonal lines in the scatterplots in Figure 11A and 11B. Furthermore, laboratorial technical issues of the DNAme measurement may contribute to the scattering. Even though numerous factors may have an impact on the DNAme estimate, astonishingly high correlation coefficients between within-subject  $\Delta$ -cAge-DNAme values at the follow-ups were exhibited in both study populations in Figure 11A and 11B.

Other epigenetic studies with cross-sectional data have shown similar signs of the epigenetic age behavior as seen in Study III. Horvath (2013) first reported in his study using cross-sectional DNAme data that the epigenetic clock-ticking rate is accelerated before adulthood and thereafter the rate slows down (Horvath 2013). Moreover, the DNA methylomic profiles of twin pairs are highly similar (77% similarity) at birth and begin to deviate from then on, and by the late adulthood, the DNA methylomic profiles of twins display only 39% similarity (at median age of 63 years, all women) (Horvath 2013).

Above all, a recent follow-up study (Simpkin et al. 2016) on children (baseline at birth, follow-up time points at 7 and 17 years of age) demonstrated that the difference between chronological and epigenetic age in the children is smaller on average at the time of birth when compared to later stages of childhood. The deviation in epigenetic age acceleration on those children increased in association with chronological age. The authors observed several factors that are related to epigenetic age acceleration in children, including *in utero* exposure to alcohol consumption and smoking. Importantly, Simpkin et al. (2016) concluded that within-subject correlation between epigenetic ages at different time points increases with increasing chronological age; in other words, at the time of birth and during childhood, the children exhibited lower correlations of epigenetic ages between two sequential time points than during later periods of life. Consequently, these results are consistent with the conclusion that the primary trajectory of epigenetic age (i.e. difference between epigenetic and chronological ages) is fixed before adulthood.



**Figure 16.** A hypothesis for development of epigenetic age trajectory during human lifespan based on findings in Study IV and Simpkin et al. (2016). Negative or positive  $\Delta$ -cAge-DNAme value correspond to old or young DNA methylome age, respectively. Before adulthood, the epigenetic aging is either decelerated (A) or accelerated (C) when compared to the situation where  $\Delta$ -cAge-DNAme is 0 (B). In this hypothesis, the trajectory i.e.  $\Delta$ -cAge-DNAme values change less during adulthood.

Early years during the human lifespan are extremely important determinants of future health span. Previous findings from the acknowledged Dunedin study in New Zealand clearly demonstrate this issue. The study comprises full birth cohort including 1037 individuals and has been followed with unique intensity (participation rate >90%) for 40 years starting from the time of birth. (Poulton et al. 2015) Moreover, specific biomarkers, for example, telomere length (Kananen et al. 2010; Naess and Kirkengen 2015) and the epigenetic data from Study III and by Simpkin et al. (2016) have now shown support for this early-years-development-paradigm.

As a conclusion, the rate of the epigenetic age acceleration varies to some extent but less in adulthood, even at the old ages, and the main difference between epigenetic age and chronological age is set to a certain general level before adulthood, probably during childhood (as shown in Figure 16). Overall, the model proposed in Figure 16 is thought provoking, because older epigenetic age is also associated with increased mortality in adulthood (B. H. Chen et al. 2016; Zhang et al. 2017). Therefore, epigenetic changes in the beginning of life gain an even more crucial meaning for the entire human lifespan.

## 6.5 Cytomegalovirus infection

In order to obtain a perspective on another aspect of immune system aging, the association of CMV<sup>±</sup> infection status with epigenetic aging was investigated in Study II. Previously, human CMV infection has been connected in epidemiological studies to aging-related conditions such as frailty and functional impairment (Wang et al. 2010; Moro-Garcia et al. 2012) as well as all-cause and cardiovascular mortality (Strandberg et al. 2009; Roberts et al. 2010; Gkrania-Klotsas et al. 2013). Furthermore, CMV infection in humans seems to drive notable changes in the T cell repertoire and functioning, and thus the infection is a possible contributor to immunosenescence (Pawelec and Derhovanessian 2011; Tu and Rao 2016).

The analysis in Study II was performed in separate chronological age categories, because it was suggested that nonagenarians and young controls probably differ in their CMV infection timeframes; the nonagenarians may have had the acute CMV infection decades ago and therefore the latent virus has had more time to facilitate inflammaging and immunosenescence. Interestingly, in both categories, the CMV<sup>+</sup> subjects turned out to be older by their epigenetic age than the CMV<sup>-</sup> subjects (Figure 12). As such, these findings support the biological significance of the DNAmAge estimate.

## 6.6 The effect of blood cell subtypes

The data in this thesis indicate that the aging-associated changes in the cell composition appear to be relatively small during many stages of lifespan. Nevertheless, based on the data, the aging-associated shift in the cell composition by the time of 90 years of age is also apparent. Furthermore, the data suggest that the epigenetic age is associated with the varying blood cell subtype distributions. In practice, these blood cell composition-related issues generate complexity for the interpretations of results from blood cell-based experiments.

In line with prior knowledge (Weiskopf et al. 2009; Pawelec et al. 2010; Tu and Rao 2016; B. H. Chen et al. 2016) the cell count data of V90 population showed that the cell composition is remodeled in the elderly (as shown in Table 5). The aging-associated shift was explored cross-sectionally in both time points (in 2010 and 2014), and the change in cell composition was seen between young controls and nonagenarians, whose age difference was close to 60 years. For instance, the proportions of CD28-CD4<sup>+</sup>, CD28-CD8<sup>+</sup> and CD14<sup>+</sup> cells differed greatly between these groups (Table 5).

A study by Chen et al. (2016) has underlined the acknowledged (Pawelec 2017) aging-associated shift in blood cell composition. The DNA methylomic profiles may be utilized to estimate cell subtype counts in blood specimens (Houseman et al. 2012; Jaffe and Irizarry 2014), and the aging-associated shift was demonstrated using this estimation technology. The estimates of blood cell subtype counts were clearly

associated with chronological age in a mortality prediction meta-analysis of 13 cohorts comprising over 13 000 participants. The study showed that, when compared to other cell subtypes, the quantities of CD8<sup>+</sup> naïve cells were decreased and CD8<sup>+</sup>CD28<sup>-</sup>CD45RA<sup>-</sup> cells were increased the most by age. That is to say, in the meta-analysis, the highest reported correlation coefficient was 0.48 for CD8<sup>+</sup>CD28<sup>-</sup>CD45RA<sup>-</sup> cells and the lowest was -0.53 for CD8<sup>+</sup> naïve cells. (B. H. Chen et al. 2016) In V90\_2010 data (Table 5), there are also clearly differing proportions of CD8<sup>+</sup> naïve cells between young controls and nonagenarians. Interestingly, the utilization of blood cell count information to the methylomic analysis improved the mortality prediction capability in the initial analysis of the report by B. H. Chen et al. (2016).

In addition to the aging-associated shift discussed above, the longitudinal data of blood cell composition provided evidence that within-subject changes in blood cell subtype proportions are relatively small in young adulthood, in middle age, as well as in the advanced age for several years or even decades. This is demonstrated in significant intra-individual correlations of the blood cell subtype distributions between sequential time points (Table 8 and 9, Figure 14C). In other words, in Study III, young adults and nonagenarians were followed for 4 or 25 years, and the results showed that within-subject changes in the immune cell landscape were relatively small during those years. As shown in Figure 14C and Table 8, in V90 population (both young adults and nonagenarians), the CD28<sup>-</sup>CD4<sup>+</sup> proportions in blood samples collected in 2010 were highly similar to those CD28<sup>-</sup>CD4<sup>+</sup> proportions in blood samples which were collected 4 years later. Other blood cell types showed similarly noticeable within-subject correlations in both V90 and YFS populations (Tables 8 and 9). For instance, the B cell and monocyte counts as well as the CD4<sup>+</sup>:CD8<sup>+</sup> cell count ratio measured from the blood samples of middle-aged participants in YFS\_2011 sample correlated clearly ( $r \geq 0.5$ ) with their corresponding cell counts in the YFS\_1986 sample collected 25 years earlier. These results suggest that the major shifts in the blood cell composition that are seen in Table 5 might take place somewhere after middle age and before advanced ages. However, the age groups in the study populations in this thesis comprised young and middle-aged adults as well as nonagenarians, while children and the age group between middle age and advanced ages were lacking. Accordingly, it remains to be experimentally verified when exactly the major shift in the blood cell composition occurs.

In Study III, the epigenetic age was associated with the varying blood cell subtype distributions (Table 6 and 7). Specifically, CD4<sup>+</sup>CD28<sup>-</sup> and CD8<sup>+</sup>CD28<sup>-</sup> cells were the most clear blood cell subtype correlates with the  $\Delta$ -cAge-DNA<sub>m</sub>Age. Overall, the blood cell subtype that correlated with  $\Delta$ -cAge-DNA<sub>m</sub>Age the most was CD4<sup>+</sup>CD28<sup>-</sup> cells; an increased frequency of these cells was associated with older DNA methylome age in the nonagenarians at both time points (Table 6; Figure 14). Moreover, CD4<sup>+</sup>CD28<sup>-</sup> cells were similarly associated with DNA<sub>m</sub>Age in the young controls (Table 6; Figure 14).

The observation that increased proportions of immunosenescence-demonstrating cell subtypes (CD4<sup>+</sup>CD28<sup>-</sup> and CD8<sup>+</sup>CD28<sup>-</sup> cells) and a reduced CD4<sup>+</sup>:CD8<sup>+</sup> cell ratio emerged as highly significant correlates of older DNAmAge is of specific importance to aging and age-related conditions. However, the association between DNAmAge and immune cell distribution is not limited to aging alone. A study by Horvath and Levine (2015) demonstrated that accelerated epigenetic aging in HIV-1 patients is accompanied by higher frequencies of NK and CD8<sup>+</sup>CD28<sup>-</sup>CD34RA<sup>-</sup> T cells and decreased frequencies of granulocytes, naïve (CD4<sup>+</sup> and CD8<sup>+</sup>) T cells. The study was performed using six DNA methylation datasets comprising brain and blood samples. The average increase in epigenetic age between cases and controls was 7.4 years in the brain tissue and 5.2 in the blood. (Horvath and Levine 2015) It is worth noting that prematurely developed aging-related conditions are more common in individuals with HIV-1 infection (Pathai et al. 2014).

While experimental evidence is limited, Horvath and Levine (2015) have already proposed a few models that might underlie the association between blood cell subtype proportions and epigenetic age acceleration (Horvath and Levine 2015). Specifically, the authors conducted these models regarding HIV-1 infected individuals but these suggestions might be generalized to a wider perspective to concern CMV infection and even aging in general. According to the authors, in one of the most promising suggestions, the burden of the virus infection on the epigenetic age acceleration is mediated through the increased quantities of senescent or exhausted T cells (e.g. HIV-1 infection → increased proportion of CD28<sup>-</sup>CD45RA<sup>-</sup> CD8<sup>+</sup> T cells in blood → higher epigenetic age measured in blood). However, the authors find weaknesses in this model, as it may not be generalized to the brain tissue, which is isolated by the blood-brain barrier, and epigenetic aging is seen in the brain at the same time. Another model by Horvath and Levine (2015) suggests that the causal relationship between the three components is missing; the quantities of exhausted T cells are increased due to virus infection and epigenetic age are independently changed as a result of virus infection. Moreover, the authors propose that the viral integration into the host genome might weaken the genomic stability, which might be further coupled with epigenomic instability, and thus the changes in the DNA methylome age are plausible. Faults in the DNA methylation maintenance machine might arise and this could be seen as epigenetic age acceleration.

In Study II, the proportion of CD4<sup>+</sup>CD28<sup>-</sup> cells correlated more strongly with the epigenetic age acceleration than the CMV serostatus. The suggested models by Horvath and Levine (2015) may also apply to this finding. Either the CMV serostatus might accelerate epigenetic aging seen in the blood cells by exhausting the immune cell pool, or these two events exist in parallel without no causality in between. However, because altered CD4<sup>+</sup>CD28<sup>-</sup> and CD8<sup>+</sup>CD28<sup>-</sup> cell proportions are so tightly linked with cytomegalovirus infection (Tu and Rao 2016), it may be difficult to differentiate and solve these issues experimentally.

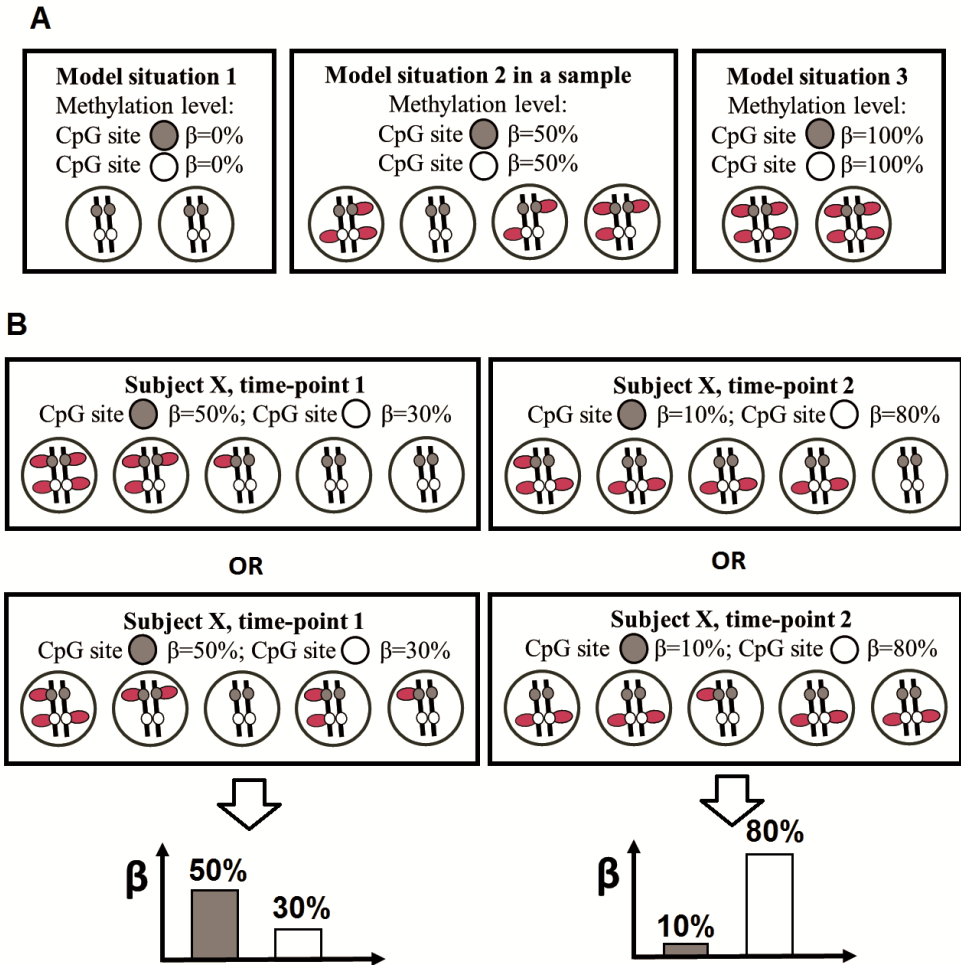
In spite of the limited knowledge of causalities, in practice, studies (Study I-III; Reinius et al. 2012; Jaffe and Irizarry 2014; Horvath and Levine 2015; Marttila et al. 2015a) have shown that the blood cell subtype proportions are potential confounding factors in EWASs or analysis focusing on epigenetic age estimate. Moreover, as demonstrated in this thesis and elsewhere (Marttila et al. 2015), CD4<sup>+</sup>CD28<sup>-</sup> cells appear to be a specifically important confounder in methylomic aging-association studies in the blood. The overall variation in genome-wide DNA methylation profile has been shown to correlate greatly with the proportion of CD4<sup>+</sup>CD28<sup>-</sup> in PBMCs (Marttila et al. 2015a). This implies that these cells have crucially different and unique methylomic landscapes when compared to other cell types within PBMCs. The mechanistic basis for the role of CD4<sup>+</sup>CD28<sup>-</sup> cells in immune system aging is not fully understood and further studies are needed. However, it may be speculated that alterations in aging-associated changes in DNA methylation profiles as well as in the epigenetic clock sites may be connected to certain central features of these cells, i.e. their cytokine secretion profile, shortened telomeres and resistance to apoptosis (Weiskopf et al. 2009; Arnold et al. 2011). In addition to aging, the expansion of these cells have been reported in certain immune-related diseases, such as multiple sclerosis, rheumatoid arthritis and acute coronary syndromes (Broux et al. 2012). Therefore, observations regarding blood cell landscape in this thesis may also have implications on the methylomic analyses of those disorders.

## 6.7 $\beta$ value vs DNA methylation status

The varying DNA methylation levels ( $\beta$  values or methylation level percentages) are the output measure used in the EWASs. It is less acknowledged in the literature that the basis of the varying  $\beta$  values is highly complex paradigm, as illustrated in a schematic presentation in Figure 17, and the problematics concern epigenetic studies as a whole. In practice,  $\beta$  value of a biological sample with numerous cells may be a result of various situations. The output value of a single CpG site from the Illumina 450Beadchip analysis is the continuous methylation value  $\beta$ , which ranges between 0% and 100%.

Value  $\beta$  is the sample-specific average of all methyl groups attached to certain DNA methylation site in the cells. In a sample of only one cell where the two DNA double strands with identical CpG sites are located, there might be three different situations. The sites may be either completely unmethylated corresponding to the measured average methylation level of 0%, only the other CpG site in the other DNA double strand is methylated corresponding to 50%, or both of the double strands are methylated corresponding to the average of 100%. It is yet to be shown how common a situation the average of 50% is in a single cell, and what the functional consequences for that cell are. Correspondingly, from a CpG site in two cells, the measured average of  $\beta$  value may be either 0%, 25%, 50%, 75% or 100%.

Based on this principle (shown in Figure 17), after increasing the number of cells to high number, the average methylation level from a single CpG site may be in a linear range between 0% and 100%.



**Figure 17.** Models describing the basis of the varying  $\beta$  value. Continuous methylation value  $\beta$  is the output from the Illumina 450Beadchip analysis, and the value is ranging between 0% and 100%.  $\beta$  value is the average of all methylation events in a typical sample with a high number of cells. In theory, there are numerous possibilities for how the  $\beta$  value is initially formed. In this schematic presentation, the two DNA double strands (black lines) are shown inside each cell (large circles). A) Here, in each DNA double strand, there are two different CpG sites (in grey and white color). The model situations 1-3 illustrate scenarios when the methyl groups (in purple) are either at the same time present (100%), partially present (50%) or absent (0%). B) Bottom panels illustrate longitudinal setting (time point 1, left; time point 2, right) where during the follow-up methylation  $\beta$  values are changed from 50% to 10% (grey CpG site) and from 30% to 80% (white CpG site).

At this moment, it is unknown, for example, how DNA methylation changes by age are established in a pool of a large number of somatic cells in such a manner that the gradually changing DNA methylation pattern of 353 CpG sites in DNAmAge estimator (Horvath 2013) reflects the chronological age (or even biological age?) of an individual. As observed in Study III and elsewhere (Horvath 2013; Zampieri et al. 2015; Jylhava et al. 2017), the DNAmAge estimate correlates with chronological age remarkably well (correlation coefficient  $> 0.7$ ), even though the route from single methylation marker states to the final epigenetic age estimate is very complex. In conclusion, regarding average DNA methylation levels measured from a tissue sample with numerous cells, the cell population contributing to the average may potentially be largely heterogenic. Consequently, it is even more fascinating that the epigenetic age estimation is predicting the chronological age of an individual from multiple human tissues.

## 6.8 Limitations

The data in this thesis add knowledge regarding epigenetic aging, immune aging and survival in human populations. In particular, the longitudinal sample of the nonagenarians (V90) with the combination of DNA methylome and FACS analyses provides unique information. However, this thesis has some limitations. Studies II and IV lacked independent validation cohorts. Moreover, the age groups in the study populations comprised young and middle-aged adults as well as nonagenarians, but children and the age group between middle age and advanced ages were lacking. The number of follow-up time points and sample sizes might have been higher. The micro array based methodology used here is enriched with certain genomic locations, and thus it is not representative of the complete human genome. That is, the arrays are biased when compared to full genome scanning methods.

Regarding blood specimens used in this thesis, instead of heterogeneous blood cells, purified cell subtypes are more suitable material. This issue was controlled with characterizations of the blood cell subtype counts. However, the cell count data were generally lacking detailed information of minor cell subtypes. The flow cytometry data comprised minor cell subtype proportions from CD4<sup>+</sup>CD28<sup>-</sup> and CD8<sup>+</sup>CD28<sup>-</sup> cells alone; and FACS-characterization with more detailed cell subtype classifications is advantageous.

In Study III, the data do not provide explanations why the proportions of the specific blood cell types are so strongly associated with the DNAmAge estimate (Figure 14, Tables 6 and 7), but the fact that the blood cell composition changes with age is also apparent in these results (Table 5). The results do not demonstrate any causal relationships between the aging-associated changes in cell composition and the actual DNA methylation changes in specific site per cell passage. However, it is very interesting that cell subtype proportions and epigenetic age estimate fluctuate in parallel as a function of age, and is worth of further research.



Specifically, the complete mechanism underlying the epigenetic clock is unclear. The epigenetic age estimate (Horvath 2013) has been shown to reflect many aspects of human health, and is considered as the measure of biological age. However, the phenomenon is difficult to prove credibly using heterogeneous human populations. Importantly, all the numerous features related to biological age may confound the analyses (Jung and Pfeifer 2015). Animal models provide more homogenous experiment conditions for analyzing the question. Nevertheless, it is not a straightforward task to generalize research findings from animals to humans. Accordingly, studies in human populations have great value but systematic summarization of research findings is challenging.

Most of all, knowledge of complete mechanisms, which influence epigenetic changes, is limited. For example, a full understanding of de novo and maintenance DNA methylation as well as DNA demethylation pathways and their interplay with each other is lacking. Regarding EWASs, the aging-associations are pure associations and DNA methylation is lacking experimental causal evidence of being an independent hallmark of aging (Lopez-Otin et al. 2013). Current knowledge (Leenen et al. 2016) suggests that DNA methylation might be merely a side effect (i.e. caused by aging) and is not one of the most crucial cellular mechanisms promoting aging process. In this manner, aging-associated DNA methylation changes might serve only as a biomarker of other functional changes in the cell. However, because of the problematics in the experimental evidencing, these issues remain to be solved.

## 7 SUMMARY AND CONCLUSIONS

This dissertation aimed to characterize (1) aging-associated and (2) mortality-associated methylation features at single CpG site resolution genome-wide. The longitudinal behavior of (3) the epigenetic age and (4) blood cell subtype counts were explored. In addition, (5) the association of CMV infection with epigenetic aging was investigated.

Based on this work, the main conclusions are as follows:

- 1) a. Clocklike-behaving CpG site methylation may be reliably detected from a cross-sectional sample with age range of less than 10 years.
  - b. Generally, aging-associated hypermethylation and hypomethylation at single CpG site resolution were related to different cellular functions, and thus the findings underline the different natures of those events. When compared to hypomethylation, hypermethylation shows signs of being a less random event.
  - c. In order to obtain replicative results from EWASs with heterogenic tissue samples such as whole blood, cell count adjustment appeared to be essential.
- 2) The mortality-predicting methylomic signature performed better than the conventional aging biomarkers and was independent from the aging-associated epigenetic drift. The signature supports the genomic-level role of NF- $\kappa$ B at the very end of the human lifespan.
- 3) Longitudinal methylomic data provided evidence that the difference between chronological and epigenetic age is surprisingly stable during several years or even decades, and when accompanied with previous reports, it may be concluded that the main trajectory of the blood DNA methylome aging rate appears to be largely set before adulthood.
- 4) a. The longitudinal data of blood cell composition also provided evidence that, in parallel with aging-associated shifts in the immune cell composition, intra-individual changes in blood cell subtype proportions are relatively small during young adulthood, middle age, as well as in advanced age for several years or even decades. These results suggest that the major shifts in the blood cell composition might occur somewhere after middle age and before advanced ages.
  - b. The cell proportions correlated with epigenetic aging.
  - c. Consequently, these blood cell composition-related issues warrant careful consideration when interpreting blood cell based results concerning individuals with varying ages.
- 5) The increased DNA methylome age of the blood cells was associated with latent CMV infection in the populations of young adults and nonagenarians; however, this finding may be a reflection of changes in the blood cell composition.

Noteworthy, this thesis provides evidence that the epigenetic age of blood is tightly associated with blood cell subtype frequencies, especially with the immunosenescence-reflecting markers, the CD4<sup>+</sup>CD28<sup>-</sup> and CD8<sup>+</sup>CD28<sup>-</sup> cells and the CD4<sup>+</sup>:CD8<sup>+</sup> ratio (Study III). The most significant correlate linked with DNA methylomic data was CD28<sup>-</sup> T cells, which is a marker of immune system exhaustion and aging.

As the most intriguing finding, the results imply that most of the differences in the epigenetic age arise before adulthood and that the changing rate of the epigenetic age is rather stable thereafter (Study III, Figures 11 and 16). However, it remains to be determined how the inter-individual differences in the epigenetic age arise. Moreover, other very important questions remain to be answered: to what extent is it possible to alter the ticking rate of the epigenetic clock by implementing lifestyle changes (diet, physical activity, etc.) during different stages of life (childhood, adulthood or later)?

In order to answer to these questions, further epigenetic analyses are needed. Long-term follow-up studies with comprehensive phenotypic information similar to, for instance, the Dunedin study setting (Poulton et al. 2015) might provide a deeper understanding of features influencing the epigenetic clock-ticking rate in human populations. In such an analysis, the birth cohort might be followed intensively from the *in utero* life through early childhood to later adulthood. These data might be further coupled with equivalent data from the family and non-relatives in the surrounding environment.

The prospective research has numerous issues to resolve. Single-cell experiments might solve many problematics including that presented in Figure 17. The hypothesized mechanistic links based on associations presented in this thesis and in the literature might be further evaluated using controlled (*in vitro*, *in vivo*) functional studies. Methods with higher genomic coverage instead of using micro array based methodology might provide less biased results (i.e. no enrichment of promoter regions). Preferably, the EWAS might be coupled with multiple omics data (transcriptomics *in cis* and *in trans*, metabolomics, proteomics). Overall, the precise functional role of DNA methylation in aging is still largely unsolved. Therefore, full-scale omics analysis, including epigenetics and covering complete genome in combination with single-cell experiments in human samples and animal models, might unravel many unclear details in the biology of aging. In this way, the role of DNA methylation in the aging process might become clearer.

## 8 ACKNOWLEDGEMENTS

This thesis work was carried out between the years 2013 and 2017 at the University of Tampere (Faculty of Medicine and Life Sciences, Microbiology and immunology, group of Mikko Hurme). The department is acknowledged for providing excellent research facilities and atmosphere.

This thesis was financially supported by the Faculty of Medicine and Life Sciences in University of Tampere, Doctoral Programme in Medicine and Life Sciences in University of Tampere, as well as the Pirkanmaa Hospital District's Science Centre. All are warmly thanked for their support.

I am where I am due to a number of important people. Especially, I want to thank my supervisor and kustos of my defence, Professor Mikko Hurme, MD, PhD, for giving me the great opportunity to work in his group and study features related to biology of aging. I value his support and guidance, which also oriented my path towards becoming an independent researcher. Furthermore, I appreciate that I got a chance to participate in other inspiring research projects alongside my thesis project.

Professor Marja Jylhä, MD, PhD, and Juulia Jylhävä, PhD, I highly value your scientific expertise and contribution to my thesis project in- and outside of the thesis committee meetings. Riikka Lund from University of Turku is especially thanked for accepting the role of opponent in my dissertation. I am very grateful for Cilla Söderhäll, PhD, (Karolinska institute) and Professor Anu Kauppinen, PhD, (University of Eastern-Finland) for helpful suggestions and comments during the pre-examination of the thesis.

I owe great gratitude for all who contributed to this thesis work and beyond. I would like to thank Professors Terho Lehtimäki, MD, PhD, Olli Raitakari, MD, PhD, and Mika Kähönen, MD, PhD, for their important role in organizing the collection and characterization of the Young Finns Study sample used in this thesis work. I thank also Professor Antti Hervonen, MD, PhD and others who were involved in sample collection of Vitality 90+ Study. I thank especially all the volunteers who participated in the studies.

I warmly thank Professor Iris Hovatta, PhD, of my position in the Neurogenomics laboratory during the years 2007 and 2011 at the University of Helsinki and THL. During those times, I realized the beauty of science and molecular biology. The research work in the group of Iris Hovatta and the inspiring company in the Molecular Neurology Research Program at the Biomedicum Helsinki made me hunger for more knowledge. Thus, I continued my studies in Master's Degree Programme in Bioinformatics at the University of Tampere. I also warmly thank Professors Markku Kulomaa, PhD, and Vesa Hytönen, PhD, for welcoming me to Tampere and allowing me to participate their projects in the fields of molecular biotechnology and protein dynamics. I am likewise grateful for Professor Marja Jylhä, MD, PhD, that I got the opportunity to supplement my biology-oriented knowledge of human aging with more sociological perspective.

From the group of Mikko Hurme, I dedicate my enormous gratitude for Saara Marttila, PhD, whose close co-working upgraded this thesis work to a higher level. I also thank Juulia Jylhävä, PhD, for the same reason. Tapio Nevalainen, MSc, is warmly thanked for his intensive contribution to this work (and for being my coffee-mate). Sinikka Repo-Koskinen with her excellent laboratory skills and whole-hearted spirit of mind has been in the central of this thesis work. I wish to thank all others who have contributed to this thesis at the laboratory and at the office.

All my present and former colleagues in different laboratories and research projects during the years are thanked for being excellent co-workers and company with all the necessary non-scientific, pseudoscientific and scientific discussions. Thanks to my most recent office-mates (Saara and Tapio), theoretical background of biology of aging and the purpose of aging during evolution is somewhat clearer to me now. New scientific developments and many aspects of science fiction (among other important topics) have been discussed in details. I have encountered so many people during my years at the FIMM, THL, University of Helsinki and University of Tampere. All these people have made my path to this point possible. I thank members of groups of Kulomaa and Hytönen and all those people who have worked or studied with me at the University of Tampere (Linda, Päivi, Tiina, Laura, Lati, Ulla, Niklas, Outi, Sampo, Rolle, Soili, Sanna, Emma, two Ninas, Marika and others). I want to thank all my loved friends at the Hovatta lab (Tessa, Pia, Kaisa, Jonas, Jenni, Juuso, Mari) during the years 2007 and 2011. Colleagues (Paula, Chris, Joni, Ilse, Brendan, Padmini, Riikka, Alexandra, Katarin, Markus, Heidi and all the others) at the fifth floor at the Biomedicum Helsinki (at the Molecular Neurology Research Program) are remembered with warm thoughts. Similarly, people (Siv, Kaisa, Anu, Ida, Tiia, among others) working at that time at the MLO/KTL and Genomic center (nowadays part of FIMM), are greatly valued. Those were fun and memorable times. All these people taught me how science is made and how important and beneficial a good working atmosphere is for the research.

Finally, I am extremely grateful for my family and friends outside research world. Without this support, this work has not been completed. *This thesis is dedicated especially to Vilma.*

***Åaura***

Tampere, January 2018

## 9 REFERENCES

- Adams JM (Jun 13, 2017) GRG World Supercentenarian Rankings List. <http://grg.org/>.
- Adler AS, Sinha S, Kawahara TL, Zhang JY, Segal E, Chang HY (2007) Motif module map reveals enforcement of aging by continual NF-kappaB activity. *Genes Dev* 21:3244-3257. doi: [gad.1588507](https://doi.org/10.1588507) [pii].
- Ahuja N, Issa JP (2000) Aging, methylation and cancer. *Histol Histopathol* 15:835-842.
- Alonso Arias R, Moro-Garcia MA, Echeverria A, Solano-Jaurrieta JJ, Suarez-Garcia FM, Lopez-Larrea C (2013) Intensity of the humoral response to cytomegalovirus is associated with the phenotypic and functional status of the immune system. *J Virol* 87:4486-4495. doi: [10.1128/JVI.02425-12](https://doi.org/10.1128/JVI.02425-12) [doi].
- Aravinthan A (2015) Cellular senescence: a hitchhiker's guide. *Hum Cell* 28:51-64. doi: [10.1007/s13577-015-0110-x](https://doi.org/10.1007/s13577-015-0110-x) [doi].
- Arbeev KG, Ukraintseva SV, Yashin AI (2016) Dynamics of biomarkers in relation to aging and mortality. *Mech Ageing Dev* 156:42-54. doi: [10.1016/j.mad.2016.04.010](https://doi.org/10.1016/j.mad.2016.04.010) [doi].
- Armstrong NJ, Mather KA, Thalamuthu A, Wright MJ, Trollor JN, Ames D, Brodaty H, Schofield PR, Sachdev PS, Kwok JB (2017) Aging, exceptional longevity and comparisons of the Hannum and Horvath epigenetic clocks. *Epigenomics* 9:689-700. doi: [10.2217/epi-2016-0179](https://doi.org/10.2217/epi-2016-0179) [doi].
- Arnold CR, Wolf J, Brunner S, Herndler-Brandstetter D, Grubeck-Loebenstien B (2011) Gain and loss of T cell subsets in old age--age-related reshaping of the T cell repertoire. *J Clin Immunol* 31:137-146. doi: [10.1007/s10875-010-9499-x](https://doi.org/10.1007/s10875-010-9499-x) [doi].
- Asada K, Abe M, Ushijima T (2013) Clinical application of the CpG island methylator phenotype to prognostic diagnosis in neuroblastomas. *J Hum Genet* 58:428-433. doi: [10.1038/jhg.2013.64](https://doi.org/10.1038/jhg.2013.64) [doi].
- Avner P, Heard E (2001) X-chromosome inactivation: counting, choice and initiation. *Nat Rev Genet* 2:59-67. doi: [10.1038/35047580](https://doi.org/10.1038/35047580) [doi].
- Axume J, Smith SS, Pogribny IP, Moriarty DJ, Caudill MA (2007) The MTHFR 677TT genotype and folate intake interact to lower global leukocyte DNA methylation in young Mexican American women. *Nutr Res* 27:1365-1317. doi: [10.1016/j.nutres.2006.12.006](https://doi.org/10.1016/j.nutres.2006.12.006) [doi].
- Bacalini MG, Friso S, Olivieri F, Pirazzini C, Giuliani C, Capri M, Santoro A, Franceschi C, Garagnani P (2014) Present and future of anti-ageing epigenetic diets. *Mech Ageing Dev* 136-137:101-115. doi: [10.1016/j.mad.2013.12.006](https://doi.org/10.1016/j.mad.2013.12.006) [doi].

Bacalini MG, Deelen J, Pirazzini C, De Cecco M, Giuliani C, Lanzarini C, Ravaoli F, Marasco E, van Heemst D, Suchiman HE et al (2017) Systemic Age-Associated DNA Hypermethylation of ELOVL2 Gene: In Vivo and In Vitro Evidences of a Cell Replication Process. *The Journals of Gerontology: Series A* 72:1015-1023.

Baglietto L, Ponzi E, Haycock P, Hodge A, Bianca Assumma M, Jung CH, Chung J, Fasanelli F, Guida F, Campanella G et al (2017) DNA methylation changes measured in pre-diagnostic peripheral blood samples are associated with smoking and lung cancer risk. *Int J Cancer* 140:50-61. doi: 10.1002/ijc.30431 [doi].

Baker DJ, Dawlaty MM, Wijshake T, Jeganathan KB, Malureanu L, van Ree JH, Crespo-Diaz R, Reyes S, Seaburg L, Shapiro V et al (2013) Increased expression of BubR1 protects against aneuploidy and cancer and extends healthy lifespan. *Nat Cell Biol* 15:96-102. doi: 10.1038/ncb2643 [doi].

Basaiawmoit RV, Rattan SI (2010) Cellular stress and protein misfolding during aging. *Methods Mol Biol* 648:107-117. doi: 10.1007/978-1-60761-756-3\_7 [doi].

Bell JT, Tsai PC, Yang TP, Pidsley R, Nisbet J, Glass D, Mangino M, Zhai G, Zhang F, Valdes A et al (2012) Epigenome-wide scans identify differentially methylated regions for age and age-related phenotypes in a healthy ageing population. *PLoS Genet* 8:e1002629. doi: 10.1371/journal.pgen.1002629 [doi].

Bibikova M, Barnes B, Tsan C, Ho V, Klotzle B, Le JM, Delano D, Zhang L, Schroth GP, Gunderson KL et al (2011) High density DNA methylation array with single CpG site resolution. *Genomics* 98:288-295. doi: 10.1016/j.ygeno.2011.07.007; 10.1016/j.ygeno.2011.07.007.

Bibikova M, Le J, Barnes B, Saedinia-Melnyk S, Zhou L, Shen R, Gunderson KL (2009) Genome-wide DNA methylation profiling using Infinium(R) assay. *Epigenomics* 1:177-200. doi: 10.2217/epi.09.14; 10.2217/epi.09.14.

Bibikova M, Lin Z, Zhou L, Chudin E, Garcia EW, Wu B, Doucet D, Thomas NJ, Wang Y, Vollmer E et al (2006) High-throughput DNA methylation profiling using universal bead arrays. *Genome Res* 16:383-393. doi: gr.4410706 [pii].

Bocklandt S, Lin W, Sehl ME, Sanchez FJ, Sinsheimer JS, Horvath S, Vilain E (2011) Epigenetic predictor of age. *PLoS One* 6:e14821. doi: 10.1371/journal.pone.0014821 [doi].

Bockmühl Y, Patchev AV, Madejska A, Hoffmann A, Sousa JC, Sousa N, Holsboer F, Almeida OFX, Spengler D (2015) Methylation at the CpG island shore region upregulates Nr3c1 promoter activity after early-life stress. *Epigenetics* 10:247-257. doi: 10.1186/s12149-015-0199-1 [pii].

Bojesen SE (2013) Telomeres and human health. *J Intern Med* 274:399-413. doi: 10.1111/joim.12083; 10.1111/joim.12083.

- Boks MP, Derks EM, Weisenberger DJ, Strengman E, Janson E, Sommer IE, Kahn RS, Ophoff RA (2009) The relationship of DNA methylation with age, gender and genotype in twins and healthy controls. *PLoS One* 4:e6767. doi: 10.1371/journal.pone.0006767 [doi].
- Bollati V, Schwartz J, Wright R, Litonjua A, Tarantini L, Suh H, Sparrow D, Vokonas P, Baccarelli A (2009) Decline in genomic DNA methylation through aging in a cohort of elderly subjects. *Mech Ageing Dev* 130:234-239. doi: 10.1016/j.mad.2008.12.003 [doi].
- Bonder MJ, Luijk R, Zhernakova DV, Moed M, Deelen P, Vermaat M, van Iterson M, van Dijk F, van Galen M, Bot J et al (2017) Disease variants alter transcription factor levels and methylation of their binding sites. *Nat Genet* 49:131-138. doi: 10.1038/ng.3721 [doi].
- Breitling LP, Saum KU, Perna L, Schottker B, Holleczeck B, Brenner H (2016) Frailty is associated with the epigenetic clock but not with telomere length in a German cohort. *Clin Epigenetics* 8:21-016-0186-5. eCollection 2016. doi: 10.1186/s13148-016-0186-5 [doi].
- Breitling LP, Yang R, Korn B, Burwinkel B, Brenner H (2011) Tobacco-smoking-related differential DNA methylation: 27K discovery and replication. *Am J Hum Genet* 88:450-457. doi: 10.1016/j.ajhg.2011.03.003 [doi].
- Brooks-Wilson AR (2013) Genetics of healthy aging and longevity. *Hum Genet* 132:1323-1338. doi: 10.1007/s00439-013-1342-z [doi].
- Broux B, Markovic-Plese S, Stinissen P, Hellings N (2012) Pathogenic features of CD4+CD28-T cells in immune disorders. *Trends Mol Med* 18:446-453. doi: 10.1016/j.molmed.2012.06.003 [doi].
- Burkle A, Moreno-Villanueva M, Bernhard J, Blasco M, Zondag G, Hoeijmakers JH, Toussaint O, Grubeck-Loebenstein B, Mocchegiani E, Collino S et al (2015) MARK-AGE biomarkers of ageing. *Mech Ageing Dev* 151:2-12. doi: 10.1016/j.mad.2015.03.006 [doi].
- Cao K, Blair CD, Faddah DA, Kieckhafer JE, Olive M, Erdos MR, Nabel EG, Collins FS (2011) Progerin and telomere dysfunction collaborate to trigger cellular senescence in normal human fibroblasts. *J Clin Invest* 121:2833-2844. doi: 10.1172/JCI43578 [doi].
- Cao R, Wang L, Wang H, Xia L, Erdjument-Bromage H, Tempst P, Jones RS, Zhang Y (2002) Role of Histone H3 Lysine 27 Methylation in Polycomb-Group Silencing. *Science* 298:1039.
- Cao-Lei L, Dancause KN, Elgbeili G, Massart R, Szyf M, Liu A, Laplante DP, King S (2015) DNA methylation mediates the impact of exposure to prenatal maternal stress on BMI and central adiposity in children at age 13(1/2) years: Project Ice Storm. *Epigenetics* 10:749-761. doi: 10.1080/15592294.2015.1063771 [doi].
- Cao-Lei L, Veru F, Elgbeili G, Szyf M, Laplante DP, King S (2016) DNA methylation mediates the effect of exposure to prenatal maternal stress on cytokine production in children at age 13(1/2) years: Project Ice Storm. *Clin Epigenetics* 8:54-016-0219-0. eCollection 2016. doi: 10.1186/s13148-016-0219-0 [doi].



- Carmona JJ, Michan S (2016) Biology of Healthy Aging and Longevity. *Rev Invest Clin* 68:7-16.
- Cerella C, Grandjettette C, Dicato M, Diederich M (2016) Roles of Apoptosis and Cellular Senescence in Cancer and Aging. *Curr Drug Targets* 17:405-415. doi: CDT-EPUB-64946 [pii].
- Chatterton Z, Hartley BJ, Seok MH, Mendeleev N, Chen S, Milekic M, Rosoklija G, Stankov A, Trencvsja-Ivanovska I, Brennan K et al (2017) In utero exposure to maternal smoking is associated with DNA methylation alterations and reduced neuronal content in the developing fetal brain. *Epigenetics Chromatin* 10:4-017-0111-y. eCollection 2017. doi: 10.1186/s13072-017-0111-y [doi].
- Chen BH, Marioni RE, Colicino E, Peters MJ, Ward-Caviness CK, Tsai PC, Roetker NS, Just AC, Demerath EW, Guan W et al (2016) DNA methylation-based measures of biological age: meta-analysis predicting time to death. *Aging (Albany NY)* 8:1844-1865. doi: 10.18632/aging.101020 [doi].
- Chen YA, Lemire M, Choufani S, Butcher DT, Grafodatskaya D, Zanke BW, Gallinger S, Hudson TJ, Weksberg R (2013) Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. *Epigenetics* 8:203-209. doi: 10.4161/epi.23470; 10.4161/epi.23470.
- Christensen BC, Houseman EA, Marsit CJ, Zheng S, Wrensch MR, Wiemels JL, Nelson HH, Karagas MR, Padbury JF, Bueno R et al (2009) Aging and environmental exposures alter tissue-specific DNA methylation dependent upon CpG island context. *PLoS Genet* 5:e1000602. doi: 10.1371/journal.pgen.1000602 [doi].
- Christensen BC, Marsit CJ, Kelsey KT (2012) Influence of Environmental Factors on the Epigenome. Michels KB. *Epigenetic Epidemiology*. , Springer, pp 197.
- Christensen K, Doblhammer G, Rau R, Vaupel JW (2009) Ageing populations: the challenges ahead. *Lancet* 374:1196-1208. doi: 10.1016/S0140-6736(09)61460-4 [doi].
- Christiansen L, Lenart A, Tan Q, Vaupel JW, Aviv A, McGue M, Christensen K (2016) DNA methylation age is associated with mortality in a longitudinal Danish twin study. *Aging Cell* 15:149-154. doi: 10.1111/acel.12421 [doi].
- Cole JH, Ritchie SJ, Bastin ME, Valdes Hernandez MC, Munoz Maniega S, Royle N, Corley J, Pattie A, Harris SE, Zhang Q et al (2017) Brain age predicts mortality. *Mol Psychiatry*. doi: 10.1038/mp.2017.62 [doi].
- Collins LJ, Schönfeld B, Chen XS (2011) The Epigenetics of Non-coding RNA. Anonymous *The Handbook of Epigenetics: The New Molecular and Medical Genetics*. , Elsevier Inc., pp 49.
- Conboy IM, Conboy MJ, Wagers AJ, Girma ER, Weissman IL, Rando TA (2005) Rejuvenation of aged progenitor cells by exposure to a young systemic environment. *Nature* 433:760-764. doi: nature03260 [pii].

Conboy IM, Rando TA (2012) Heterochronic parabiosis for the study of the effects of aging on stem cells and their niches. *Cell Cycle* 11:2260-2267. doi: 10.4161/cc.20437 [doi].

Cournil A, Robine J, Maier H, Gampe J, Vaupel JW (2010) The International Database on Longevity: Structure and contents. Maier H. *Supercentenarians, Demographic Research Monographs*. Verlag Berlin Heidelberg, Springer.

Crimmins E, Vasunilashorn S, Kim JK, Alley D (2008) Biomarkers related to aging in human populations. *Adv Clin Chem* 46:161-216.

Crimmins EM (2015) Lifespan and Healthspan: Past, Present, and Promise. *Gerontologist* 55:901-911. doi: 10.1093/geront/gnv130 [doi].

Crimmins EM, Beltran-Sanchez H (2011) Mortality and morbidity trends: is there compression of morbidity? *J Gerontol B Psychol Sci Soc Sci* 66:75-86. doi: 10.1093/geronb/gbq088 [doi].

Dato S, Rose G, Crocco P, Monti D, Garagnani P, Franceschi C, Passarino G (2017) The genetics of human longevity: an intricacy of genes, environment, culture and microbiome. *Mech Ageing Dev*. doi: S0047-6374(16)30255-X [pii].

de Magalhaes JP, Curado J, Church GM (2009) Meta-analysis of age-related gene expression profiles identifies common signatures of aging. *Bioinformatics* 25:875-881. doi: 10.1093/bioinformatics/btp073 [doi].

Deelen J, Beekman M, Uh HW, Broer L, Ayers KL, Tan Q, Kamatani Y, Bennet AM, Tamm R, Trompet S et al (2014a) Genome-wide association meta-analysis of human longevity identifies a novel locus conferring survival beyond 90 years of age. *Hum Mol Genet* 23:4420-4432. doi: 10.1093/hmg/ddu139 [doi].

Deelen J, Beekman M, Codd V, Trompet S, Broer L, Hägg S, Fischer K, Thijssen PE, Suchiman HED, Postmus I et al (2014b) Leukocyte telomere length associates with prospective mortality independent of immune-related parameters and known genetic markers. *International Journal of Epidemiology*. doi: 10.1093/ije/dyt267.

Doi A, Park IH, Wen B, Murakami P, Aryee MJ, Irizarry R, Herb B, Ladd-Acosta C, Rho J, Loewer S et al (2009) Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. *Nat Genet* 41:1350-1353. doi: 10.1038/ng.471 [doi].

Dong X, Milholland B, Vijg J (2016) Evidence for a limit to human lifespan. *Nature* 538:257-259. doi: 10.1038/nature19793 [doi].

Dozmorov MG (2015) Polycomb repressive complex 2 epigenomic signature defines age-associated hypermethylation and gene expression changes. *Epigenetics* 10:484-495. doi: 1040619 [pii].

- Dufour E, Boulay J, Rincheval V, Sainsard-Chanet A (2000) A causal link between respiration and senescence in *Podospora anserina*. *Proc Natl Acad Sci U S A* 97:4138-4143. doi: 10.1073/pnas.070501997 [doi].
- Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z (2009) GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* 10:48-2105-10-48. doi: 10.1186/1471-2105-10-48 [doi].
- Elhamamsy AR (2017) Role of DNA methylation in imprinting disorders: an updated review. *J Assist Reprod Genet.* doi: 10.1007/s10815-017-0895-5 [doi].
- Elhamamsy AR (2016) DNA methylation dynamics in plants and mammals: overview of regulation and dysregulation. *Cell Biochem Funct* 34:289-298. doi: 10.1002/cbf.3183 [doi].
- Fang M, Chen D, Yang CS (2007) Dietary polyphenols may affect DNA methylation. *J Nutr* 137:223S-228S. doi: 137/1/223S [pii].
- Fang MZ, Chen D, Sun Y, Jin Z, Christman JK, Yang CS (2005) Reversal of hypermethylation and reactivation of p16INK4a, RARBeta, and MGMT genes by genistein and other isoflavones from soy. *Clin Cancer Res* 11:7033-7041. doi: 11/19/7033 [pii].
- Fang MZ, Wang Y, Ai N, Hou Z, Sun Y, Lu H, Welsh W, Yang CS (2003) Tea polyphenol (-)-epigallocatechin-3-gallate inhibits DNA methyltransferase and reactivates methylation-silenced genes in cancer cell lines. *Cancer Res* 63:7563-7570.
- Ferrari C (2004) Beta regression for modelling rates and proportions. *Journal of Applied Statistics* 31:799-815.
- Fischer K, Kettunen J, Wurtz P, Haller T, Havulinna AS, Kangas AJ, Soininen P, Esko T, Tammesoo ML, Magi R et al (2014) Biomarker profiling by nuclear magnetic resonance spectroscopy for the prediction of all-cause mortality: an observational study of 17,345 persons. *PLoS Med* 11:e1001606. doi: 10.1371/journal.pmed.1001606 [doi].
- Florath I, Butterbach K, Muller H, Bewerunge-Hudler M, Brenner H (2014) Cross-sectional and longitudinal changes in DNA methylation with age: an epigenome-wide analysis revealing over 60 novel age-associated CpG sites. *Hum Mol Genet* 23:1186-1201. doi: 10.1093/hmg/ddt531; 10.1093/hmg/ddt531.
- Flores I, Cayuela ML, Blasco MA (2005) Effects of telomerase and telomere length on epidermal stem cell behavior. *Science* 309:1253-1256. doi: 1115025 [pii].
- Fraga MF, Ballestar E, Paz MF, Ropero S, Setien F, Ballestar ML, Heine-Suner D, Cigudosa JC, Urioste M, Benitez J et al (2005) Epigenetic differences arise during the lifetime of monozygotic twins. *Proc Natl Acad Sci U S A* 102:10604-10609. doi: 0500398102 [pii].
- Fraga MF, Esteller M (2007) Epigenetics and aging: the targets and the marks. *Trends in Genetics* 23:413-418. doi: <http://dx.doi.org/10.1016/j.tig.2007.05.008>.

- Franceschi C, Campisi J (2014) Chronic inflammation (inflammaging) and its potential contribution to age-associated diseases. *J Gerontol A Biol Sci Med Sci* 69 Suppl 1:S4-9. doi: 10.1093/gerona/glu057 [doi].
- Francisco Cribari-Neto AZ (2010) Beta Regression in R. *Journal of Statistical Software* 34(2), 1-24. <http://www.jstatsoft.org/v34/i02/>.
- Freire-Aradas A, Phillips C, Lareu MV (2017) Forensic individual age estimation with DNA: From initial approaches to methylation tests. *Forensic Sci Rev* 29:121-144. doi: FSRv29n2-1 [pii].
- Gao X, Jia M, Zhang Y, Breitling LP, Brenner H (2015) DNA methylation changes of whole blood cells in response to active smoking exposure in adults: a systematic review of DNA methylation studies. *Clin Epigenetics* 7:113-015-0148-3. eCollection 2015. doi: 10.1186/s13148-015-0148-3 [doi].
- Garagnani P, Bacalini MG, Pirazzini C, Gori D, Giuliani C, Mari D, Di Blasio AM, Gentilini D, Vitale G, Collino S et al (2012) Methylation of ELOVL2 gene as a new epigenetic marker of age. *Aging Cell* 11:1132-1134. doi: 10.1111/acel.12005 [doi].
- Geginat J, Paroni M, Facciotti F, Gruarin P, Kastirr I, Caprioli F, Pagani M, Abrignani S (2013) The CD4-centered universe of human T cell subsets. *Semin Immunol* 25:252-262. doi: 10.1016/j.smim.2013.10.012 [doi].
- Gentilini D, Mari D, Castaldi D, Remondini D, Ogliari G, Ostan R, Bucci L, Sirchia SM, Tabano S, Cavagnini F et al (2013) Role of epigenetics in human aging and longevity: genome-wide DNA methylation profile in centenarians and centenarians' offspring. *Age (Dordr)* 35:1961-1973. doi: 10.1007/s11357-012-9463-1 [doi].
- Gervin K, Andreassen BK, Hjorthaug HS, Carlsen KC, Carlsen KH, Undlien DE, Lyle R, Munthe-Kaas MC (2016) Intra-individual changes in DNA methylation not mediated by cell-type composition are correlated with aging during childhood. *Clin Epigenetics* 8:110. doi: 10.1186/s13148-016-0277-3 [doi].
- Gkrania-Klotsas E, Langenberg C, Sharp SJ, Luben R, Khaw KT, Wareham NJ (2013) Seropositivity and higher immunoglobulin g antibody levels against cytomegalovirus are associated with mortality in the population-based European prospective investigation of Cancer-Norfolk cohort. *Clin Infect Dis* 56:1421-1427. doi: 10.1093/cid/cit083 [doi].
- Goebeler S, Jylha M, Hervonen A (2003) Medical history, cognitive status and mobility at the age of 90. A population-based study in Tampere, Finland. *Aging Clin Exp Res* 15:154-161.
- Goodwin K, Viboud C, Simonsen L (2006) Antibody response to influenza vaccination in the elderly: A quantitative review. *Vaccine* 24:1159-1169. doi: <https://doi.org/10.1016/j.vaccine.2005.08.105>.

Gopalan S, Carja O, Fagny M, Patin E, Myrick JW, McEwen LM, Mah SM, Kobor MS, Froment A, Feldman MW et al (2017) Trends in DNA Methylation with Age Replicate Across Diverse Human Populations. *Genetics* 206:1659.

Greer EL, Maures TJ, Hauswirth AG, Green EM, Leeman DS, Maro GS, Han S, Banko MR, Gozani O, Brunet A (2010) Members of the H3K4 trimethylation complex regulate lifespan in a germline-dependent manner in *C. elegans*. *Nature* 466:383-387. doi: 10.1038/nature09195 [doi].

Gruber R, Koch H, Doll BA, Tegtmeier F, Einhorn TA, Hollinger JO (2006) Fracture healing in the elderly patient. *Exp Gerontol* 41:1080-1093. doi: S0531-5565(06)00285-3 [pii].

Guéant J (2016) One carbon metabolism, a complex metabolic network involved in pathomechanisms of inherited disorders, birth defects and age-related pathologies. *Biochimie* 126:1-2. doi: <http://dx.doi.org/10.1016/j.biochi.2016.05.008>.

Han S, Brunet A (2012) Histone methylation makes its mark on longevity. *Trends Cell Biol* 22:42-49. doi: 10.1016/j.tcb.2011.11.001 [doi].

Hannum G, Guinney J, Zhao L, Zhang L, Hughes G, Sada S, Klotzle B, Bibikova M, Fan J, Gao Y et al (2013) Genome-wide Methylation Profiles Reveal Quantitative Views of Human Aging Rates. *Mol Cell* 49:359-367. doi: <http://dx.doi.org/10.1016/j.molcel.2012.10.016>.

Harman D (1992) Free radical theory of aging. *Mutat Res* 275:257-266. doi: 0921-8734(92)90030-S [pii].

Harrison DE, Strong R, Sharp ZD, Nelson JF, Astle CM, Flurkey K, Nadon NL, Wilkinson JE, Frenkel K, Carter CS et al (2009) Rapamycin fed late in life extends lifespan in genetically heterogeneous mice. *Nature* 460:392-395. doi: 10.1038/nature08221 [doi].

Hauser MT, Aufsatz W, Jonak C, Luschnig C (2011) Transgenerational epigenetic inheritance in plants. *Biochim Biophys Acta* 1809:459-468. doi: 10.1016/j.bbagr.2011.03.007 [doi].

Hawse WF, Morel PA (2014) An immunology primer for computational modelers. *J Pharmacokinet Pharmacodyn* 41:389-399. doi: 10.1007/s10928-014-9384-y [doi].

He YF, Li BZ, Li Z, Liu P, Wang Y, Tang Q, Ding J, Jia Y, Chen Z, Li L et al (2011) Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* 333:1303-1307. doi: 10.1126/science.1210944 [doi].

Heijmans BT, Tobi EW, Stein AD, Putter H, Blauw GJ, Susser ES, Slagboom PE, Lumey LH (2008) Persistent epigenetic differences associated with prenatal exposure to famine in humans. *Proc Natl Acad Sci U S A* 105:17046-17049. doi: 10.1073/pnas.0806560105 [doi].

Helenius M, Hanninen M, Lehtinen SK, Salminen A (1996) Changes associated with aging and replicative senescence in the regulation of transcription factor nuclear factor-kappa B. *Biochem J* 318 ( Pt 2):603-608.

Hernandez DG, Nalls MA, Gibbs JR, Arepalli S, van der Brug M, Chong S, Moore M, Longo DL, Cookson MR, Traynor BJ et al (2011) Distinct DNA methylation changes highly correlated with chronological age in the human brain. *Hum Mol Genet* 20:1164-1172. doi: 10.1093/hmg/ddq561 [doi].

Herskind AM, McGue M, Holm NV, Sorensen TI, Harvald B, Vaupel JW (1996) The heritability of human longevity: a population-based study of 2872 Danish twin pairs born 1870-1900. *Hum Genet* 97:319-323.

Heyn H, Li N, Ferreira HJ, Moran S, Pisano DG, Gomez A, Diez J, Sanchez-Mut JV, Setien F, Carmona FJ et al (2012) Distinct DNA methylomes of newborns and centenarians. *Proc Natl Acad Sci U S A* 109:10522-10527. doi: 10.1073/pnas.1120658109 [doi].

Hjelmborg Jv, Iachine I, Skytthe A, Vaupel JW, McGue M, Koskenvuo M, Kaprio J, Pedersen NL, Christensen K (2006) Genetic influence on human lifespan and longevity. *Hum Genet* 119:312. doi: 10.1007/s00439-006-0144-y.

Hoerl AE KR (2000) Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* 42:80-86.

Holliday R (2006) Epigenetics: a historical overview. *Epigenetics* 1:76-80. doi: 2762 [pii].

Holliday R, Grigg GW (1993) DNA methylation and mutation. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* 285:61-67. doi: [https://doi.org/10.1016/0027-5107\(93\)90052-H](https://doi.org/10.1016/0027-5107(93)90052-H).

Holliday R, Pugh JE (1975) DNA modification mechanisms and gene activity during development. *Science* 187:226-232.

Horvath S (2013) DNA methylation age of human tissues and cell types. *Genome Biol* 14:R115. doi: gb-2013-14-10-r115 [pii].

Horvath S, Erhart W, Brosch M, Ammerpohl O, von Schonfels W, Ahrens M, Heits N, Bell JT, Tsai PC, Spector TD et al (2014) Obesity accelerates epigenetic aging of human liver. *Proc Natl Acad Sci U S A* 111:15538-15543. doi: 10.1073/pnas.1412759111 [doi].

Horvath S, Garagnani P, Bacalini MG, Pirazzini C, Salvioli S, Gentilini D, Di Blasio AM, Giuliani C, Tung S, Vinters HV et al (2015) Accelerated epigenetic aging in Down syndrome. *Aging Cell* 14:491-495. doi: 10.1111/accel.12325 [doi].

Horvath S, Gurven M, Levine ME, Trumble BC, Kaplan H, Allayee H, Ritz BR, Chen B, Lu AT, Rickabaugh TM et al (2016) An epigenetic clock analysis of race/ethnicity, sex, and coronary heart disease. *Genome Biol* 17:171-016-1030-0. doi: 10.1186/s13059-016-1030-0 [doi].

Horvath S, Langfelder P, Kwak S, Aaronson J, Rosinski J, Vogt TF, Eszes M, Faull RL, Curtis MA, Waldvogel HJ et al (2016) Huntington's disease accelerates epigenetic aging of human

- brain and disrupts DNA methylation levels. *Aging (Albany NY)* 8:1485-1512. doi: 10.18632/aging.101005 [doi].
- Horvath S, Levine AJ (2015) HIV-1 Infection Accelerates Age According to the Epigenetic Clock. *J Infect Dis.* doi: jiv277 [pii].
- Horvath S, Pirazzini C, Bacalini MG, Gentilini D, Di Blasio AM, Delledonne M, Mari D, Arosio B, Monti D, Passarino G et al (2015) Decreased epigenetic age of PBMCs from Italian semi-supercentenarians and their offspring. *Aging (Albany NY)* 7:1159-1170. doi: 100861 [pii].
- Horvath S, Ritz BR (2015) Increased epigenetic age and granulocyte counts in the blood of Parkinson's disease patients. *Aging (Albany NY)* 7:1130-1142. doi: 100859 [pii].
- Horvath S, Zhang Y, Langfelder P, Kahn RS, Boks MP, van Eijk K, van den Berg LH, Ophoff RA (2012) Aging effects on DNA methylation modules in human brain and blood tissue. *Genome Biol* 13:R97. doi: 10.1186/gb-2012-13-10-r97.
- Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, Wiencke JK, Kelsey KT (2012) DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics* 13:86-2105-13-86. doi: 10.1186/1471-2105-13-86 [doi].
- Irizarry RA, Ladd-Acosta C, Wen B, Wu Z, Montano C, Onyango P, Cui H, Gabo K, Rongione M, Webster M et al (2009) The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nat Genet* 41:178-186. doi: 10.1038/ng.298 [doi].
- Ito S, Shen L, Dai Q, Wu SC, Collins LB, Swenberg JA, He C, Zhang Y (2011) Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* 333:1300-1303. doi: 10.1126/science.1210597 [doi].
- Jaffe AE, Irizarry RA (2014) Accounting for cellular heterogeneity is critical in epigenome-wide association studies. *Genome Biol* 15:R31-2014-15-2-r31. doi: 10.1186/gb-2014-15-2-r31 [doi].
- Jakobsson A, Westerberg R, Jacobsson A (2006) Fatty acid elongases in mammals: Their regulation and roles in metabolism. *Progress in Lipid Research* 45:237-249. doi: <https://doi.org/10.1016/j.plipres.2006.01.004>.
- Jin C, Li J, Green CD, Yu X, Tang X, Han D, Xian B, Wang D, Huang X, Cao X et al (2011) Histone demethylase UTX-1 regulates *C. elegans* life span by targeting the insulin/IGF-1 signaling pathway. *Cell Metab* 14:161-172. doi: 10.1016/j.cmet.2011.07.001 [doi].
- Johansson A, Enroth S, Gyllensten U (2013) Continuous Aging of the Human DNA Methylome Throughout the Human Lifespan. *PLoS One* 8:e67378. doi: 10.1371/journal.pone.0067378 [doi].
- Jones PA, Liang G (2012) The human Epigenome. Michels KB. *Epigenetic Epidemiology.* , Springer, pp 5-20.

Jones MJ, Goodman SJ, Kobor MS (2015) DNA methylation and healthy human aging. *Aging Cell*. doi: 10.1111/accel.12349 [doi].

Jones PA (2012) Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet* 13:484-492. doi: 10.1038/nrg3230 [doi].

Jones PA, Liang G (2009) Rethinking how DNA methylation patterns are maintained. *Nat Rev Genet* 10:805-811. doi: 10.1038/nrg2651 [doi].

Jung M, Pfeifer GP (2015) Aging and DNA methylation. *BMC Biol* 13:10.1186/s12915-015-0118-4. doi: 118 [pii].

Juo YY, Johnston FM, Zhang DY, Juo HH, Wang H, Pappou EP, Yu T, Easwaran H, Baylin S, van Engeland M et al (2014) Prognostic value of CpG island methylator phenotype among colorectal cancer patients: a systematic review and meta-analysis. *Ann Oncol* 25:2314-2327. doi: 10.1093/annonc/mdu149 [doi].

Jurk D, Wilson C, Passos JF, Oakley F, Correia-Melo C, Greaves L, Saretzki G, Fox C, Lawless C, Anderson R et al (2014) Chronic inflammation induces telomere dysfunction and accelerates ageing in mice. *Nat Commun* 2:4172. doi: 10.1038/ncomms5172 [doi].

Jylhava J (2014) Determinants of longevity: genetics, biomarkers and therapeutic approaches. *Curr Pharm Des* 20:6058-6070. doi: CPD-EPUB-59650 [pii].

Jylhava J, Hurme M (2010) Gene variants as determinants of longevity: focus on the inflammatory factors. *Pflugers Arch* 459:239-246. doi: 10.1007/s00424-009-0726-3 [doi].

Jylhava J, Kananen L, Raitanen J, Marttila S, Nevalainen T, Hervonen A, Jylha M, Hurme M (2016) Methyloomic predictors demonstrate the role of NF-kappaB in old-age mortality and are unrelated to the aging-associated epigenetic drift. *Oncotarget* 7:19228-19241. doi: 10.18632/oncotarget.8278 [doi].

Jylhava J, Kotipelto T, Raitala A, Jylha M, Hervonen A, Hurme M (2011) Aging is associated with quantitative and qualitative changes in circulating cell-free DNA: the Vitality 90+ study. *Mech Ageing Dev* 132:20-26. doi: 10.1016/j.mad.2010.11.001 [doi].

Jylhava J, Pedersen NL, Hagg S (2017) Biological Age Predictors. *EBioMedicine*. doi: S2352-3964(17)30142-1 [pii].

Jylhava J, Raitanen J, Marttila S, Hervonen A, Jylha M, Hurme M (2014) Identification of a prognostic signature for old-age mortality by integrating genome-wide transcriptomic data with the conventional predictors: the Vitality 90+ Study. *BMC Med Genomics* 7:54-8794-7-54. doi: 10.1186/1755-8794-7-54 [doi].

Kaerberlein MR, Martin GM, Kaerberlein TL (2016) *Handbook of the biology of aging*. , Springer.



Kananen L, Marttila S, Nevalainen T, Kummola L, Junttila I, Mononen N, Kahonen M, Raitakari OT, Hervonen A, Jylha M et al (2016) The trajectory of the blood DNA methylome ageing rate is largely set before adulthood: evidence from two longitudinal studies. *Age (Dordr)* 38:65-016-9927-9. Epub 2016 Jun 14. doi: 10.1007/s11357-016-9927-9 [doi].

Kananen L, Surakka I, Pirkola S, Suvisaari J, Lonqvist J, Peltonen L, Ripatti S, Hovatta I (2010) Childhood adversities are associated with shorter telomere length at adult age both in individuals with an anxiety disorder and controls. *PLoS One* 5:e10826. doi: 10.1371/journal.pone.0010826 [doi].

Kaushal A, Zhang H, Karmaus WJJ, Everson TM, Marsit CJ, Karagas MR, Tsai S, Wen H, Wang S (2017) Genome-wide DNA methylation at birth in relation to in utero arsenic exposure and the associated health in later life. *Environ Health* 16:50. doi: 10.1186/s12940-017-0262-0.

Kohli RM, Zhang Y (2013) TET enzymes, TDG and the dynamics of DNA demethylation. *Nature* 502:472-479. doi: 10.1038/nature12750 [doi].

Kozlenkov A, Jaffe AE, Timashpolsky A, Apontes P, Rudchenko S, Barbu M, Byne W, Hurd YL, Horvath S, Dracheva S (2017) DNA Methylation Profiling of Human Prefrontal Cortex Neurons in Heroin Users Shows Significant Difference between Genomic Contexts of Hyper- and Hypomethylation and a Younger Epigenetic Age. *Genes (Basel)* 8:10.3390/genes8060152. doi: E152 [pii].

Krabbe KS, Pedersen M, Bruunsgaard H (2004) Inflammatory mediators in the elderly. *Exp Gerontol* 39:687-699. doi: 10.1016/j.exger.2004.01.009 [doi].

Lara J, Cooper R, Nissan J, Ginty AT, Khaw KT, Deary IJ, Lord JM, Kuh D, Mathers JC (2015) A proposed panel of biomarkers of healthy ageing. *BMC Med* 13:222-015-0470-9. doi: 10.1186/s12916-015-0470-9 [doi].

Lee SJ, Go AS, Lindquist K, Bertenthal D, Covinsky KE (2008) Chronic conditions and mortality among the oldest old. *Am J Public Health* 98:1209-1214. doi: 10.2105/AJPH.2007.130955 [doi].

Leenen FA, Muller CP, Turner JD (2016) DNA methylation: conducting the orchestra from exposure to phenotype? *Clin Epigenetics* 8:92-016-0256-8. eCollection 2016. doi: 10.1186/s13148-016-0256-8 [doi].

Leonard AE, Kelder B, Bobik EG, Chuang L, Lewis CJ, Kopchick JJ, Mukerji P, Huang Y (2002) Identification and expression of mammalian long-chain PUFA elongation enzymes. *Lipids* 37:733-740. doi: 10.1007/s11745-002-0955-6.

Levine ME, Hosgood HD, Chen B, Absher D, Assimes T, Horvath S (2015) DNA methylation age of blood predicts future onset of lung cancer in the women's health initiative. *Aging (Albany NY)* 7:690-700. doi: 100809 [pii].

Levine ME, Lu AT, Bennett DA, Horvath S (2015) Epigenetic age of the pre-frontal cortex is associated with neuritic plaques, amyloid load, and Alzheimer's disease related cognitive functioning. *Aging (Albany NY)* 7:1198-1211. doi: 100864 [pii].

Levine ME, Lu AT, Chen BH, Hernandez DG, Singleton AB, Ferrucci L, Bandinelli S, Salfati E, Manson JE, Quach A et al (2016) Menopause accelerates biological aging. *Proc Natl Acad Sci U S A* 113:9327-9332. doi: 10.1073/pnas.1604558113 [doi].

Li Y, Yin Y, Mariuzza RA (2013) Structural and biophysical insights into the role of CD4 and CD8 in T cell activation. *Front Immunol* 4:206. doi: 10.3389/fimmu.2013.00206 [doi].

Lin Q, Weidner CI, Costa IG, Marioni RE, Ferreira MR, Deary IJ, Wagner W (2016) DNA methylation levels at individual age-associated CpG sites can be indicative for life expectancy. *Aging (Albany NY)* 8:394-401. doi: 100908 [pii].

Liu Y, Murphy SK, Murtha AP, Fuemmeler BF, Schildkraut J, Huang Z, Overcash F, Kurtzberg J, Jirtle R, Iversen ES et al (2012) Depression in pregnancy, infant birth weight and DNA methylation of imprint regulatory elements. *Epigenetics* 7:735-746. doi: 10.4161/epi.20734 [doi].

Liu Z, Xie Z, Jones W, Pavlovicz RE, Liu S, Yu J, Li PK, Lin J, Fuchs JR, Marcucci G et al (2009) Curcumin is a potent DNA hypomethylation agent. *Bioorg Med Chem Lett* 19:706-709. doi: 10.1016/j.bmcl.2008.12.041 [doi].

Lopez-Otin C, Blasco MA, Partridge L, Serrano M, Kroemer G (2013) The hallmarks of aging. *Cell* 153:1194-1217. doi: 10.1016/j.cell.2013.05.039; 10.1016/j.cell.2013.05.039.

Lowe D, Horvath S, Raj K (2016) Epigenetic clock analyses of cellular senescence and ageing. *Oncotarget* 7:8524-8531. doi: 10.18632/oncotarget.7383 [doi].

Lu SC (2000) S-Adenosylmethionine. *Int J Biochem Cell Biol* 32:391-395. doi: [http://dx.doi.org/10.1016/S1357-2725\(99\)00139-9](http://dx.doi.org/10.1016/S1357-2725(99)00139-9).

Ma X, Wang YW, Zhang MQ, Gazdar AF (2013) DNA methylation data analysis and its application to cancer research. *Epigenomics* 5:301-316. doi: 10.2217/epi.13.26 [doi].

Marino G, Ugalde AP, Fernandez AF, Osorio FG, Fueyo A, Freije JM, Lopez-Otin C (2010) Insulin-like growth factor 1 treatment extends longevity in a mouse model of human premature aging by restoring somatotroph axis function. *Proc Natl Acad Sci U S A* 107:16268-16273. doi: 10.1073/pnas.1002696107 [doi].

Marioni RE, Shah S, McRae AF, Chen BH, Colicino E, Harris SE, Gibson J, Henders AK, Redmond P, Cox SR et al (2015) DNA methylation age of blood predicts all-cause mortality in later life. *Genome Biol* 16:25-015-0584-6. doi: 10.1186/s13059-015-0584-6 [doi].

Marioni RE, Shah S, McRae AF, Ritchie SJ, Muniz-Terrera G, Harris SE, Gibson J, Redmond P, Cox SR, Pattie A et al (2015) The epigenetic clock is correlated with physical and cognitive fitness in the Lothian Birth Cohort 1936. *Int J Epidemiol*. doi: dyu277 [pii].

Mariuswalter W (2016) DNAm\_landscape.png (Retrieved 8.12.2017).  
[https://commons.wikimedia.org/wiki/File:DNAm\\_landscape.png](https://commons.wikimedia.org/wiki/File:DNAm_landscape.png).

Martino D, Loke YJ, Gordon L, Ollikainen M, Cruickshank MN, Saffery R, Craig JM (2013) Longitudinal, genome-scale analysis of DNA methylation in twins from birth to 18 months of age reveals rapid epigenetic change in early life and pair-specific effects of discordance. *Genome Biol* 14:R42-2013-14-5-r42. doi: 10.1186/gb-2013-14-5-r42 [doi].

Martino DJ, Tulic MK, Gordon L, Hodder M, Richman TR, Metcalfe J, Prescott SL, Saffery R (2011) Evidence for age-related and individual-specific changes in DNA methylation profile of mononuclear cells during early immune development in humans. *Epigenetics* 6:1085-1094. doi: 10.4161/epi.6.9.16401 [doi].

Martos SN, Tang WY, Wang Z (2015) Elusive inheritance: Transgenerational effects and epigenetic inheritance in human environmental disease. *Prog Biophys Mol Biol* 118:44-54. doi: 10.1016/j.pbiomolbio.2015.02.011 [doi].

Marttila S, Kananen L, Hayrynen S, Jylhava J, Nevalainen T, Hervonen A, Jylha M, Nykter M, Hurme M (2015) Ageing-associated changes in the human DNA methylome: genomic locations and effects on gene expression. *BMC Genomics* 16:1381-015-1381-z. Epub 2015 Mar 14. doi: 10.1186/s12864-015-1381-z [doi].

Marttila S, Kananen L, Jylhava J, Nevalainen T, Hervonen A, Jylha M, Hurme M (2015) Length of paternal lifespan is manifested in the DNA methylome of their nonagenarian progeny. *Oncotarget* 6:30557-30567. doi: 10.18632/oncotarget.5905 [doi].

Maunakea AK, Nagarajan RP, Bilienky M, Ballinger TJ, D'Souza C, Fouse SD, Johnson BE, Hong C, Nielsen C, Zhao Y et al (2010) Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature* 466:253-257. doi: 10.1038/nature09165 [doi].

McClay JL, Aberg KA, Clark SL, Nerella S, Kumar G, Xie LY, Hudson AD, Harada A, Hultman CM, Magnusson PK et al (2014) A methylome-wide study of aging using massively parallel sequencing of the methyl-CpG-enriched genomic fraction from blood in over 700 subjects. *Hum Mol Genet* 23:1175-1185. doi: 10.1093/hmg/ddt511 [doi].

Mohandas T, Sparkes RS, Shapiro LJ (1981) Reactivation of an inactive human X chromosome: evidence for X inactivation by DNA methylation. *Science* 211:393-396.

Molofsky AV, Slutsky SG, Joseph NM, He S, Pardal R, Krishnamurthy J, Sharpless NE, Morrison SJ (2006) Increasing p16INK4a expression decreases forebrain progenitors and neurogenesis during ageing. *Nature* 443:448-452. doi: nature05091 [pii].

Moore AZ, Hernandez DG, Tanaka T, Pilling LC, Nalls MA, Bandinelli S, Singleton AB, Ferrucci L (2016) Change in Epigenome-Wide DNA Methylation Over 9 Years and Subsequent Mortality: Results From the InCHIANTI Study. *J Gerontol A Biol Sci Med Sci* 71:1029-1035. doi: 10.1093/gerona/glv118 [doi].

Morgan HD, Sutherland HG, Martin DI, Whitelaw E (1999) Epigenetic inheritance at the agouti locus in the mouse. *Nat Genet* 23:314-318. doi: 10.1038/15490 [doi].

Moro-Garcia MA, Alonso-Arias R, Lopez-Vazquez A, Suarez-Garcia FM, Solano-Jaurrieta JJ, Baltar J, Lopez-Larrea C (2012) Relationship between functional ability in older people, immune system status, and intensity of response to CMV. *Age (Dordr)* 34:479-495. doi: 10.1007/s11357-011-9240-6 [doi].

Morrow G, Samson M, Michaud S, Tanguay RM (2004) Overexpression of the small mitochondrial Hsp22 extends *Drosophila* life span and increases resistance to oxidative stress. *FASEB J* 18:598-599. doi: 10.1096/fj.03-0860fje [doi].

Murphy K, Travers P, Walport M (2008) Chapter 1: Basic concepts in immunology. Schanck D, Masson S, Lawrence E, Lucas G, Goatly B. *Janeway's immunobiology*. 7, Garland Science, Taylor&Francis Group, pp 1-37.

Naess AB, Kirkengen AL (2015) Is childhood stress associated with shorter telomeres? *Tidsskr Nor Laegeforen* 135:1356-1360. doi: 10.4045/tidsskr.14.1194 [doi].

Nevalainen T, Kananen L, Marttila S, Jylhava J, Mononen N, Kahonen M, Raitakari OT, Hervonen A, Jylha M, Lehtimaki T et al (2017) Obesity accelerates epigenetic aging in middle-aged but not in elderly individuals. *Clin Epigenetics* 9:20-016-0301-7. eCollection 2017. doi: 10.1186/s13148-016-0301-7 [doi].

NIH (2017) National Institutes of Health, Epigenetic mechanisms.jpg. <https://commons.wikimedia.org/w/index.php?curid=9789221>. Accessed 12/5/2017.

Nuotio J, Oikonen M, Magnussen CG, Jokinen E, Laitinen T, Hutri-Kahonen N, Kahonen M, Lehtimaki T, Taittonen L, Tossavainen P et al (2014) Cardiovascular risk factors in 2011 and secular trends since 2007: the Cardiovascular Risk in Young Finns Study. *Scand J Public Health* 42:563-571. doi: 10.1177/1403494814541597 [doi].

Official Statistics of Finland, OSF (2015) Population projection [e-publication]. Last updated 30.10.2015.

ISSN=1798-5153. 2015, Quality description: Population projection 2015–2065 . Helsinki: Statistics Finland [referred: 7.12.2017]. [http://www.stat.fi/til/vaenn/2015/vaenn\\_2015\\_2015-10-30\\_laa\\_001\\_en.html](http://www.stat.fi/til/vaenn/2015/vaenn_2015_2015-10-30_laa_001_en.html).

Osorio FG, Barcena C, Soria-Valles C, Ramsay AJ, de Carlos F, Cobo J, Fueyo A, Freije JM, Lopez-Otin C (2012) Nuclear lamina defects cause ATM-dependent NF-kappaB activation and link accelerated aging to a systemic inflammatory response. *Genes Dev* 26:2311-2324. doi: 10.1101/gad.197954.112 [doi].

- Pacaud R, Brocard E, Lalier L, Hervouet E, Vallette FM, Cartron P (2014) The DNMT1/PCNA/UHRF1 disruption induces tumorigenesis characterized by similar genetic and epigenetic signatures. *Scientific Reports* 4:4230.
- Pan H, Finkel T (2017) Key Proteins and Pathways that Regulate Lifespan. *Journal of Biological Chemistry*. doi: 10.1074/jbc.R116.771915.
- Pathai S, Bajillan H, Landay AL, High KP (2014) Is HIV a Model of Accelerated or Accentuated Aging? *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences* 69:833-842. doi: 10.1093/gerona/glt168.
- Pawelec G (2017) Does the human immune system ever really become "senescent"? version 1; referees: 5 approved]. *F1000Research* 6. doi: 10.12688/f1000research.11297.1.
- Pawelec G, Derhovanessian E (2011) Role of CMV in immune senescence. *Virus Res* 157:175-179. doi: 10.1016/j.virusres.2010.09.010 [doi].
- Pawelec G, Larbi A, Derhovanessian E (2010) Senescence of the human immune system. *J Comp Pathol* 142 Suppl 1:S39-44. doi: 10.1016/j.jcpa.2009.09.005 [doi].
- Pawelec G, McElhane JE, Aiello AE, Derhovanessian E (2012) The impact of CMV infection on survival in older humans. *Curr Opin Immunol* 24:507-511. doi: 10.1016/j.coi.2012.04.002 [doi].
- Perna L, Zhang Y, Mons U, Holleczeck B, Saum KU, Brenner H (2016) Epigenetic age acceleration predicts cancer, cardiovascular, and all-cause mortality in a German case cohort. *Clin Epigenetics* 8:64-016-0228-z. eCollection 2016. doi: 10.1186/s13148-016-0228-z [doi].
- Pidsley R, Wong CC, Volta M, Lunnon K, Mill J, Schalkwyk LC (2013) A data-driven approach to preprocessing Illumina 450K methylation array data. *BMC Genomics* 14:293-2164-14-293. doi: 10.1186/1471-2164-14-293; 10.1186/1471-2164-14-293.
- Poulton R, Moffitt TE, Silva PA (2015) The Dunedin Multidisciplinary Health and Development Study: overview of the first 40 years, with an eye to the future. *Soc Psychiatry Psychiatr Epidemiol* 50:679-693. doi: 10.1007/s00127-015-1048-8.
- Pourghesari B, Khan N, Best D, Bruton R, Nayak L, Moss PA (2007) The cytomegalovirus-specific CD4+ T-cell response expands with age and markedly alters the CD4+ T-cell repertoire. *J Virol* 81:7759-7765. doi: JVI.01262-06 [pii].
- Quach A, Levine ME, Tanaka T, Lu AT, Chen BH, Ferrucci L, Ritz B, Bandinelli S, Neuhauser ML, Beasley JM et al (2017) Epigenetic clock analysis of diet, exercise, education, and lifestyle factors. *Aging (Albany NY)* 9:419-446. doi: 10.18632/aging.101168 [doi].
- Raitakari OT, Juonala M, Ronnema T, Keltikangas-Jarvinen L, Rasanen L, Pietikainen M, Hutri-Kahonen N, Taittonen L, Jokinen E, Marniemi J et al (2008) Cohort profile: the

cardiovascular risk in Young Finns Study. *Int J Epidemiol* 37:1220-1226. doi: 10.1093/ije/dym225 [doi].

Rakyan VK, Down TA, Maslau S, Andrew T, Yang TP, Beyan H, Whittaker P, McCann OT, Finer S, Valdes AM et al (2010) Human aging-associated DNA hypermethylation occurs preferentially at bivalent chromatin domains. *Genome Res* 20:434-439. doi: 10.1101/gr.103101.109 [doi].

Rasmussen KD, Helin K (2016) Role of TET enzymes in DNA methylation, development, and cancer. *Genes Dev* 30:733-750. doi: 10.1101/gad.276568.115 [doi].

Rector T, Taylor B, Sultan S, Shaikat A, Adabag S, Nelson D, Capecchi T, MacDonald R, Greer N, Wilt T (2016) Estimation of life expectancy, Calculators. . doi: NBK424560 [bookaccession].

Reinius LE, Acevedo N, Joerink M, Pershagen G, Dahlen SE, Greco D, Soderhall C, Scheynius A, Kere J (2012) Differential DNA methylation in purified human blood cells: implications for cell lineage and studies on disease susceptibility. *PLoS One* 7:e41361. doi: 10.1371/journal.pone.0041361 [doi].

Reynolds LM, Taylor JR, Ding J, Lohman K, Johnson C, Siscovick D, Burke G, Post W, Shea S, Jacobs DR, Jr et al (2014) Age-related variations in the methylome associated with gene expression in human monocytes and T cells. *Nat Commun* 5:5366. doi: 10.1038/ncomms6366 [doi].

Riggs AD (2002) X chromosome inactivation, differentiation, and DNA methylation revisited, with a tribute to Susumu Ohno. *Cytogenet Genome Res* 99:17-24. doi: 71569 [doi].

Roberts ET, Haan MN, Dowd JB, Aiello AE (2010) Cytomegalovirus antibody levels, inflammation, and mortality among elderly Latinos over 9 years of follow-up. *Am J Epidemiol* 172:363-371. doi: 10.1093/aje/kwq177 [doi].

Roseboom TJ, Painter RC, van Abeelen AF, Veenendaal MV, de Rooij SR (2011) Hungry in the womb: what are the consequences? Lessons from the Dutch famine. *Maturitas* 70:141-145. doi: 10.1016/j.maturitas.2011.06.017 [doi].

Salminen A, Huuskonen J, Ojala J, Kauppinen A, Kaarniranta K, Suuronen T (2008) Activation of innate immunity system during aging: NF- $\kappa$ B signaling is the molecular culprit of inflamm-aging. *Ageing Res Rev* 7:83-105. doi: S1568-1637(07)00056-6 [pii].

Salomon JA, Wang H, Freeman MK, Vos T, Flaxman AD, Lopez AD, Murray CJ (2012) Healthy life expectancy for 187 countries, 1990-2010: a systematic analysis for the Global Burden Disease Study 2010. *Lancet* 380:2144-2162. doi: 10.1016/S0140-6736(12)61690-0 [doi].

Schroder K, Tschoop J (2010) The Inflammasomes. *Cell* 140:821-832. doi: <https://doi.org/10.1016/j.cell.2010.01.040>.

Schübeler D (2015) ESCI award lecture: regulation, function and biomarker potential of DNA methylation. *Eur J Clin Invest* 45:288-293. doi: 10.1111/eci.12403.

Schuermann D, Weber AR, Schär P (2016) Active DNA demethylation by DNA repair: Facts and uncertainties. *DNA Repair* 44:92-102. doi: <http://dx.doi.org/10.1016/j.dnarep.2016.05.013>.

Sharma A (2015) Transgenerational epigenetic inheritance: resolving uncertainty and evolving biology. *Biomol Concepts* 6:87-103. doi: 10.1515/bmc-2015-0005 [doi].

Sharma A (2013) Transgenerational epigenetic inheritance: focus on soma to germline information transfer. *Prog Biophys Mol Biol* 113:439-446. doi: 10.1016/j.pbiomolbio.2012.12.003 [doi].

Sharma A (2017) Transgenerational epigenetics: Integrating soma to germline communication with gametic inheritance. *Mechanisms of Ageing and Development* 163:15-22. doi: <https://doi.org/10.1016/j.mad.2016.12.015>.

Shkolnikov V, Barbieri M, Wilmoth J (2016) The Human Mortality Database. <http://www.mortality.org>.

Siebold AP, Banerjee R, Tie F, Kiss DL, Moskowitz J, Harte PJ (2010) Polycomb Repressive Complex 2 and Trithorax modulate *Drosophila* longevity and stress resistance. *Proc Natl Acad Sci U S A* 107:169-174. doi: 10.1073/pnas.0907739107 [doi].

Simmen MW (2008) Genome-scale relationships between cytosine methylation and dinucleotide abundances in animals. *Genomics* 92:33-40. doi: <https://doi.org/10.1016/j.ygeno.2008.03.009>.

Simpkin AJ, Hemani G, Suderman M, Gaunt TR, Lyttleton O, Mcardle WL, Ring SM, Sharp GC, Tilling K, Horvath S et al (2016) Prenatal and early life influences on epigenetic age in children: a study of mother-offspring pairs from two cohort studies. *Hum Mol Genet* 25:191-201. doi: 10.1093/hmg/ddv456 [doi].

Singh T, Newman AB (2011) Inflammatory markers in population studies of aging. *Ageing Res Rev* 10:319-329. doi: 10.1016/j.arr.2010.11.002 [doi].

Skytthe A, Pedersen NL, Kaprio J, Stazi MA, Hjelmborg JV, Iachine I, Vaupel JW, Christensen K (2003) Longevity studies in GenomEUtwin. *Twin Res* 6:448-454. doi: 10.1375/136905203770326457 [doi].

Slieker RC, van Iterson M, Luijk R, Beekman M, Zhernakova DV, Moed MH, Mei H, van Galen M, Deelen P, Bonder MJ et al (2016) Age-related accrual of methylomic variability is linked to fundamental ageing mechanisms. *Genome Biol* 17:191. doi: 10.1186/s13059-016-1053-6 [doi].

- Smith ZD, Meissner A (2013) DNA methylation: roles in mammalian development. *Nat Rev Genet* 14:204-220.
- Spiers H, Hannon E, Wells S, Williams B, Fernandes C, Mill J (2016) Age-associated changes in DNA methylation across multiple tissues in an inbred mouse model. *Mech Ageing Dev* 154:20-23. doi: S0047-6374(16)30006-9 [pii].
- Spruijt CG, Vermeulen M (2014) DNA methylation: old dog, new tricks? *Nat Struct Mol Biol* 21:949-954. doi: 10.1038/nsmb.2910 [doi].
- Starnawska A, Tan Q, Lenart A, McGue M, Mors O, Børglum AD, Christensen K, Nyegaard M, Christiansen L (2017) Blood DNA methylation age is not associated with cognitive functioning in middle-aged monozygotic twins. *Neurobiology of Aging* 50:60-63. doi: <http://dx.doi.org/10.1016/j.neurobiolaging.2016.10.025>.
- Steggenga WT, Boekschoten MV, Lute C, Hooiveld GJ, de Groot PJ, Morris TJ, Teschendorff AE, Butcher LM, Beck S, Muller M (2014) Genome-wide age-related changes in DNA methylation and gene expression in human PBMCs. *Age (Dordr)* 36:9648-014-9648-x. Epub 2014 May 2. doi: 10.1007/s11357-014-9648-x [doi].
- Strandberg TE, Pitkala KH, Tilvis RS (2009) Cytomegalovirus antibody level and mortality among community-dwelling older adults with stable cardiovascular disease. *JAMA* 301:380-382. doi: 10.1001/jama.2009.4 [doi].
- Stubbs TM, Bonder MJ, Stark AK, Krueger F, BI Ageing Clock Team, von Meyenn F, Stegle O, Reik W (2017) Multi-tissue DNA methylation age predictor in mouse. *Genome Biol* 18:68-017-1203-5. doi: 10.1186/s13059-017-1203-5 [doi].
- Su D, Wang X, Campbell MR, Porter DK, Pittman GS, Bennett BD, Wan M, Englert NA, Crowl CL, Gimple RN et al (2016) Distinct Epigenetic Effects of Tobacco Smoking in Whole Blood and among Leukocyte Subtypes. *PLoS One* 11:e0166486. doi: 10.1371/journal.pone.0166486 [doi].
- Supek F, Bosnjak M, Skunca N, Smuc T (2011) REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One* 6:e21800. doi: 10.1371/journal.pone.0021800 [doi].
- Swindell WR (2009) Genes and gene expression modules associated with caloric restriction and aging in the laboratory mouse. *BMC Genomics* 10:585-2164-10-585. doi: 10.1186/1471-2164-10-585 [doi].
- Swindell WR, Masternak MM, Kopchick JJ, Conover CA, Bartke A, Miller RA (2009) Endocrine regulation of heat shock protein mRNA levels in long-lived dwarf mice. *Mech Ageing Dev* 130:393-400. doi: 10.1016/j.mad.2009.03.004 [doi].
- Sylwester AW, Mitchell BL, Edgar JB, Taormina C, Pelte C, Ruchti F, Sleath PR, Grabstein KH, Hosken NA, Kern F et al (2005) Broadly targeted human cytomegalovirus-specific CD4+



and CD8+ T cells dominate the memory compartments of exposed subjects. *J Exp Med* 202:673-685. doi: jem.20050882 [pii].

Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, Agarwal S, Iyer LM, Liu DR, Aravind L et al (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* 324:930-935. doi: 10.1126/science.1170116 [doi].

Talamonti E, Pauter AM, Asadi A, Fischer AW, Chiurchià V, Jacobsson A (2017) Impairment of systemic DHA synthesis affects macrophage plasticity and polarization: implications for DHA supplementation during inflammation. *Cell Mol Life Sci* 74:2815-2826. doi: 2498 [pii].

Talens RP, Christensen K, Putter H, Willemsen G, Christiansen L, Kremer D, Suchiman HE, Slagboom PE, Boomsma DI, Heijmans BT (2012) Epigenetic variation during the adult lifespan: cross-sectional and longitudinal data on monozygotic twin pairs. *Aging Cell* 11:694-703. doi: 10.1111/j.1474-9726.2012.00835.x [doi].

Teschendorff AE, West J, Beck S (2013) Age-associated epigenetic drift: implications, and a case of epigenetic thrift? *Hum Mol Genet* 22:R7-R15. doi: 10.1093/hmg/ddt375 [doi].

Thum T (2014) Non-coding RNAs in ageing. *Ageing Res Rev* 17:1-2. doi: 10.1016/j.arr.2014.08.001 [doi].

Tilstra JS, Robinson AR, Wang J, Gregg SQ, Clauson CL, Reay DP, Nasto LA, St Croix CM, Usas A, Vo N et al (2012) NF-kappaB inhibition delays DNA damage-induced senescence and aging in mice. *J Clin Invest* 122:2601-2612. doi: 10.1172/JCI45785 [doi].

Tobi EW, Lumey LH, Talens RP, Kremer D, Putter H, Stein AD, Slagboom PE, Heijmans BT (2009) DNA methylation differences after exposure to prenatal famine are common and timing- and sex-specific. *Hum Mol Genet* 18:4046-4053. doi: 10.1093/hmg/ddp353 [doi].

Tu W, Rao S (2016) Mechanisms Underlying T Cell Immunosenescence: Aging and Cytomegalovirus Infection. *Front Microbiol* 7:2111. doi: 10.3389/fmicb.2016.02111 [doi].

Valentini E, Zampieri M, Malavolta M, Bacalini MG, Calabrese R, Guastafierro T, Reale A, Franceschi C, Hervonen A, Koller B et al (2016) Analysis of the machinery and intermediates of the 5hmC-mediated DNA demethylation pathway in aging on samples from the MARK-AGE Study. *Aging (Albany NY)* 8:1896-1922. doi: 10.18632/aging.101022 [doi].

van Dongen J, Nivard MG, Willemsen G, Hottenga JJ, Helmer Q, Dolan CV, Ehli EA, Davies GE, van Ijzerman M, Breeze CE et al (2016) Genetic and environmental influences interact with age and sex in shaping the human methylome. *Nat Commun* 7:11115. doi: 10.1038/ncomms11115 [doi].

Varela-Rey M, Woodhoo A, Martinez-Chantar ML, Mato JM, Lu SC (2013) Alcohol, DNA methylation, and cancer. *Alcohol Res* 35:25-35.

Veru F, Laplante DP, Luheshi G, King S (2014) Prenatal maternal stress exposure and immune function in the offspring. *Stress* 17:133-148. doi: 10.3109/10253890.2013.876404 [doi].

Vina J, Borrás C, Abdelaziz KM, García-Valles R, Gómez-Cabrera MC (2013) The free radical theory of aging revisited: the cell signaling disruption theory of aging. *Antioxid Redox Signal* 19:779-787. doi: 10.1089/ars.2012.5111 [doi].

Wagner W (2017) Epigenetic aging clocks in mice and men. *Genome Biol* 18:107. doi: 10.1186/s13059-017-1245-8.

Walker GA, Lithgow GJ (2003) Lifespan extension in *C. elegans* by a molecular chaperone dependent upon insulin-like signals. *Aging Cell* 2:131-139.

Wang GC, Kao WH, Murakami P, Xue QL, Chiou RB, Detrick B, McDyer JF, Semba RD, Casolaro V, Walston JD et al (2010) Cytomegalovirus infection and the risk of mortality and frailty in older women: a prospective observational cohort study. *Am J Epidemiol* 171:1144-1152. doi: 10.1093/aje/kwq062 [doi].

Weidner CI, Lin Q, Koch CM, Eisele L, Beier F, Ziegler P, Bauerschlag DO, Jockel KH, Erbel R, Muhleisen TW et al (2014) Aging of blood can be tracked by DNA methylation changes at just three CpG sites. *Genome Biol* 15:R24-2014-15-2-r24. doi: 10.1186/gb-2014-15-2-r24 [doi].

Weiskopf D, Weinberger B, Grubeck-Loebenstein B (2009) The aging of the immune system. *Transpl Int* 22:1041-1050. doi: 10.1111/j.1432-2277.2009.00927.x [doi].

Wertheimer AM, Bennett MS, Park B, Uhrlaub JL, Martinez C, Pulko V, Currier NL, Nikolich-Zugich D, Kaye J, Nikolich-Zugich J (2014) Aging and cytomegalovirus (CMV) infection differentially and jointly affect distinct circulating T cell subsets in humans(). *J Immunol* 192:2143-2155. doi: 10.4049/jimmunol.1301721 [doi].

West J, Beck S, Wang X, Teschendorff AE (2013) An integrative network algorithm identifies age-associated differential methylation interactome hotspots targeting stem-cell differentiation pathways. *Sci Rep* 3:1630. doi: 10.1038/srep01630 [doi].

Westra HJ, Peters MJ, Esko T, Yaghootkar H, Schurmann C, Kettunen J, Christiansen MW, Fairfax BP, Schramm K, Powell JE et al (2013) Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat Genet* 45:1238-1243. doi: 10.1038/ng.2756 [doi].

Xia S, Zhang X, Zheng S, Khanabdali R, Kalionis B, Wu J, Wan W, Tai X (2016) An Update on Inflamm-Aging: Mechanisms, Prevention, and Treatment. *J Immunol Res* 2016:10.1155/2016/8426874. doi: 10.1155/2016/8426874 [doi].

Xie Q, Bai Q, Zou LY, Zhang QY, Zhou Y, Chang H, Yi L, Zhu JD, Mi MT (2014) Genistein inhibits DNA methylation and increases expression of tumor suppressor genes in human breast cancer cells. *Genes Chromosomes Cancer* 53:422-431. doi: 10.1002/gcc.22154 [doi].

Xu G, Wong J (2015) Oxidative DNA demethylation mediated by Tet enzymes. *National Science Review* 2:318-328.

Xu W, Wang F, Yu Z, Xin F (2016) Epigenetics and Cellular Metabolism. *Genetics & Epigenetics* 8:43-51. doi: 10.4137/GEG.S32160.

Xu Z, Taylor JA (2014) Genome-wide age-related DNA methylation changes in blood and other tissues relate to histone modification, expression and cancer. *Carcinogenesis* 35:356-364. doi: 10.1093/carcin/bgt391 [doi].

Yang J, Huang T, Petralia F, Long Q, Zhang B, Argmann C, Zhao Y, Mobbs CV, Schadt EE, Zhu J et al (2015) Synchronized age-related gene expression changes across multiple tissues in human and the link to complex diseases. *Scientific Reports* 5:15145.

Yang SB, Tien AC, Boddupalli G, Xu AW, Jan YN, Jan LY (2012) Rapamycin ameliorates age-dependent obesity associated with increased mTOR signaling in hypothalamic POMC neurons. *Neuron* 75:425-436. doi: 10.1016/j.neuron.2012.03.043 [doi].

Yuan T, Jiao Y, de Jong S, Ophoff RA, Beck S, Teschendorff AE (2015) An integrative multi-scale analysis of the dynamic DNA methylation landscape in aging. *PLoS Genet* 11:e1004996. doi: 10.1371/journal.pgen.1004996 [doi].

Zambelli F, Pesole G, Pavese G (2009) Pscan: finding over-represented transcription factor binding site motifs in sequences from co-regulated or co-expressed genes. *Nucleic Acids Res* 37:W247-52. doi: 10.1093/nar/gkp464 [doi].

Zampieri M, Ciccarone F, Calabrese R, Franceschi C, Burkle A, Caiafa P (2015) Reconfiguration of DNA methylation in aging. *Mech Ageing Dev.* doi: S0047-6374(15)00007-X [pii].

Zhang Y, Hapala J, Brenner H, Wagner W (2017) Individual CpG sites that are associated with age and life expectancy become hypomethylated upon aging. *Clin Epigenetics* 9:10.1186/s13148-017-0315-9. doi: 315 [pii].

Zhang Y, Wilson R, Heiss J, Breitling LP, Saum KU, Schottker B, Holleczeck B, Waldenberger M, Peters A, Brenner H (2017) DNA methylation signatures in peripheral blood strongly predict all-cause mortality. *Nat Commun* 8:14617. doi: 10.1038/ncomms14617 [doi].

Zhang Y, Yang R, Burwinkel B, Breitling LP, Brenner H (2014) F2RL3 methylation as a biomarker of current and lifetime smoking exposures. *Environ Health Perspect* 122:131-137. doi: 10.1289/ehp.1306937 [doi].

## 10 ORIGINAL COMMUNICATIONS

The original publications are reprinted based of the following permissions:

- I** Reprinted under the terms of the Creative Commons Attribution Licence
- II** Reprinted under the terms of the Creative Commons Attribution Licence
- III** Reprinted with kind permission of the American Aging Association
- IV** Reprinted under the terms of the Creative Commons Attribution Licence

RESEARCH ARTICLE

Open Access



# Aging-associated DNA methylation changes in middle-aged individuals: the Young Finns study

L. Kananen<sup>1,2\*</sup>, S. Marttila<sup>1,2</sup>, T. Nevalainen<sup>1,2</sup>, J. Jylhävä<sup>1,2</sup>, N. Mononen<sup>3</sup>, M. Kähönen<sup>4</sup>, O. T. Raitakari<sup>5</sup>, T. Lehtimäki<sup>3,6</sup> and M. Hurme<sup>1,2,3</sup>

## Abstract

**Background:** Chronological aging-associated changes in the human DNA methylome have been studied by multiple epigenome-wide association studies (EWASs). Certain CpG sites have been identified as aging-associated in multiple studies, and the majority of the sites identified in various studies show common features regarding location and direction of the methylation change. However, as a whole, the sets of aging-associated CpGs identified in different studies, even with similar tissues and age ranges, show only limited overlap. In this study, we further explore and characterize CpG sites that show close relationship between their DNA methylation level and chronological age during adulthood and which bear the relationship regardless of blood cell type heterogeneity.

**Results:** In this study, with a multivariable regression model adjusted for cell type heterogeneity, we identified 1202 aging-associated CpG sites (a-CpGs, FDR < 5 %), in whole blood in a population with an especially narrow age range (40 - 49 years). Repeatedly reported a-CpGs located in genes *ELOVL2*, *FHL2*, *PENK* and *KLF14* were also identified. Regions with aging-associated hypermethylation were enriched regarding several gene ontology (GO) terms (especially in the cluster of developmental processes), whereas hypomethylated sites showed no enrichment. The genes with higher numbers of a-CpG hits were more often hypermethylated with advancing age. The comparison analysis revealed that of the 1202 a-CpGs identified in the present study, 987 were identified as differentially methylated also between nonagenarians and young adults in a previous study (The Vitality 90+ study), and importantly, the directions of changes were identical in the previous and in the present study.

**Conclusions:** Here we report that aging-associated DNA methylation features can be identified in a middle-aged population with an age range of only 9 years. A great majority of these sites have been previously reported as aging-associated in a population aged 19 to 90 years. Aging is associated with different types of changes in DNA methylation, clock-like as well as random. We speculate that the a-CpGs identified here in a population with a narrow age-range represent clock-like changes, as they showed concordant methylation behavior in population spanning whole adulthood as well.

**Keywords:** Aging-associated, DNA methylation, EWAS, CpG sites, Adulthood, Hypermethylation, Blood cell type heterogeneity

\* Correspondence: laura.kananen@uta.fi

<sup>1</sup>Department of Microbiology and Immunology, School of Medicine, University of Tampere, Tampere, Finland

<sup>2</sup>Gerontology Research Center, Tampere, Finland

Full list of author information is available at the end of the article



## Background

The epigenome includes DNA methylation (DNAm), post-translational histone modifications and chromatin remodeling. Tens of millions of nucleotides referred to as CpG sites, which are prone to DNAm, exist in the haploid human genome. Furthermore, the genome-wide DNAm profile is maintained through cell divisions. DNA methyltransferases apply methyl groups on CpG sites to form 5-methylcytosine, whereas demethylation may occur either passively due to dysfunction of the transferring enzyme or actively through 5-hydroxymethylcytosine formation. Genomic regions spanning approximately 0.5 kilobases with a high density of CpG sites are called CpG islands, and these are commonly localized near transcription start sites. CpG sites in such islands are often less methylated; thus, the genes are available for initiation of transcription. Moreover, DNAm plays crucial roles in gene expression by not only blocking the promoter region but also altering the activities of regulatory elements, such as enhancers and insulators. Alternatively, gene body methylation may influence alternative splicing [1, 2]. Thus, the cell identity is in part determined and maintained by a cell type-specific genome-wide methylation pattern, which may therefore be used in the laboratory as a marker to characterize the cell types [3–5].

The genome-wide DNAm profile of the cell changes; DNAm patterns are altered in diseases, such as Alzheimer disease, cancer and type 2 diabetes, and are also influenced by the accumulating effects of environmental factors such as toxin exposure and diet [1, 6, 7]. Single CpG sites undergo hypo- and hypermethylation either randomly by stochastic factors or via more systematic mechanisms [1]. For example, exposure to environmental factors such as smoking induces hypomethylation of a well-characterized single CpG site in the gene *F2RL3*; this represents an example of a non-random change in DNAm because the magnitude of the change is dose and exposure-time dependent [8, 9].

Furthermore, the epigenome is modified by the biological aging process. As also Heyn et al. [10] reported and Zampieri et al. [1] reviewed, in general, aging induces a decrease in average DNA methylation level genome-wide (global hypomethylation). This was demonstrated by whole-genome bisulfite sequencing of newborns and centenarians with as high as ~90 % genomic coverage. The comparison of methylation states between the two extremes of the human lifespan also revealed how the systematic methylation patterns of the CpG sites are eventually lost and how inter-individual differences increase with advanced age. In addition, hypermethylation in regions near promoters can cause down-regulation of essential genes that influence vitally important pathways; Heyn et al. [10] reported that aging-accelerated hypermethylation events occurred in

13 % of the CpG sites among the millions of sites in the genome. Therefore, methylation alterations may be considered as one important factor in the development of aging-associated diseases [1, 10].

Many studies have addressed the aging-associated DNAm changes in blood cells using Illumina array technology-based methods, which cover 27000 or 485000 CpG sites in the genome [1]. The methylation levels of specific CpG sites are known to be associated with chronological aging in a wide variety of tissues [11–13]. However, as a whole, the sets of aging-associated CpGs identified in different studies, even with comparable tissues and age ranges, show limited overlap. Only few EWASs on age have taken the cell type heterogeneity into account [14–17]. We and others [4] hypothesize that lack of cell type adjustment may have potentially distorted the results obtained, and this may have contributed to the lack of concordance observed between the studies.

In this study, we aimed to discover and characterize regions where the DNAm levels are associated with chronological age (a-CpGs) in a middle-aged population (aged 40–49 years) through analysis where the cell type heterogeneity was adjusted for. Middle-aged individuals were selected from the Young Finns Study (YFS) [18] follow-up in 2011; the selection in the present study is a balanced sample (i.e. the number of subjects in each age group was equal and the groups had similar sex-distribution), and it therefore provides an excellent opportunity to inspect the effects of aging on DNA methylome. Furthermore, this sample comprises individuals in an extremely narrow age range of only nine years. The subjects' DNA methylomes were characterized using Illumina Infinium HumanMethylation450 BeadChips and the cell type heterogeneity and sex were adjusted for in the analysis.

Additionally, our findings were interpreted together with compatible data obtained using the same 450BeadChip technology, including our previous results obtained from an EWAS on age (The Vitality 90 + Study, V90+), in which the subjects' ages ranged from 19 to 90 years [15], as well as other results compiled by Steegenga et al. [19]. The results from the YFS were interpreted by considering that rates of aging-associated DNAm changes fluctuate, especially during the growth period before adulthood and at the end of the lifespan [11, 20]. Accordingly, the a-CpGs found in the YFS that overlap with those established from adult samples with wider age ranges, such as V90+ study, may be speculated to be DNAm regions with constant rate of change throughout adulthood. Thus, we aimed to explore the a-CpGs where level of methylation changes in a clocklike fashion throughout adulthood from those that show a more random aging-associated pattern.

## Results

### Aging-associated alterations in DNA methylation

In this study, the genome-wide DNAm levels in whole blood samples of middle-aged individuals were measured using 450BeadChip technology. The sample heterogeneity (i.e., the proportions of CD8T and CD4T cells, monocytes, granulocytes, and NK and B cells) were estimated by comparing DNAm profiles to the reference dataset [4] (Additional file 1: Figure S1). The cell type proportions were verified as important determinants of variation in DNAm using Spearman's correlation analysis, in which the cell type proportions were correlated with the main principal components (PCs). The PCs were defined with principal component analysis (PCA) from the DNAm data without cell subtype adjustment (Additional file 2: Table S1a). The analysis revealed that PC1 to PC6 together explained a large proportion (24 %) of the variance in the DNA methylome data. Among those PCs, several PCs had considerable large ( $-0.5 > r > 0.5$ ) correlation coefficients; thus, adjustments for the cell type proportion in the analysis were mandatory. The hypothesis whether DNAm level of a CpG site is associated with chronological age was tested at each CpG

site using generalized linear regression analysis ('beta regression'), where sex and cell type proportions were adjusted for.

We found 1202 a-CpGs (i.e. CpG sites where age was a statistically significant variable in the multivariable regression model,  $FDR < 5\%$ ) in middle-aged individuals (aged 40–49 years), of which 622 (52 %) were hypomethylated and 580 (48 %) were hypermethylated with advancing age. These hypo- and hypermethylated sites were annotated on 440 and 437 genes, respectively. Lists of the most significant aging-associations in YFS are shown in Tables 1 and 2 and in Additional file 3: Table S4. Frequently reported CpG sites (summarized by Steegenga et al. [19]) located in the *ELOVL2* (cg16867657, cg24724428 and cg21572722), three sites in the *FHL2* (cg06639320, cg22454769 and cg24079702), two sites in the *PENK* (cg16219603, cg16419235), and two sites in the *KLF14* (cg08097417, cg09499629 and cg07955995) were also identified as hypermethylated in the present study.

Interestingly, similar to correlation analysis results shown in Additional file 2: Table S1a, the cell type proportions were important determinants of variation in

**Table 1** The top 20 hypermethylated a-CpGs in middle-aged individuals. The hypermethylated and hypomethylated a-CpGs are shown separately in Tables 1 and 2, respectively. The top-ranking hypermethylated a-CpGs were selected with the following criteria: 1) direction of the association based on the value of beta regression (denoted as 'betareg') estimate of age; 2) more than one hit identified per gene ( $q$ -value  $< 0.05$  which corresponds to false discovery rate  $< 5\%$ ) and 3) the top-ranking  $p$ -values. The full list of a-CpGs is shown in Additional file 3: Table S4. The  $q$ -value denotes the Benjamini-Hochberg-corrected  $p$ -value

ProbelID	Gene name	CHR	Coordinate	Betareg estimate of age	$q$ -value
cg16867657	<i>ELOVL2</i>	6	11152863	0.022	0.00E+00
cg24724428	<i>ELOVL2</i>	6	11152874	0.021	4.80E-07
cg21572722	<i>ELOVL2</i>	6	11152880	0.013	3.46E-06
cg06639320	<i>FHL2</i>	2	105382171	0.018	3.46E-06
cg00059225	<i>GLRA1</i>	5	151284550	0.013	5.13E-06
cg08097417	<i>KLF14</i>	7	130069673	0.020	1.87E-05
cg22454769	<i>FHL2</i>	2	105382199	0.021	5.03E-05
cg07553761	<i>TRIM59</i>	3	161650671	0.016	6.12E-05
cg01588592	<i>ETV3L</i>	1	155335949	0.011	1.14E-04
cg11176990	<i>LOC375196</i>	2	39041037	0.014	1.54E-03
cg09499629	<i>KLF14</i>	7	130069676	0.018	1.54E-03
cg22158769	<i>LOC375196</i>	2	39041043	0.020	2.43E-03
cg18898125	<i>NEFM</i>	8	24826286	0.012	2.49E-03
cg21911021	<i>ZIK1</i>	19	62786823	0.020	3.07E-03
cg27217742	<i>RGS12</i>	4	3335078	0.013	3.07E-03
cg17737681	<i>DLX1</i>	2	172660382	0.015	3.29E-03
cg24079702	<i>FHL2</i>	2	105382203	0.015	5.99E-03
cg16219603	<i>PENK</i>	8	57523140	0.013	7.00E-03
cg23930856	<i>TFAP2B</i>	6	50919683	0.013	7.22E-03
cg11152943	<i>TRAPPC9</i>	8	141318170	0.013	7.57E-03

**Table 2** The top 20 hypomethylated a-CpGs in middle-aged individuals. The hypermethylated and hypomethylated a-CpGs are shown separately in Tables 1 and 2, respectively. The top-ranking hypomethylated a-CpGs were selected with the following criteria: 1) direction of the association based on the value of beta regression (denoted as 'betareg') estimate of age; 2) more than one hit identified per gene (q-value < 0.05 which corresponds to false discovery rate < 5 %) and 3) the top-ranking p-values. The full list of a-CpGs is shown in Additional file 3: Table S4. The q-value denotes the Benjamini-Hochberg-corrected p-value

ProbelID	Gene name	CHR	Coordinate	Betareg estimate of age	q-value
cg00791074	<i>MTHFD1L</i>	6	151227862	-0.018	7.51E-04
cg18618815	<i>COL1A1</i>	17	45630323	-0.018	5.99E-03
cg14169886	<i>PRDM16</i>	1	3101709	-0.014	5.99E-03
cg01820374	<i>LAG3</i>	12	6752344	-0.014	9.24E-03
cg19421125	<i>LAG3</i>	12	6753117	-0.022	1.02E-02
cg14829066	<i>NTRK3</i>	15	86360145	-0.013	1.49E-02
cg03290281	<i>C6orf195</i>	6	2577606	-0.021	1.49E-02
cg05561193	<i>DCLK2</i>	4	151218492	-0.017	1.96E-02
cg20249566	<i>NWD1</i>	19	16691739	-0.024	1.97E-02
cg23928726	<i>PEX10</i>	1	2334858	-0.014	1.97E-02
cg20007894	<i>SCAND3</i>	6	28648421	-0.019	2.08E-02
cg16355231	<i>PEX10</i>	1	2334839	-0.019	2.14E-02
cg15058210	<i>HDAC4</i>	2	239861814	-0.018	2.16E-02
cg06030846	<i>TMEM108</i>	3	134581182	-0.011	2.16E-02
cg25994988	<i>UBASH3B</i>	11	122157592	-0.011	2.16E-02
cg18345924	<i>NCAM2</i>	21	21294102	-0.016	2.18E-02
cg00638021	<i>COL1A1</i>	17	45622061	-0.013	2.26E-02
cg19344626	<i>NWD1</i>	19	16691749	-0.024	2.36E-02
cg01288258	<i>ITFG2</i>	12	2792128	-0.011	2.41E-02
cg05221385	<i>TAF10</i>	11	6590080	-0.010	2.43E-02

DNAmet levels of the 1202 a-CpGs as well (Additional file 2: Table S1b). In this second correlation analysis, the PCs were defined with PCA from DNA methylation data of the 1202 a-CpGs (aging-associated CpG sites, FDR < 5 %); methylation data in PCA were not adjusted for the cell subtype heterogeneity. Correlation analysis revealed that PC1-PC6 determined more than 50 % of variance in methylation levels of these a-CpGs and these PCs correlated clearly with age and the cell counts. It is also worth of mentioning that of the 1202 a-CpGs in our initial aging-association analysis, there were 526 multivariable regression models (corresponding 526 CpG sites) where all cell count variables (monocytes, granulocytes, NK, CD8T and CD4T cells) were detected as statistically significant (FDR < 5 %) predictors of DNA methylation levels.

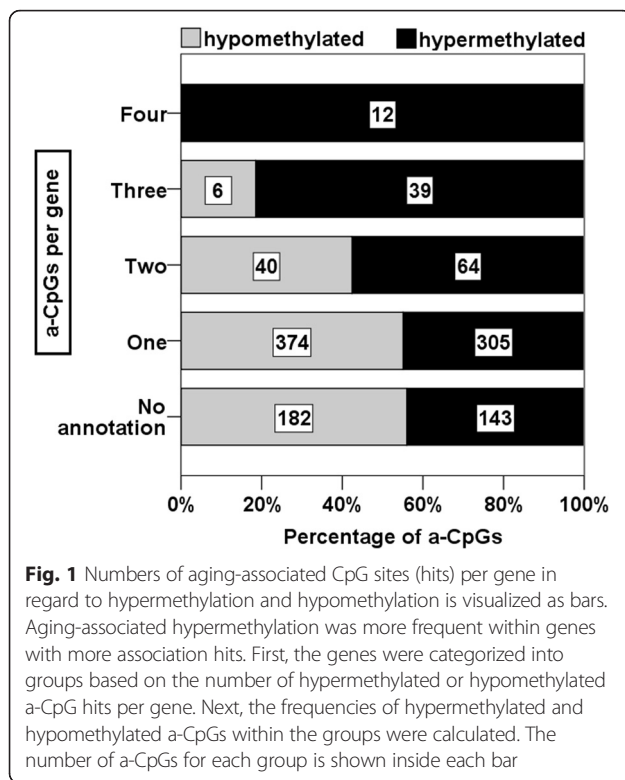
The importance of the cell count considerations was explored with an additional set of regression models, where the DNA methylation level in each CpG site genome-wide was explained with age and sex only while the cell counts were not adjusted for. In this analysis, only 56 sites were classified as aging-associated (FDR < 5 %) and these sites were all included to the original

pool of 1202 a-CpGs. The 56 a-CpGs are pointed out in the Additional file 3: Table S4.

Aging-associated hypermethylation and hypomethylation differ in their features. The exploration of aging-associations in the YFS revealed that hypermethylation was more frequent within genes with more association hits as shown in Additional file 4: Table S5 and Fig. 1). Specifically, there were 70 genes in total either with more than one hypomethylated or more than one hypermethylated a-CpGs per gene. Of those, 22 genes comprised more than one hypomethylated a-CpGs per gene and 48 genes comprised more than one hypermethylated a-CpGs per gene as shown in Additional file 4: Table S5.

Next, the genomic locations of the a-CpGs were investigated, revealing that 388 of the 1202 a-CpGs were located on CpG islands rather than island shores, shelves or non-island regions, and a majority (N = 331) of those were hypermethylated (Additional file 1: Figure S2). The remaining sites were distributed to shores, shelves and non-island regions with opposite manner as shown in Additional file 1: Figure S2; the aging-associated hypomethylation was more abundant on those regions. The a-CpG locations on genes were also investigated; no





enrichment of a-CpGs was detected in the regions of 3' untranslated regions (UTRs), 5'UTRs or close distances to transcription start sites or gene bodies (Additional file 1: Figure S3a and b). The distributions of the a-CpGs on chromosomes were also investigated; hypermethylated a-CpGs were over-represented on chromosome 18, whereas hypomethylated sites were not enriched on any chromosome (hypergeometric test, nominal  $p$ -value of 0.05) (Additional file 1: Figure S3c). In addition, we ensured using visual examination that there were no spatial local cluster(s) of a-CpGs on Chr-18.

#### Sex specificity of the aging-associated CpG sites

To evaluate the sex specificity of the aging-associations, an interaction model with variables corresponding to sex, age and the interaction of sex and age (age\*sex) was constructed. No sex-specific a-CpGs were identified, as analysis revealed that no interaction term had a false discovery rate (FDR) below 5 % ( $q$ -value < 0.05) in the interaction models. Furthermore, we analyzed women ( $N = 111$ ) and men ( $N = 73$ ) separately as well: sex-specific a-CpGs were explored among all CpG sites with an multivariable regression model ('beta regression') where age and cell type proportion variables were used to predict DNA methylation level in each CpG site. These analyses revealed that there were 105 and 173 a-CpGs (FDR < 5 %) among men and women, respectively; these CpG sites were all included to our original pool of

1202 a-CpGs which were detected using whole sample ( $N = 184$ ). Importantly, as shown in Additional file 1: Figure S5, when the directions of change among the 1202 a-CpGs were cross-compared between men and women (without  $p$ -value cut-off), all sites, except one, showed concordant behavior regarding hypermethylation or hypomethylation during aging (i.e. whether the estimate of age variable in the regression model was negative or positive value). This behavior was also identical to the directions of change among the 1202 a-CpGs in the initial analysis ( $N = 184$ ). As a conclusion, these results were in line with our interaction analysis: there were no significantly sex-specific a-CpGs among middle-aged individuals.

#### Functional roles of a-CpGs in the YFS

The gene ontology (GO) functions and processes of the genes with a-CpGs were investigated using the Gene Ontology enRiChment anaLysis and visualiZation (GORilla) tool [21]. The analysis was conducted separately for genes with hypermethylated a-CpGs and for hypomethylated a-CpGs ( $N = 440$  and  $N = 437$ , respectively). The analysis revealed an unambiguous differences between hypo- and hypermethylated a-CpGs, as 73 GO process terms and to 8 GO function terms were enriched to genes with hypermethylated a-CpGs (Tables 3 and 4, respectively; Additional file 2: Table S2.), whereas there was no enrichment of terms among the genes with hypomethylated a-CpGs (Bonferroni-adjusted  $p$ -value threshold of 0.05). The most statistically significant processes were anatomical structure development (GO:0048856,  $p = 1.02 \times 10^{-11}$ ) and morphogenesis (GO:0009653,  $p = 5.02 \times 10^{-10}$ ), both of which cluster under the term 'developmental process'.

In addition, Pscan [22] was used to predict whether there were common regulators for groups of genes. The hypermethylation-associated genes were predicted to be regulated by 11 common transcription factors (Additional file 2: Table S3), several of which were zinc coordinating. For hypomethylation-associated genes, no common transcription factors were found. A large proportion of the 11 regulators of genes with hypermethylated a-CpGs in the YFS were zinc coordinating, and four (E2F1, EGR1, SP1, TFAP2A) were identical to those identified in the V90+ study [15].

#### Comparisons to other studies

In the explorative cross-comparison analysis, the a-CpGs identified in middle-aged individuals of the YFS were compared to aging-associated DNA methylome alterations between nonagenarians and 19–30-year-old individuals evidenced in our previous study (the V90+ study) [15]. The a-CpGs identified in the V90+ study were strongly associated with aging while the cell type heterogeneity was adjusted for in the analysis. A total of

**Table 3** Several GO process terms were enriched within genes with hypermethylated a-CpGs in the analysis with GOrilla [21, 43]. This table represents the main clusters of processes (53 redundant GO terms were filtered out of 73 terms using REVIGO [44]). The full list of processes is shown in Additional file 2: Table S2

GO term	Description of the process	<i>p</i> -value (-log10)
GO:0048856	Anatomical structure development	10.9914
GO:0050794	Regulation of cellular process	8.9788
GO:0007389	Pattern specification process	8.2343
GO:0032502	Developmental process	8.2041
GO:0009893	Positive regulation of metabolic process	8.0511
GO:0044708	Single-organism behavior	7.5544
GO:0035108	Limb morphogenesis	7.5544
GO:0003002	Regionalization	7.3585
GO:0051239	Regulation of multicellular organismal process	7.301
GO:0006357	Regulation of transcription from RNA polymerase II promoter	7.2248
GO:0065007	Biological regulation	7.1675
GO:0007610	Behavior	7.08
GO:0048598	Embryonic morphogenesis	7.0778
GO:0048518	Positive regulation of biological process	6.8761
GO:0048519	Negative regulation of biological process	6.7122
GO:0008285	Negative regulation of cell proliferation	6.4921
GO:0048523	Negative regulation of cellular process	5.8827
GO:0010842	Retina layer formation	5.8041
GO:0051961	Negative regulation of nervous system development	5.7423
GO:0032774	RNA biosynthetic process	5.4225

the 1202 a-CpGs established in the YFS cohort, 999 a-CpGs were also aging-associated in the V90+ sample (FDR < 5 %, Additional file 3: Table S4). Of these 999 a-CpGs, 464 (46 %) were hypermethylated, and 535 (54 %) were hypomethylated with advancing age. Furthermore, in 987 of the overlapping 999 a-CpGs the direction of the aging-associated change was the same: in the present and

in the V90+ study, 455 a-CpGs were hypermethylated, and 532 were hypomethylated with advancing age (Fig. 2).

Finally, a-CpGs that were characterized from whole blood samples as aging-associated using 450 BeadChip technology and previously reported by Hannum et al. (number of hits, 89) [13], Garagnani et al. (number of hits, 9) [12] and Florath et al. (number of hits, 162) [23] and presented as summary table in Steegenga et al [19] were further compared with our data. The corresponding age of the samples ranged between 19–101, 9–83 and 50–75 years, respectively. The comparison revealed 21 common CpG sites out of the 999 a-CpGs in two or more studies in addition to the YFS and the V90+ study (Fig. 3).

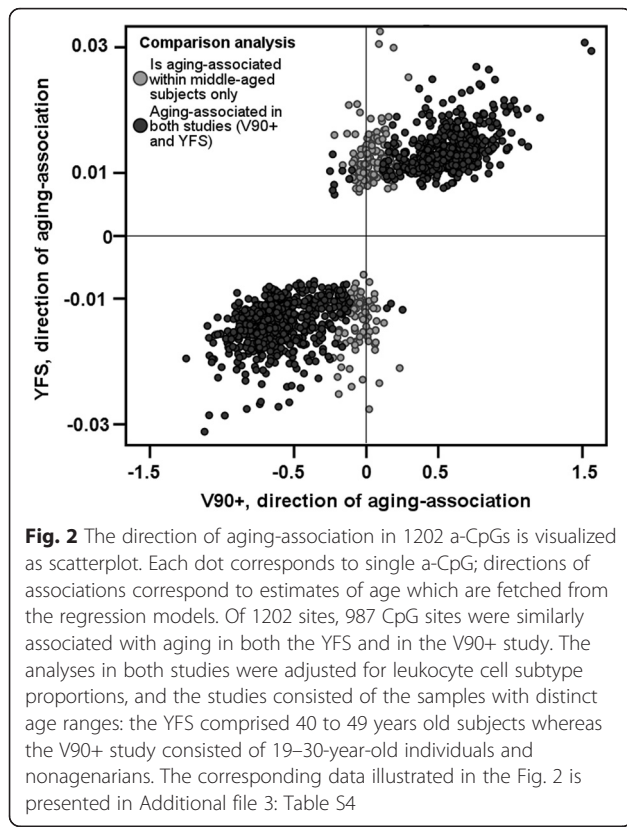
## Discussion

In this study, we identified 1202 a-CpGs where the DNAm level was associated with aging in middle-aged individuals (i.e. with an age range of 40 to 49 years), in whom the growth and development of youth has ended yet old age and its associated diseases had not begun. Of the 1202 a-CpGs, 622 (52 %) were hypomethylated, and 580 (48 %) were hypermethylated with advancing age, with annotations on 440 and 437 different genes, respectively. In general, the functional features of these aging-associated sites are mostly similar to those identified from cohorts with larger age differences. Our study highlights also that a large number of sites undergo aging-associated DNAm level changes throughout adulthood and we speculate that a great proportion of those probably change with a clock-like manner.

A large fraction of the DNAm sites are altered during the lifespan, as shown by previous studies performed using 450BeadChip technology [15, 24] and whole-genome bisulfite sequencing [10]. Furthermore, the rates of these changes may fluctuate at different stages of the lifespan. Studies have shown that a-CpGs behave differently during the growth period before adulthood and at the end of the lifespan [11, 20]. Nonetheless, there are genes (*ELOVL2*, *SFMBT1*, *KLF14*, *PENK*, and *FHL2*) with CpG sites that are consistently detected as being aging-associated despite of differences in sample tissue

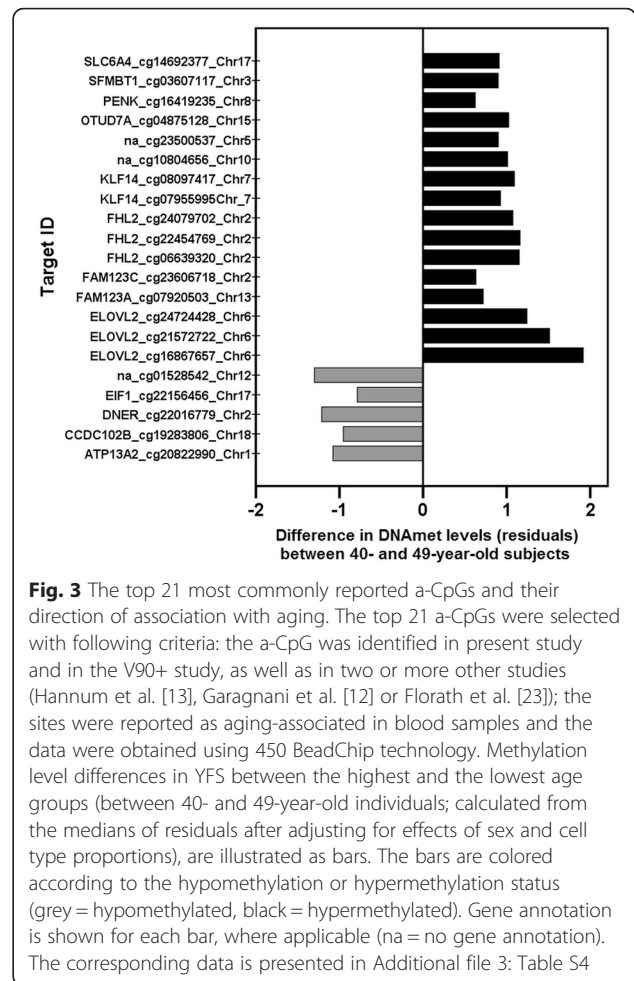
**Table 4** GO function terms were enriched within genes with hypermethylated a-CpGs in the analysis with GOrilla. Table contains the full list of enriched GO function terms (Bonferroni-adjusted  $p < 0.05$ ) obtained from analysis with GOrilla [21, 43]

GO term	Description of the function	<i>p</i> -value (-log10)
GO:0043565	Sequence-specific DNA binding	10.001
GO:0000981	Sequence-specific DNA binding RNA polymerase II transcription factor activity	7.322
GO:0001071	Nucleic acid binding transcription factor activity	6.721
GO:0003700	Sequence-specific DNA binding transcription factor activity	6.721
GO:0003677	DNA binding	6.625
GO:0005326	Neurotransmitter transporter activity	5.148
GO:0005488	Binding	4.967



types or age distributions [11–13, 15]; notably, these genes were also identified in the present study as being aging-associated (Tables 1 and 2; Additional file 3: Table S4). However, a recent meta-analysis on three DNAm levels data sets obtained using 450BeadChip illustrated discrepancies in the lists of regions where DNAm levels were altered during the entire human lifespan, ranging from 0 to 100 years of age [2]. Because blood sample heterogeneity has been shown to have a great impact on EWASs [4, 15], our speculation is that the discrepancies might be due to the presence of different cell types.

In the primary analysis, we aimed to identify a-CpGs in middle-aged individuals representing general population with age range of only one decade. Then, we cross-compared the results to those obtained with similar analysis pipeline from a population aged 19 to 90 years (Vitality 90+ study) [15]. Among the 1202 a-CpGs characterized from the YFS with an age range of nine years, 987 sites had an identical association direction as detected in the Vitality 90+ study, as shown in Fig. 2 and in Additional file 3: Table S4. We hypothesize that sites displaying aging-associated methylation changes in both populations possibly represent sites where the change in DNA methylation follows a clock-like pattern. We further speculate that the non-overlapping CpG sites identified in the population with a wider age range (19 to 90 years of age) may possibly represent



sites where the aging-associated change is accelerated in either early or late adulthood; the a-CpGs identified only when comparing group of nonagenarians to young adults may represent changes that reflect e.g. aging-associated pathologies or accumulation of aging-associated impairments.

As aging influences the immune system of men and women differently and as the risk rates of several diseases between sexes are unequal [25, 26], 1) an interaction analysis was performed to address the sex specificity of a-CpGs, and 2) the aging-associations were also evaluated in separate analyses among men and women. These analyses revealed no sex-specific single a-CpGs; thus, the identified a-CpGs are universally altered in both men and women. These results are in accordance with our previous results from the V90+ study, in which the DNAm states of nonagenarians were compared with 19–30-year-old individuals [15], and with results published by others [24, 27]. However, studies have shown that as a whole, the DNA methylomes of males age more rapidly than those of females [13, 28].

Aging-accelerated hypomethylation may be thought as an erosion-like event, whereas hypermethylation may be thought as an actively guided process. In practice, the difference between these features is manifested, for example, through the enrichment of GO terms for groups of genes and for signaling pathways [1, 15]. The distinct roles of the methylation status were demonstrated in the present study with the numbers of a-CpG hits in a gene, as we observed notable enrichment of hypermethylation events located in genes with more than one a-CpG (Fig. 1). The functional roles of genes with a-CpGs were established by GO term enrichment analysis, which revealed obvious difference between hypo- and hypermethylated a-CpGs, even though the analysis was conducted with an equal number of genes in the GO term analyses. A high number of GO terms were enriched to genes with hypermethylated a-CpGs (Tables 3 and 4; Additional file 2: Table S2), whereas there was no GO term enrichment within genes with hypomethylated a-CpGs. The most statistically significant processes enriched to genes with hypermethylated a-CpGs were 'anatomical structure development' and 'morphogenesis', both of which cluster under the term 'developmental process'. The enrichment of hypermethylated a-CpGs to these processes has been reported previously [14, 23, 24, 29]. Reynolds [30] and Yuan [16] reported also that the CpG sites hypermethylated during aging are enriched to common processes and exhibit shared features, whereas hypomethylated a-CpGs are a less homogenous group. Furthermore, age-associated hypermethylation interactome hotspots have been reported [31].

In addition to the details mentioned above, we observed other similar hypermethylation characteristics in the YFS, as those reported in previous studies [1, 15]. For example, the majority (85 % out of 388) of a-CpGs localized in CpG-islands (instead of shores, shelves or other regions) were hypermethylated, and an excess of hypermethylated a-CpGs were also found on chromosome 18. However, there was no enrichment of a-CpGs on chromosome 19. In the V90+ study, the hypermethylated a-CpGs located in the genes encoding zinc-associated proteins were more abundant on chromosome 19 [15], where zinc-finger genes are clustered. The zinc-finger genes (such as *ZNF154*) located in chromosome 19 are proposed to be repressors of endogenous retroviruses (ERVs) [32], and the repressor activity may be disturbed by hypermethylation. Interestingly, CpG sites located in the gene *ZNF154* and almost all other genes encoding zinc-fingers on chromosome 19 were absent from our pool of 1202 a-CpGs. Thus, as the hypermethylation of CpG sites located in genes encoding zinc-fingers was observed in the oldest age group, we hypothesize that rates of methylation level changes at the CpG sites located in ERV repressor genes

(e.g. *ZNF154*) may fluctuate throughout the lifespan and that the rates may be enhanced in association with other senescence-related factors. Therefore, it is possible that DNAm-based dysfunction of the repression system might explain the increased expression of ERVs in old age [33]. Future studies are required to address these questions.

To further inspect the roles of the genes with aging-accelerated DNAm changes, analysis of the common regulators (transcription factors) of groups of genes with hypermethylated and hypomethylated a-CpGs was conducted with Pscan [22]. The results were again surprisingly concordant with those in the V90+ study. There were 11 regulators with unique identifiers for hypermethylated a-CpGs (Additional file 2: Table S3), whereas hypomethylated a-CpGs had no common regulators. A great proportion of the 11 regulators of genes with hypermethylated a-CpGs in the YFS were zinc coordinating, and four (E2F1, EGR1, SP1, and TFAP2A) were identical to those identified in the V90+ study results [15]. Overall, the results from analysis of the functional roles of the genes with a-CpGs were surprisingly well in line with the observations from the V90+ study and supported the proposition that aging-associated hypermethylation is a more tightly regulated process, whereas aging-associated hypomethylation is induced more by environmental effects and stochastic factors.

Finally, we demonstrated the lack of concordance in previously reported pools of a-CpGs by comparing three published lists of overlapping a-CpGs produced using 450BeadChips from whole blood samples from subjects with age ranges of 50–75, 19–101 and 9–83 [12, 13, 23]. Although 987 of the a-CpGs in the YFS showed similar association directions as in the V90+ study (Fig. 2 and Additional file 3: Table S4), we observed only 61 overlapping a-CpGs in the YFS and the V90+ study, which were also reported as aging-associated in one or more other robustly compatible studies (same sample type and array technology). Of these, only 21 a-CpGs were observed in two or more of the studies in the comparison (Fig. 3). To the best of our knowledge [4, 15], the main factor that contributes to the DNAm profiles in blood cells is cell type heterogeneity; thus, we speculate that the lack of cell type adjustments may account for the majority of disparity in the cross-comparisons. The results of aging-association analysis and combined PCA-correlation analysis in this study supports our speculation. Cell type heterogeneity should be taken into account when analyzing samples composed of mixed cell types, but a limited number of such studies have been conducted [4, 14–17].

Notably, our study had an obvious limitation, it would substantially benefit from being a follow-up; therefore, future studies are needed. Nevertheless, the analysis is

powered by well-designed sample characteristics because each age group was matched by sex and sample size and because adjustments were made for cell type heterogeneity. Thus, the analysis was sensitive enough to detect DNAm changes within an age range spanning nine years.

## Conclusions

Here we report that aging-associated DNA methylation changes can be identified in a middle-aged population with a narrow age range of 9 years. Aging-associated DNAm changes are not uniform, but occur due to different reasons, at different rates and directions in different parts of the genome and are not alike in all cell types. Thus, due to this diverse nature of aging-associated DNA methylation changes, all confounding factors should be accounted for in the analysis, in order to obtain comparable results. Our results support the notion that cell type heterogeneity should be adjusted for when analyzing tissues consisting of mixed cell types. Moreover, our results imply that considerable proportion of DNAm changes show clock-like behavior throughout adulthood.

## Methods

### Study population

The Young Finns study (YFS) comprises a series of six cohorts, representing general population, born in 1962, 1965, 1968, 1971, 1974 and 1977 from five cities with university hospitals in Finland (Helsinki, Kuopio, Oulu, Tampere and Turku) [18]. A subsample of 184 individuals was randomly assigned from a follow-up in 2011. The sample collection in 2011 is described in more detail elsewhere [34]. The categories of age in the methylation analysis were 40, 43, 46 and 49 years old, with group sizes of 50, 44, 55 and 35, in which 58 %, 68.2, 56.4 and 60 % were women, respectively. All of the participants were of western European descent. The study followed the guidelines of the Declaration of Helsinki and was approved by the Ethical Review Committee of Turku University Hospital. All participants provided informed consent.

### DNA methylome quantification

#### Sample preparations

Leukocyte DNA of the YFS cohort was obtained from EDTA-blood samples using a Wizard® Genomic DNA Purification Kit (Promega Corporation, Madison, WI, USA) according to the manufacturer's instructions. Genome-wide DNA methylation levels were obtained using Illumina Infinium HumanMethylation450 BeadChips [35–37] in the Core Facility at the Institute of Molecular Medicine Finland (FIMM), University of Helsinki according to the protocol by Illumina.

The methylation data set was preprocessed identically with a previously described analysis pipeline which was used in the DNA methylation analysis of the V90+ study samples [15, 38, 39]. Briefly, methylation signal data was preprocessed as a methylumi object using R software ( $R \geq 2.15.3$ ) with array-specific algorithms implemented in the R package `wateRmelon` [40] and `BMIQ` [38]. The resulting  $\beta$  values ranged linearly from 0 (non-methylated, 0 %) to 1 (completely methylated, 100 %). The quality of DNA samples and methylation data was carefully ensured by standard examinations with principal component analysis (PCA) and visualizations with density plots, boxplots and dotplots. Three of the YFS samples were excluded due to atypically low probe intensities compared with control probe intensities.

The YFS sample was lacking leukocyte cell type characterizations; thus, the proportions were determined by the estimation algorithm implemented in the `estimateCellCounts` function of the `minfi` Bioconductor package [4] using R software ( $R \geq 2.15.3$ ). The algorithm utilizes the selection of 600 control probes that represents specific signatures of CD8T and CD4T cells, monocytes, granulocytes, and NK and B cells (Additional file 1: Figure S1). The reference data used in the estimation is available in the `FlowSorted.Blood.450K` Bioconductor package [4].

### Quality control of the DNA methylome data

As the cell type proportions contribute to most of the variation in genome-wide DNAm [4, 15], the significance of the estimated cell counts in the DNAm data was investigated by PCA, and the main PCs of DNAm were correlated with the cell counts (Additional file 2: Table S1a). Spearman's correlation analysis indicated a clear connection between methylation profiles and estimated cell proportions. Thus, the estimated cell counts as well as the genome-wide methylation data was shown to behave as expected.

As part of the quality control step, a well-known CpG site with phenotype association was selected. Smoking is strongly associated with the hypomethylation of cg03636183, located in the gene *F2RL3* [8, 9]; our data from the YFS replicated this finding, as we observed a difference between daily smokers and others (Wilcoxon rank sum-test,  $P = 2.4 \times 10^{-6}$ ; Additional file 1: Figure S4). Analysis with multivariable regression model (function `lm()` in R) revealed that the cell type heterogeneity, age or sex of the samples did not alter the finding of cg0363618.

### Detection of aging-associated methylation regions

Aging-associated CpG sites, the a-CpGs, were explored using a generalized linear regression model, referred to as the 'variable dispersion beta regression' in an iterative

manner for each methylation locus (CpG site). The age (categories of 40, 43, 46 and 49) was employed as a variable to predict the site-specific methylation outcome in the form of a  $\beta$  value (ranging from 0 to 1); this was done in each equation using the mean model and a linker function of *logit*. The cellular heterogeneity was adjusted in the initial multivariable regression analyses: in addition to age and sex variables, variables corresponding to each estimated blood cell subtype proportion (CD8T and CD4T cells, monocytes, granulocytes, NK and B cells; all ranging linearly from 0 to 1) were included to the regression models as predictors of DNA methylation level. Additionally, sex-specific a-CpGs were explored among all CpG sites using two approaches: 1) with an interaction model where age, sex, sex\*age and cell type proportion variables were used to predict DNA methylation level, and 2) with a regression model where age and cell type proportion variables were used to predict DNA methylation level separately for men and women. Furthermore, to explore the relevance of the cell count considerations in the regression analyses, an additional set of age-association analyses was performed. In these regression models, the DNA methylation level of each CpG site was explained with age and gender variables only and the cell proportions were not adjusted for. The analyses were performed using R software ( $R > 2.15.3$ ), and the regression analyses were mainly conducted with algorithms implemented in the betareg package [41]. The nominal Benjamini-Hochberg adjusted  $p$ -value ( $q$ -value) was set to 0.05. The a-CpGs were annotated based on the assembly provided by the R package, FDb.InfiniumMethylation.hg19 [42]. For the purpose of visualization in Fig. 3, standardized weighted residual values of the methylation levels were extracted for each CpG site from regression models in which only sex and cell type proportion variables were set as predictors.

#### Analysis of the functional roles of a-CpGs

The enriched gene ontology (GO) terms of the genes with a-CpGs were discovered using GOrilla [21, 43], and the significant terms were further clustered by REVIGO [44]. The GOrilla analysis was performed for the process, function and component categories with two un-ranked lists, of which the first list comprised genes with hypomethylated or hypermethylated a-CpGs (Additional file 3: Table S4), and the second comprised the genes in the background ( $N = 20,902$ ; analysis date, 9.3.2015). Furthermore, the prediction of common transcription factors of the groups of genes with either hypermethylated or hypomethylated a-CpGs (as two separate analyses) was conducted using Pscan with the default settings (JASPAR database; analysis date, 10.3.2015) [22]. The nominal

$p$ -value was set to at the Bonferroni-corrected value of 0.05 in each analysis.

#### Availability of supporting data

The methylation data presented in this manuscript have been submitted to the Gene Expression Omnibus (GEO) database (<http://www.ncbi.nlm.nih.gov/geo/>) under the accession number GSE69270.

#### Additional files

**Additional file 1: Figures S1-S5.** 1) A figure of estimated proportions of CD8T, CD4T, NK, B cell, monocyte and granulocyte cells of peripheral blood samples in YFS. Proportions are visualized as boxplots, categorized by age group and organized to separate panels by sex. 2) A figure of aging-associated CpG site locations in regard to CpG islands (CGIs). Number of aging-associated CpG sites are visualized with stacked bars. 3) A figure (a-c) presenting locations of a-CpGs. 4) A figure showing results for association of DNA methylation level in cg03636183 with smoking. 5) A figure presenting sex specificity of the aging-associated CpG sites (a-CpGs). (DOCX 357 kb)

**Additional file 2: Tables S1-S3.** 1) Two summary tables (a and b) of the results from Spearman correlation analyses between age, the cell counts and the first principal components (PCs). PCs were defined from either the whole methylation data or 1202 a-CpGs using PCA. 2) A table of the GO terms of the bio processes that are enriched to genes with aging-associated CpG-sites. 3) A table of common transcription factors for genes with hypermethylated a-CpGs characterized using Pscan. (DOCX 31 kb)

**Additional file 3: Table S4.** A full table of 1202 a-CpGs with detailed information. (XLSX 183 kb)

**Additional file 4: Table S5.** A summary table where the 70 genes with more than one hypomethylated or more than one hypermethylated a-CpGs per gene are presented. (XLSX 15 kb)

#### Abbreviations

a-CpGs: aging-associated CpG sites (FDR < 5 %); DNAm: DNA methylation; ERV: endogenous retrovirus; EWAS: epigenome-wide association studies; PC: principal component; UTR: untranslated region; V90+: The Vitality 90+ Study; YFS: Young Finns study.

#### Competing interests

The authors declare that they have no competing interests.

#### Authors' contributions

LK processed the data and performed analyses and was responsible for writing the manuscript. NM, JJ and TN were responsible of the experiments. MH, MK and TL provided reagents and materials and LK, SM, JJ, TL and MH contributed to the design of the study. OTR, MK and TL were responsible for recruiting the subjects in the study. All authors contributed to the writing of manuscript and read and approved the final manuscript.

#### Acknowledgements

The Young Finns Study has been financially supported by the Academy of Finland: grants 286284 (T.L.), 132704 (M.H.), 134309 (Eye), 126925, 121584, 124282, 129378 (Salve), 117787 (Gendi), and 41071 (Skidi); the Social Insurance Institution of Finland; Kuopio, Tampere and Turku University Hospital Medical Funds (grant X51001 for T.L.); Juho Vainio Foundation; Paavo Nurmi Foundation; Finnish Foundation of Cardiovascular Research (T.L.); Finnish Cultural Foundation; Tampere Tuberculosis Foundation (T.L., M.H.); Emil Aaltonen Foundation (T.L.); and Yrjö Jahnsson Foundation (T.L., M.H.) This work was also supported by grants from the Competitive Research Fund of Pirkanmaa Hospital District (9M017, 9N013 to M.H.). We thank Nina Peltonen and Sinikka Repo-Koskinen for their excellent technical assistance.

**Author details**

<sup>1</sup>Department of Microbiology and Immunology, School of Medicine, University of Tampere, Tampere, Finland. <sup>2</sup>Gerontology Research Center, Tampere, Finland. <sup>3</sup>Fimlab Laboratories, Tampere, Finland. <sup>4</sup>Departments of Clinical Physiology, Tampere University Hospital and University of Tampere School of Medicine, Tampere, Finland. <sup>5</sup>Research Centre of Applied and Preventive Cardiovascular Medicine and the Department of Clinical Physiology and Nuclear Medicine, University of Turku and Turku University Hospital, Turku, Finland. <sup>6</sup>Department of Clinical Chemistry, University of Tampere School of Medicine, Tampere, Finland.

Received: 5 June 2015 Accepted: 1 February 2016

Published online: 09 February 2016

**References**

- Zampieri M, Ciccarone F, Calabrese R, Franceschi C, Burkle A, Caiafa P. Reconfiguration of DNA methylation in aging. *Mech Ageing Dev*. 2015.
- Bacalini MG, Boattini A, Gentilini D, Giampieri E, Pirazzini C, Giuliani C, et al. A meta-analysis on age-associated changes in blood DNA methylation: results from an original analysis pipeline for Infinium 450k data. *Aging (Albany NY)*. 2015;7(2):97–109.
- Accomando WP, Wiencke JK, Houseman EA, Nelson HH, Kelsey KT. Quantitative reconstruction of leukocyte subsets using DNA methylation. *Genome Biol*. 2014;15(3):R50.
- Jaffe AE, Irizarry RA. Accounting for cellular heterogeneity is critical in epigenome-wide association studies. *Genome Biol*. 2014;15(2):R31.
- Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics*. 2012;13:86.
- Johnson AA, Akman K, Calimport SR, Wuttke D, Stolzing A, de Magalhaes JP. The role of DNA methylation in aging, rejuvenation, and age-related disease. *Rejuvenation Res*. 2012;15(5):483–94.
- Bacalini MG, Friso S, Olivieri F, Pirazzini C, Giuliani C, Capri M, et al. Present and future of anti-ageing epigenetic diets. *Mech Ageing Dev*. 2014;136–137:101–15.
- Zhang Y, Yang R, Burwinkel B, Breitling LP, Brenner H. F2RL3 methylation as a biomarker of current and lifetime smoking exposures. *Environ Health Perspect*. 2014;122(2):131–7.
- Breitling LP, Yang R, Korn B, Burwinkel B, Brenner H. Tobacco-smoking-related differential DNA methylation: 27K discovery and replication. *Am J Hum Genet*. 2011;88(4):450–7.
- Heyn H, Li N, Ferreira HJ, Moran S, Pisano DG, Gomez A, et al. Distinct DNA methylomes of newborns and centenarians. *Proc Natl Acad Sci U S A*. 2012; 109(26):10522–7.
- Horvath S. DNA methylation age of human tissues and cell types. *Genome Biol*. 2013;14(10):R115.
- Garagnani P, Bacalini MG, Pirazzini C, Gori D, Giuliani C, Mari D, et al. Methylation of ELOVL2 gene as a new epigenetic marker of age. *Aging Cell*. 2012;11(6):1132–4.
- Hannum G, Guinney J, Zhao L, Zhang L, Hughes G, Sada S, et al. Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol Cell*. 2013;49(2):359–67.
- 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.
- Marttila S, Kananen L, Hayrynen S, Jylhava J, Nevalainen T, Hervonen A, et al. Ageing-associated changes in the human DNA methylome: genomic locations and effects on gene expression. *BMC Genomics*. 2015;16(1):179. Epub 2015 Mar 14.
- Yuan T, Jiao Y, de Jong S, Ophoff RA, Beck S, Teschendorff AE. An integrative multi-scale analysis of the dynamic DNA methylation landscape in aging. *PLoS Genet*. 2015;11(2), e1004996.
- Teschendorff AE, Menon U, Gentry-Maharaj A, Ramus SJ, Weisenberger DJ, Shen H, et al. Age-dependent DNA methylation of genes that are suppressed in stem cells is a hallmark of cancer. *Genome Res*. 2010; 20(4):440–6.
- Raitakari OT, Juonala M, Ronnemaa T, Keltikangas-Jarvinen L, Rasanen L, Pietikainen M, et al. Cohort profile: the cardiovascular risk in Young Finns study. *Int J Epidemiol*. 2008;37(6):1220–6.
- Steegenga WT, Boekschoten MV, Lute C, Hooiveld GJ, de Groot PJ, Morris TJ, et al. Genome-wide age-related changes in DNA methylation and gene expression in human PBMCs. *Age (Dordr)*. 2014;36(3):9648. Epub 2014 May 2.
- Alisch RS, Barwick BG, Chopra P, Myrick LK, Satten GA, Conneely KN, et al. Age-associated DNA methylation in pediatric populations. *Genome Res*. 2012;22(4):623–32.
- Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics*. 2009;10:48.
- Zambelli F, Pesole G, Pavesi G. Pscan: finding over-represented transcription factor binding site motifs in sequences from co-regulated or co-expressed genes. *Nucleic Acids Res*. 2009;37(Web Server issue):W247–52.
- Florath I, Butterbach K, Muller H, Bewerunge-Hudler M, Brenner H. Cross-sectional and longitudinal changes in DNA methylation with age: an epigenome-wide analysis revealing over 60 novel age-associated CpG sites. *Hum Mol Genet*. 2014;23(5):1186–201.
- Johansson A, Enroth S, Gyllenstein U. Continuous aging of the human DNA methylome throughout the human lifespan. *PLoS One*. 2013;8(6), e67378.
- Gubbels Bupp MR. Sex, the aging immune system, and chronic disease. *Cell Immunol*. 2015.
- Oertelt-Prigione S. The influence of sex and gender on the immune response. *Autoimmun Rev*. 2012;11(6-7):A479–85.
- McClay JL, Aberg KA, Clark SL, Nerella S, Kumar G, Xie LY, et al. A methylome-wide study of aging using massively parallel sequencing of the methyl-CpG-enriched genomic fraction from blood in over 700 subjects. *Hum Mol Genet*. 2014;23(5):1175–85.
- Weidner CI, Lin Q, Koch CM, Eisele L, Beier F, Ziegler P, et al. Aging of blood can be tracked by DNA methylation changes at just three CpG sites. *Genome Biol*. 2014;15(2):R24.
- Hernandez DG, Nalls MA, Gibbs JR, Arepalli S, van der Brug M, Chong S, et al. Distinct DNA methylation changes highly correlated with chronological age in the human brain. *Hum Mol Genet*. 2011;20(6):1164–72.
- Reynolds LM, Taylor JR, Ding J, Lohman K, Johnson C, Siscovick D, et al. Age-related variations in the methylome associated with gene expression in human monocytes and T cells. *Nat Commun*. 2014;5:5366.
- West J, Beck S, Wang X, Teschendorff AE. An integrative network algorithm identifies age-associated differential methylation interactome hotspots targeting stem-cell differentiation pathways. *Sci Rep*. 2013;3: 1630.
- Lukic S, Nicolas JC, Levine AJ. The diversity of zinc-finger genes on human chromosome 19 provides an evolutionary mechanism for defense against inherited endogenous retroviruses. *Cell Death Differ*. 2014;21(3):381–7.
- Taruscio D, Mantovani A. Factors regulating endogenous retroviral sequences in human and mouse. *Cytogenet Genome Res*. 2004;105(2-4): 351–62.
- Nuotio J, Oikonen M, Magnussen CG, Jokinen E, Laitinen T, Hutri-Kahonen N, et al. Cardiovascular risk factors in 2011 and secular trends since 2007: the cardiovascular risk in Young Finns study. *Scand J Public Health*. 2014; 42(7):563–71.
- Bibikova M, Lin Z, Zhou L, Chudin E, Garcia EW, Wu B, et al. High-throughput DNA methylation profiling using universal bead arrays. *Genome Res*. 2006;16(3):383–93.
- Bibikova M, Le J, Barnes B, Saedinia-Melnyk S, Zhou L, Shen R, et al. Genome-wide DNA methylation profiling using Infinium(R) assay. *Epigenomics*. 2009;1(1):177–200.
- Bibikova M, Barnes B, Tsan C, Ho V, Klotzle B, Le JM, et al. High density DNA methylation array with single CpG site resolution. *Genomics*. 2011; 98(4):288–95.
- Teschendorff AE, Marabita F, Lechner M, Bartlett T, Tegner J, Gomez-Cabrero D, et al. A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. *Bioinformatics*. 2013;29(2):189–96.
- Marabita F, Almgren M, Lindholm ME, Ruhmann S, Fagerstrom-Billai F, Jagodic M, et al. An evaluation of analysis pipelines for DNA methylation profiling using the Illumina HumanMethylation450 BeadChip platform. *Epigenetics*. 2013;8(3):333–46.
- Pidsley R, Wong CC Y, Volta M, Lunnon K, Mill J, Schalkwyk LC. A data-driven approach to preprocessing Illumina 450K methylation array data. *BMC Genomics*. 2013;14:293.
- Cribari-Neto F, Zeileis A. Beta regression in R. *JSS*. 2010;34:1–24 [http://www.jstatsoft.org/v34/i02/].

42. Triche T, Jr. InfiniumMethylation.hg19: Annotation package for Illumina Infinium DNA methylation probes. R package version 2.2.0.
43. Eden E, Lipson D, Yogev S, Yakhini Z. Discovering motifs in ranked lists of DNA sequences. *PLoS Comput Biol.* 2007;3(3), e39.
44. Supek F, Bosnjak M, Skunca N, Smuc T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One.* 2011;6(7), e21800.

Submit your next manuscript to BioMed Central  
and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)







## Short communication

## Cytomegalovirus infection accelerates epigenetic aging



Laura Kananen<sup>a,b,1</sup>, Tapio Nevalainen<sup>a,b,\*</sup>, Juulia Jylhävä<sup>a,b</sup>, Saara Marttila<sup>a,b</sup>, Antti Hervonen<sup>b,c</sup>,  
Marja Jylhä<sup>b,c</sup>, Mikko Hurme<sup>a,b,d</sup>

<sup>a</sup> Department of Microbiology and Immunology, School of Medicine, University of Tampere, Finland

<sup>b</sup> Gerontology Research Center, University of Tampere, Finland

<sup>c</sup> School of Health Sciences, University of Tampere, Finland

<sup>d</sup> Fimlab Laboratories, Tampere, Finland

## ARTICLE INFO

## Article history:

Received 16 April 2015

Received in revised form 23 June 2015

Accepted 16 October 2015

Available online 17 October 2015

## Keywords:

DNA methylation

Aging

Epigenetic age

Cytomegalovirus

## ABSTRACT

Epigenetic mechanisms such as DNA methylation (DNAm) have a central role in the regulation of gene expression and thereby in cellular differentiation and tissue homeostasis. It has recently been shown that aging is associated with profound changes in DNAm. Several of these methylation changes take place in a clock-like fashion, i.e. correlating with the calendar age of an individual. Thus, the epigenetic clock based on these kind of DNAm changes could provide a new biomarker for human aging process, i.e. being able to separate the calendar and biological age. Information about the correlation of the time indicated by this clock to the various aspects of immunosenescence is still missing. As chronic cytomegalovirus (CMV) infection is probably one of the major driving forces of immunosenescence, we now have analyzed the correlation of CMV seropositivity with the epigenetic age in the Vitality 90+ cohort 1920 (122 nonagenarians and 21 young controls, CMV seropositivity rates 95% and 57%, respectively). The data showed that CMV seropositivity was associated with a higher epigenetic age in both of these age groups (median 26.5 vs. 24.0 ( $p < 0.02$ , Mann–Whitney U-test) in the young controls and 76.0 vs. 70.0 ( $p < 0.01$ ) in the nonagenarians). Thus, these data provide a new aspect to the CMV associated pathological processes.

© 2015 Elsevier Inc. All rights reserved.

## 1. Introduction

Infection with the human cytomegalovirus (CMV) takes place usually in early childhood and its seroprevalence increases with age being 70–100% in elderly individuals in different populations. CMV stays in the body in a latent form, but it can be reactivated e.g. due to physical stress or immune suppression. CMV infections are associated with several aging-associated pathologies such as increased rate of inflammation, changes in the proportions of immune cell types, and consequently, also with increased mortality of elderly individuals (reviewed in Pawelec and Derhovanessian, 2011, Pawelec et al., 2012). However, the published data are not consistent and the causality of these associations is not yet established. Now, we have adopted a new approach to study the effects of chronic CMV infection, its effect on the epigenetic age.

Epigenetic mechanisms such as DNA methylation (DNAm) have a central role in the regulation of gene expression and thereby in cellular differentiation and tissue homeostasis. It has recently been shown that aging is associated with profound changes in DNAm. At the genome-wide level, thousands of the methylation-sensitive cytosines in the cytosine/guanine (CpG) nucleotide pairs are either hypo- or hypermethylated (reviewed in Zampieri et al., 2015). Surprisingly, several of these methylation changes take place in a clock-like fashion, i.e. correlating with the calendar age of an individual. Horvath (2013) constructed an algorithm based on 353 CpG sites demonstrating this kind of behavior. This “epigenetic clock” is thus an interesting new candidate for determination of the biological age of an individual.

There are already a couple of reports, where the age indicated by this epigenetic clock has been used to analyze the correlations with aging-associated phenotypes. In Down syndrome, where the aging process is accelerated, the epigenetic age is clearly higher than the calendar age (Horvath et al., 2015). Obesity seems to increase the epigenetic age in liver, thus being in line with the suspected aging-accelerating effect of obesity (Horvath et al., 2015). In elderly individuals increased epigenetic age correlates with both physical and mental fitness (Marioni et al., 2015b). There is also one study showing that the epigenetic age correlates with mortality. Marioni et al. (2015a) observed that 69–79 year old individuals had a 16% increased mortality during a 4–10 year

\* Corresponding author at: Department of Microbiology and Immunology, School of Medicine, University of Tampere, FIN-33014 Tampere, Finland.

E-mail addresses: [laura.kananen@uta.fi](mailto:laura.kananen@uta.fi) (L. Kananen), [tapio.l.nevalainen@uta.fi](mailto:tapio.l.nevalainen@uta.fi) (T. Nevalainen), [juulia.jylhava@uta.fi](mailto:juulia.jylhava@uta.fi) (J. Jylhävä), [saara.marttila@uta.fi](mailto:saara.marttila@uta.fi) (S. Marttila), [antti.hervonen@uta.fi](mailto:antti.hervonen@uta.fi) (A. Hervonen), [marja.jylha@uta.fi](mailto:marja.jylha@uta.fi) (M. Jylhä), [mikko.a.hurme@uta.fi](mailto:mikko.a.hurme@uta.fi) (M. Hurme).

<sup>1</sup> These authors contributed equally to this work.

follow-up if the epigenetic age was 5 years higher than the calendar age. Now we have determined the epigenetic age in a cohort of both young and old individuals and analyzed the effect of CMV seropositivity on this figure.

## 2. Materials and methods

The study population consisted of 122 nonagenarians and 21 young controls (healthy laboratory personnel). It is a subcohort from the Vitality 90+ study, which is an ongoing prospective population-based project, where all 90 year-old individuals living in the city of Tampere are invited to participate. Biological samples have been collected in 2000 and 2010 (the present subcohort). The recruitment and characterization of the participants have recently been described (Goebeler et al., 2003). The study protocol was approved by the ethics committee of the city of Tampere. The study has been conducted according to the principles expressed in the declaration of Helsinki. Collection of the blood samples and DNA purification from these same individuals has recently been described in detail (Marttila et al., 2015).

Genome-wide DNA methylation was analyzed using the Illumina Infinium HumanMethylation 450k BeadChip Array. Data processing was performed as described (Marttila et al., 2015). The 'DNA methylome age' (DNAm age) in the PBMCs was determined by the methodology presented in a recent study (Horvath, 2013). The formula of DNAm age was discovered by utilization of 21,000 probes that are present in HumanMethylation450 as well as in HumanMethylation27 BeadChips. The predictor was trained with 8000 samples of various tissue types in 82 Illumina DNA methylation array data sets. Based on the training results, the 'epigenetic clock' i.e. the regression model was built with 353 CpG-sites of which methylation level explains the most of the age variation. The algorithm is available at <http://labs.genetics.ucla.edu/horvath/dnamage/>. The anti-CMV titer was measured from the plasma samples using an enzyme-linked immunosorbent assay (Enzygnost Anti-CMV/IgG, Siemens Healthcare, Marburg, Germany). Seropositivity was defined as a titer >230 according to manufacturer's instructions.

As alterations in the leukocyte pool are known to be associated with aging and aging phenotypes, we assessed the relationship between DNAm age and leukocyte proportions using linear regression. The leukocyte subtypes available in our data were CD3<sup>+</sup> cells (of live-gated cells), CD4<sup>+</sup> cells (of CD3<sup>+</sup> cells), CD8<sup>+</sup> cells (of CD3<sup>+</sup> cells), CD4<sup>+</sup>CD28<sup>-</sup> cells (of CD4<sup>+</sup> cells), CD8<sup>+</sup>CD28<sup>-</sup> cells (of CD8<sup>+</sup> cells), CD4<sup>+</sup>CD25<sup>high</sup> (of CD4<sup>+</sup> cells), CD14<sup>+</sup> cells (of live-gated cells) and the CD4<sup>+</sup>/CD8<sup>+</sup> cell ratio (Marttila et al., 2011). All the cell proportions that exhibited statistically significant association with the DNAm age in the univariate assessment, were analyzed in a multivariate linear regression model. CMV serostatus was also added to the multivariate model as a covariate.

## 3. Results

The correlation of the epigenetic age and calendar age of the nonagenarians ( $n = 122$ ) and young controls ( $n = 21$ ) is shown in Fig. 1. In the controls, the correlation is relatively good, but the nonagenarians show a clearly lower epigenetic age than the calendar age. The effect of CMV seropositivity on the epigenetic age is shown in Fig. 2. The epigenetic age was significantly higher in the CMV seropositives (median 26.5 vs. 24.0 in the controls,  $p < 0.02$ , Mann–Whitney U-test, and 76.0 vs. 70.0,  $p < 0.01$ , in the nonagenarians). Results demonstrating the relationships between DNAm age and leukocyte proportions are shown in Supplementary Table 1.

## 4. Discussion

The role of CMV in aging-associated pathologies has extensively been analyzed (reviewed in Pawelec and Derhovanessian, 2011,

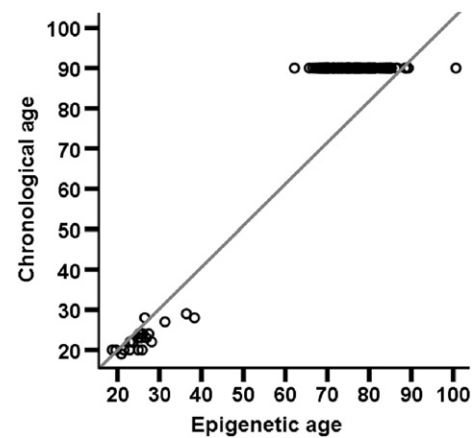


Fig. 1. Chronological age as a function of the epigenetic age in 21 young controls and 122 nonagenarians.

Pawelec et al., 2012). However, the data are not consistent, mainly because several decades are required to its effects to occur and naturally, because of the complexity of the aging process. The data presented here add a new piece to this puzzle: the effect of CMV on the epigenetic age.

This study has some limitations. The number of individuals is relatively small and therefore the associations of the observed differences in the epigenetic age with the known parameters of CMV induced immunosenescence cannot be reliably analyzed. However, one of the hallmarks of CMV seropositivity in aged individuals is the change in the proportions of T cell subsets. Although according to the data from Horvath (2013), the epigenetic age does not vary significantly in sorted blood cells and the clock is accurate also in other cell types, we addressed this issue in our data. The assessment of the relationship between DNAm age, CMV serostatus and leukocyte subpopulations demonstrated that the proportion of CD4<sup>+</sup>CD28<sup>-</sup> explains most of the variation in the DNAm age and replaces the CMV serostatus and all the other cell types as a determinant. This finding suggests that there is a complex interplay between CMV and the CD4<sup>+</sup>CD28<sup>-</sup> cell population in relation to the DNAm age. This finding is somewhat unexpected as the CD4<sup>+</sup>CD28<sup>-</sup> cells are not the primary target of CMV. Nevertheless, as CMV seropositivity is strongly associated with lower CD4<sup>+</sup>CD28<sup>-</sup> cell proportion in our data ( $p \leq 0.001$ , Mann–Whitney U-test), we speculate that for some (unknown) reason CMV and CD4<sup>+</sup>CD28<sup>-</sup> cells are so tightly interlinked that it is virtually impossible

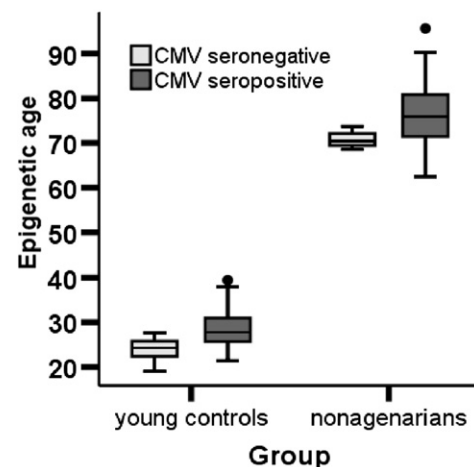


Fig. 2. The epigenetic age is illustrated as boxplots and categorized by age group (young controls  $N = 21$ , nonagenarians  $N = 122$ ) and CMV seropositivity/–negativity. The line in the middle in each box corresponds to value of median and the box contains 50% of the age values in a group.

to disentangle their individual effects on the aging of the human methylome. However, as the cell type typically considered to reflect the effect of CMV on aging-related phenotypes, i.e., the CD8<sup>+</sup>CD28<sup>-</sup>, did not explain the variation in the DNAm age better than CMV serostatus, we suggest that the relationship between CMV and DNA methylation age does not involve major CMV-induced alterations in the lymphocyte pool.

The biological mechanisms of the clock-like methylation pattern of some CpG sites are not clear. It has been suggested that it is an indication of a general epigenetic maintenance system required to keep the epigenome stable (Horvath, 2013). This interpretation is supported by the finding that in childhood and adolescence the epigenetic age/calendar age ratio is logarithmic, i.e. during active differentiation and growth efficient maintenance is required (Horvath, 2013). In aged individuals the epigenetic clock seems to slow down. This is probably due to decreased cell proliferation, regeneration or to slower metabolism of the elderly, i.e. being a biomarker of biological aging. The median epigenetic age of the nonagenarians in this cohort was 14 years lower than the calendar age. This difference is even larger than that observed in the study of Marioni et al. (2015a) in 66–79 year old individuals, maybe indicating that the clock is continuously slowing down with advancing age. However, at present we cannot exclude the possibility that this observed slowing down is not due to these biological mechanisms, but is rather due to simple selection, i.e. individuals with a low epigenetic age have better chances to attain the age of 90 years. Obviously, larger cohorts, and preferably in a follow-up setting, are required to analyze this phenomenon and its significance.

It has already been demonstrated that the host cell methylation state and sensitivity to CMV infection are correlated (Esteki-Zadeh et al., 2012). Additionally, CMV infection seems to cause global hypomethylation in some cell lines (Esteki-Zadeh et al., 2012). Therefore, it is maybe not surprising that these CMV infection—associated methylomic changes are a part of this putative epigenetic maintenance mechanism. Thus, it is also possible that other infections associated with epigenetic stress would have an increasing effect on the epigenetic age.

#### Author contributions

LK and TN participated in the data analysis and writing of the manuscript. JJ and SM participated in the sample collection. AH and MJ were responsible for recruiting the study population. MH provided reagents and materials for the study and participated in the writing of the manuscript. All authors read and approved the final manuscript.

#### Competing financial interests

The authors declare no competing financial interests.

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.exger.2015.10.008>.

#### Acknowledgments

This study was supported by grants from the Competitive Research Fund of Pirkanmaa Hospital District (9M017, 9N013 to MH, 9N012 to AH) and the Academy of Finland (132704 to MH). The authors would like to thank Sinikka Repo-Koskinen, Katri Välimaa and Sanna Tuomisaari for their skillful technical assistance.

#### References

- Esteki-Zadeh, A., Karimi, M., Straat, K., Ammerpohl, O., Zeitelhofer, M., Jagodic, M., et al., 2012. Human cytomegalovirus infection is sensitive to the host cell DNA methylation state and alters global DNA methylation capacity. *Epigenetics* 7 (6), 585–593. <http://dx.doi.org/10.4161/epi.20075>.
- Goebele, S., Jylha, M., Hervonen, A., 2003. Medical history, cognitive status and mobility at the age of 90. A population-based study in Tampere, Finland. *Aging Clin. Exp. Res.* 15 (2), 154–161.
- Horvath, S., 2013. DNA methylation age of human tissues and cell types. *Genome Biol.* 14 (10), R115 (doi:gb-2013-14-10-r115 [pii]).
- Horvath, S., Garagnani, P., Bacalini, M.G., Pirazzini, C., Salvioli, S., Gentilini, D., et al., 2015. Accelerated epigenetic aging in Down syndrome. *Aging Cell* <http://dx.doi.org/10.1111/ace1.12325>.
- Marioni, R.E., Shah, S., McRae, A.F., Chen, B.H., Colicino, E., Harris, S.E., et al., 2015a. DNA methylation age of blood predicts all-cause mortality in later life. *Genome Biol.* 16 (1), 25 (doi:s13059-015-0584-6 [pii]).
- Marioni, R.E., Shah, S., McRae, A.F., Ritchie, S.J., Muniz-Terrera, G., Harris, S.E., et al., 2015b. The epigenetic clock is correlated with physical and cognitive fitness in the Lothian birth cohort 1936. *Int. J. Epidemiol.* (doi:dyu277 [pii]).
- Marttila, S., Jylhava, J., Pesu, M., Hamalainen, S., Jylha, M., Hervonen, A., Hurme, M., 2011. IL-7 concentration is increased in nonagenarians but is not associated with markers of T cell immunosenescence. *Exp. Gerontol.* 46 (12), 1000–1002. <http://dx.doi.org/10.1016/j.exger.2011.09.004>.
- Marttila, S., Kananen, L., Hayrynen, S., Jylhava, J., Nevalainen, T., Hervonen, A., et al., 2015. Ageing-associated changes in the human DNA methylome: genomic locations and effects on gene expression. *BMC Genomics* 16 (1). <http://dx.doi.org/10.1186/s12864-015-1381-z> (1381–015–1381-z. Epub 2015 Mar 14).
- Pawelec, G., Derhovanessian, E., 2011. Role of CMV in immune senescence. *Virus Res.* 157 (2), 175–179. <http://dx.doi.org/10.1016/j.virusres.2010.09.010>.
- Pawelec, G., McElhaney, J.E., Aiello, A.E., Derhovanessian, E., 2012. The impact of CMV infection on survival in older humans. *Curr. Opin. Immunol.* 24 (4), 507–511. <http://dx.doi.org/10.1016/j.coi.2012.04.002>.
- Zampieri, M., Ciccarone, F., Calabrese, R., Franceschi, C., Burkle, A., Caiafa, P., 2015. Reconfiguration of DNA methylation in aging. *Mech. Ageing Dev.* (doi:S0047-6374(15)00007-X [pii]).

# The trajectory of the blood DNA methylome ageing rate is largely set before adulthood: evidence from two longitudinal studies

L. Kananen · S. Marttila · T. Nevalainen · L. Kummola · I. Junttila · N. Mononen · M. Kähönen · O. T. Raitakari · A. Hervonen · M. Jylhä · T. Lehtimäki · M. Hurme · J. Jylhävä

Received: 19 November 2015 / Accepted: 31 May 2016 / Published online: 14 June 2016  
© American Aging Association 2016

**Abstract** The epigenetic clock, defined as the DNA methylome age (DNAmAge), is a candidate biomarker of ageing. In this study, we aimed to characterize the behaviour of this marker during the human lifespan in more detail using two follow-up cohorts (the Young Finns study, calendar age i.e. cAge range at baseline 15–24 years, 25-year-follow-up,  $N = 183$ ; The Vitality 90+ study, cAge range at baseline 19–90 years, 4-year-follow-up,  $N = 48$ ). We also aimed to assess the relationship between DNAmAge estimate and the blood cell distributions, as both of these measures are known to change as a function of age. The subjects’

DNAmAges were determined using Horvath’s calculator of epigenetic cAge. The estimate of the DNA methylome age acceleration ( $\Delta$ -cAge-DNAmAge) demonstrated remarkable stability in both cohorts: the individual rank orders of the DNAmAges remained largely unchanged during the follow-ups. The blood cell distributions also demonstrated significant intra-individual correlation between the baseline and follow-up time points. Interestingly, the immunosenescence-associated features (CD8+CD28– and CD4+CD28– cell proportions and the CD4/CD8 cell ratio) were tightly associated with the estimate of the DNA methylome

---

**Electronic supplementary material** The online version of this article (doi:10.1007/s11357-016-9927-9) contains supplementary material, which is available to authorized users.

---

L. Kananen (✉) · S. Marttila · T. Nevalainen · M. Hurme · J. Jylhävä  
Department of Microbiology and Immunology, School of Medicine, University of Tampere, Tampere, Finland  
e-mail: laura.kananen@uta.fi

L. Kananen · S. Marttila · T. Nevalainen · A. Hervonen · M. Jylhä · M. Hurme · J. Jylhävä  
Gerontology Research Center, Tampere, Finland

L. Kummola · I. Junttila  
School of Medicine, University of Tampere, Tampere, Finland

I. Junttila · N. Mononen · T. Lehtimäki · M. Hurme  
Department of Clinical Chemistry, Fimlab Laboratories, Tampere, Finland

M. Kähönen  
Department of Clinical Physiology, Tampere University Hospital and University of Tampere School of Medicine, Tampere, Finland

O. T. Raitakari  
Research Centre of Applied and Preventive Cardiovascular Medicine and the Department of Clinical Physiology and Nuclear Medicine, University of Turku and Turku University Hospital, Turku, Finland

A. Hervonen · M. Jylhä  
School of Health Sciences, University of Tampere, Tampere, Finland

T. Lehtimäki  
Department of Clinical Chemistry, University of Tampere School of Medicine, Tampere, Finland

age. In summary, our data demonstrate that the general level of  $\Delta$ -cAge-DNAmeAge is fixed before adulthood and appears to be quite stationary thereafter, even in the oldest-old ages. Moreover, the blood DNAmeAge estimate seems to be tightly associated with ageing-associated shifts in blood cell composition, especially with those that are the hallmarks of immunosenescence. Overall, these observations contribute to the understanding of the longitudinal aspects of the DNAmeAge estimate.

**Keywords** DNA methylation · DNAmeAge · Epigenetic clock · Follow-up · Immunosenescence

### Abbreviations

cAge	Calendar age
DNAmeAge	DNA methylome age
FACS	Fluorescence-activated cell sorting analysis
MAD	The median of the differences between DNAmeAge and cAge
PBMC	Peripheral blood mononuclear cell
PCA	Principal component analysis
WBL	White blood leukocyte
YFS	Young Finns study
V90	Vitality 90+ study
$\Delta$ -cAge-DNAmeAge	The difference between calendar age and DNA methylome age

### Background

In the field of biogerontology, one important objective has been to develop a universal biomarker of ageing that describes an individual's biological age as opposed to calendar age (cAge). For example, one promising biomarker has been telomere length, which is conveniently obtainable. However, the correlation between telomere length and cAge age is only moderate (Benetos et al. 2001), and the relevance of telomere length as an unambiguous biomarker of ageing is questionable (Mather et al. 2011). Recently, based on large-scale epigenome-wide association studies, the DNA methylation levels at certain CpG sites (Garagnani et al. 2012; Horvath 2013; Hannum et al. 2013; Steegenga et al. 2014; Kananen et al. 2016) have been shown to be strongly associated with cAge in multiple tissues; i.e. methylation at these sites shows clocklike behaviour. These findings have

been further refined to produce DNA methylome age (DNAmeAge) calculation algorithms. These algorithms provide surprisingly accurate age predictions, with correlation coefficients between cAge and DNAmeAge as high as 0.9 and median absolute deviations (MADs, the median of the differences between DNAmeAge and cAge) of age predictions as low as 3 years (Bocklandt et al. 2011; Hannum et al. 2013; Horvath 2013; Weidner et al. 2014; Florath et al. 2014). Horvath (2013)) generated a multi-tissue age predictor algorithm (available at <https://dnamage.genetics.ucla.edu/>) based on elastic net regression analysis using 21,000 probes from the HumanMethylation450 and HumanMethylation27 BeadChips. This predictor algorithm was trained with 8000 samples of various tissue types and cAge ranges in 82 Illumina DNA methylation array data sets. On the grounds of the training results, the 'epigenetic clock', i.e. the regression model, was constructed and the model consists of 353 CpG sites whose methylation level best explains the variation in cAge. Using this algorithm, the difference between cAge and DNAmeAge estimate ( $\Delta$ -cAge-DNAmeAge) might be considered as a quantitative measure of a person's biological ageing.

Horvath's epigenetic clock ticking rate has been shown to be highest during organismal growth (Horvath 2013). However, observations in non-proliferative and immortalized tissues have indicated that DNAmeAge is different from mitotic age and that DNAmeAge might rather be a measure of the magnitude of epigenetic maintenance (Horvath 2013; Horvath and Levine 2015), although the available evidence is limited. In addition, analysis of twins provided some evidence that the heritability of DNAmeAge is 100 % at the time of birth but decreases to 39 % by late adulthood (Horvath 2013). Therefore, non-genetic factors appear to influence the epigenetic clock ticking rate, but it remains unclear when and how these inter-individual differences in the epigenetic clock arise and how stable these differences during distinct stages of the human lifespan are.

In recent studies, accelerated biological ageing, as indicated by DNAmeAge estimate, has been demonstrated regarding several ageing-associated phenotypes (physical and mental fitness (Marioni et al. 2015b), mortality (Marioni et al. 2015a) or phenotypes with ageing-resembling features (HIV1 (Horvath and Levine 2015) and Down syndrome (Horvath et al. 2015a)). In addition, obesity induces an increase in DNAmeAge in hepatocytes (Horvath et al. 2014).

Nevertheless, some of these results are conflicting; e.g. military personnel suffering from more severe post-traumatic stress symptoms have been demonstrated to have a lower DNAmAge (Boks et al. 2015). However, because of the limited number of longitudinal studies on DNAmAge, the behaviour and utility of the DNAmAge as a biomarker of ageing are currently unknown. As different blood cell types display distinct DNA methylation profiles (Houseman et al. 2012; Jaffe and Irizarry 2014) and the immune cell distribution changes as a function of age (Weiskopf et al. 2009; Pawelec et al. 2010), the relationship between alterations in methylation patterns and blood cell distributions is specifically relevant to ageing and age-related phenomena.

The immune cell repertoire in individuals aged 70 years and older demonstrates a notable decline; the classic manifestation of immunosenescence is the inability to generate appropriate responses to vaccines and infections (Weiskopf et al. 2009). The phenomena underlying immunosenescence include clonal expansion of CD4+CD28− and CD8+CD28− cells, which lack the costimulatory surface receptor CD28, a key factor in T cell activation. CD28− cells exhibit shortened telomeres and produce excessive levels of proinflammatory cytokines. Furthermore, the lifespan of CD28− cells is prolonged and these cells may survive for years in the periphery because they undergo apoptosis less frequently than e.g. CD28+ cells. (Weiskopf et al. 2009; Arnold et al. 2011) Although CD28− cells constitute a minor proportion of the total leukocyte pool, they are key players in immune system ageing. Thus, the CD28− cells should be considered when examining the implications of ageing on DNA methylation in leukocytes.

Horvath's DNAmAge has been reported to be relatively independent from variation in the composition of the major blood cell types, but these considerations apply only to middle-aged men and to samples from individuals with a mean cAge between 66 and 79 years (Horvath 2013; Marioni et al. 2015a). Furthermore, few population-based cross-sectional studies have examined genome-wide methylation patterns accounting for blood cell type distributions, yet typically only the major blood cell types have been analysed (Lam et al. 2012; Zilbauer et al. 2013; Jaffe and Irizarry 2014; Marttila et al. 2015). Thus, longitudinal studies including more detailed cell type characterizations and covering wider age ranges of individuals of both sexes are needed. Hence, using two Finnish follow-up

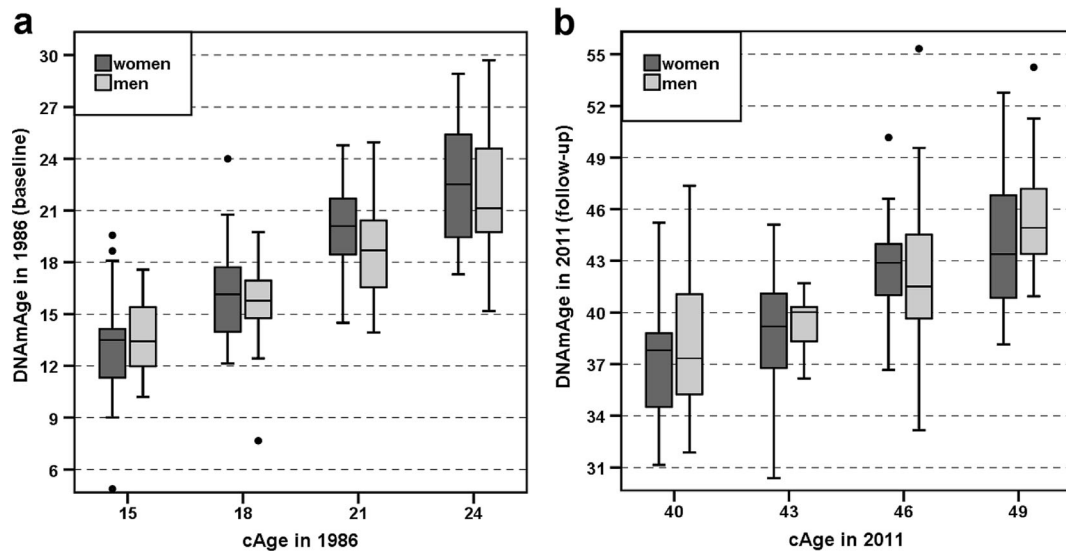
cohorts, The Vitality 90+ study (V90) (age of nonagenarians at baseline of 90 years and age range of young controls of 19–34 years, 4-year follow-up,  $N = 48$ ) and The Young Finns study (YFS) (age range at baseline of 15–24 years, 25-year follow-up,  $N = 183$ ), we aimed to elucidate the longitudinal characteristics of the DNAmAge estimate. Specifically, we addressed the individual trajectories of the DNA methylome age acceleration estimate throughout adulthood in both sexes and examined how the DNAmAge estimate parallels with the age-related fluctuation of blood cell composition. With specific relevance to ageing, our data included also the oldest-old individuals and we assessed the distributions of the specific immunosenescence-representing parameters: the CD4+/CD8+ cell ratio and CD4+CD28− and CD8+CD28− lymphocytes.

## Results

### DNA methylome age in individuals in the YFS

Horvath's DNAmAge was determined for each individual using a multi-tissue age predictor (Horvath 2013). First, the resulting DNAmAges of the YFS participants were evaluated separately in relation to cAge and gender using samples collected in years 1986 and 2011. As shown in Fig. 1a, b and Supplementary Table 1 (Additional File 2), the gender-wise medians of DNAmAge were less than the corresponding cAge at both time points. The median  $\pm$  MAD DNAmAges of the subjects in the YFS\_1986 cohort and the corresponding values of the same subjects in the YFS\_2011 cohort are shown in Supplementary Table 1 (Additional File 2). Spearman's rank sum correlation coefficients between cAge and DNAmAge were 0.785 ( $P = 8.10 \times 10^{-25}$ ) among women and 0.702 ( $P = 1.25 \times 10^{-11}$ ) among men in the YFS\_1986 cohort. The corresponding values in the YFS\_2011 cohort were 0.635 ( $P = 4.24 \times 10^{-14}$ ) among women and 0.591 ( $P = 7.36 \times 10^{-8}$ ) among men.

The changes in DNAmAge from 1986 to 2011 (i.e. how many years the DNA methylome of a person had aged in 25 years: DNAmAge in 2011 – DNAmAge in 1986) were  $23.23 \pm 2.74$  years among women ( $N = 113$ ),  $23.92 \pm 3.10$  years among men ( $N = 70$ ) and  $23.26 \pm 2.97$  among all individuals ( $N = 183$ ). This



**Fig. 1** **a, b** The DNAmAge values in the YFS samples at baseline and at follow-up after 25 years are visualized as *boxplots*. The YFS\_1986 sample ( $N = 183$ ) is shown in panel **a**, and the YFS\_2011 sample ( $N = 183$ ) is shown in panel **b**. The subjects were categorized by cAge and gender. The *boxes* for women and

men are coloured *dark* and *light grey*, respectively. The differences in the median DNAmAge values between men and women at each cAge group were not prominent (Mann-Whitney  $U$  test,  $P > 0.05$ ). More detailed information can be found in Additional Files 1 and 2. (Colour figure online)

change corresponds to 3.7 years of methylome ageing in 4 calendar years. The changes in DNAmAge did not significantly differ between men and women (Mann-Whitney  $U$  test,  $P = 0.076$ ).

#### DNA methylome ages in individuals in the V90

Similarly to the analysis using the YFS data, the resulting DNAmAges of the V90 participants were first evaluated separately in relation to cAge and gender using samples collected in years 2010 and 2014. Spearman's rank sum correlation coefficients between cAge and DNAmAge were 0.594 ( $P = 3.6 \times 10^{-11}$ ) among women and 0.661 ( $P = 3.4 \times 10^{-6}$ ) among men in the V90\_2010 cohort. The corresponding values in V90\_2014 were 0.925 ( $P = 5.4 \times 10^{-24}$ ) among women and 0.903 ( $P = 1.2 \times 10^{-10}$ ) among men. As shown in Table 1, Fig. 2a, b and Supplementary Figs. 3 and 4 (Additional File 2), the median DNAmAges were less than the corresponding cAges within most, but not all, categories: the median DNAmAges of the subjects aged 19–34 years in the V90\_2014 cohort were higher than their cAges.

To determine whether DNA methylomes of men, women, young, or old subjects aged differently during the follow-up period, the changes in DNAmAge from 2010 to 2014 (DNAmAge in 2014 – DNAmAge in

2010) were calculated. Over this period, the methylome of women aged  $4.18 \pm 3.04$  years ( $N = 30$ ) and that of men aged  $6.31 \pm 2.60$  years ( $N = 18$ ) (Mann-Whitney  $U$  test,  $P = 0.255$ , total  $N = 48$ ). Within the cAge groups (categorized as nonagenarians and young controls), the changes of DNAmAge were  $7.39 \pm 2.11$  years among the young controls ( $N = 7$ ) and  $4.53 \pm 3.52$  years among the nonagenarians ( $N = 41$ ; Mann-Whitney  $U$  test,  $P = 0.020$ ) during the 4-year follow-up period. There was no difference in the change in DNAmAge between nonagenarian men ( $N = 14$ ) and women ( $N = 27$ ) during the follow-up period ( $P = 0.582$ ). The corresponding gender-wise analysis of young controls in the V90 cohort was discarded due to the small sample size of subjects participating at both time points.

#### $\Delta$ -cAge-DNAmAge in the YFS participants

The difference between calendar age and DNA methylome age ( $\Delta$ -cAge-DNAmAge) in whole blood leukocytes (WBLs) was calculated for all study subjects. In the YFS, the  $\Delta$ -cAge-DNAmAge was  $3.24 \pm 3.17$  (median  $\pm$  MAD) among women and  $2.95 \pm 3.20$  among men in 1986 (baseline, 15–24-year-old subjects), and in 2011 (follow-up, 40–49-year-old subjects), the corresponding values were  $1.50 \pm 2.73$  among women and  $2.07 \pm 2.77$  among men. There was no statistically

**Table 1** Summary of DNAmAges and PBMC counts for the V90 sample in 2010 and 2014

Sample collection year	Group	Number	cAge	DNAmAge of PBMCs	CD3+ of live-gated cells (%)	CD4+ of CD3+ (%)	CD8+ of CD3+ (%)	CD4+/CD8+ ratio	CD28- of CD4+ (%)	CD28- of CD8+ (%)	CD19 of live-gated cells (%)	CD14+ of live-gated cells (%)
2010	19–28 years, women	14	22.5 (3.0)	22.3 (2.5)	71.6 (3.3)	59.0 (4.5)	30.1 (6.0)	2.1 (0.5)	0.2 (0.1)	14.7 (6.8)	na	2.0 (1.11)
	20–29 years, men	7	24.0 (4.5)	22.5 (4.4)	67.0 (3.0)	60.2 (2.7)	31.8 (4.7)	1.9 (0.5)	0.3 (0.4)	18.1 (13.3)	na	2.4 (0.44)
	90 years, women	84	90.0 (0)	81.7 (7.3)	60.5 (9.6)	65.6 (15.6)	27.1 (13.0)	2.4 (1.7)	9.3 (9.7)	65.3 (17.5)	na	8.5 (4.52)
	90 years, men	31	90.0 (0)	87.2 (5.3)	54.3 (10.2)	57.2 (12.5)	28.7 (13.0)	2.0 (1.1)	9.2 (8.5)	69.0 (16.0)	na	7.5 (4.30)
2014	19–34 years, women	27	26.0 (4.4)	29.4 (6.8)	67.0 (7.0)	58.9 (5.6)	30.2 (6.8)	1.9 (0.5)	2.3 (2.4)	30.0 (10.8)	5.8 (2.5)	4.4 (4.6)
	22–32 years, men	13	28.0 (4.4)	35.1 (9.2)	59.0 (6.2)	55.1 (4.2)	32.8 (4.7)	1.6 (0.3)	1.4 (1.9)	39.2 (17.0)	8.0 (4.4)	8.5 (9.9)
	94 years, women	27	94 (0)	89.6 (7.2)	58.8 (11.7)	49.1 (18.2)	35.6 (19.1)	1.3 (1.1)	14.3 (12.2)	74.4 (16.5)	2.8 (2.3)	8.4 (8.7)
	94 years, men	14	94 (0)	91.5 (11.8)	62.5 (8.9)	58.3 (19.3)	28.1 (14.6)	2.1 (1.5)	10.0 (11.6)	75.0 (16.1)	2.2 (0.7)	10.8 (11.6)

The values were categorized by cAge group and gender. The table shows the median (median absolute deviation) values. Data not available are denoted as 'na'

significant difference in the median  $\Delta$ -cAge-DNAmAge between genders in either YFS\_1986 ( $P = 0.314$ ) or YFS\_2011 ( $P = 0.654$ ). The  $\Delta$ -cAge-DNAmAge values are shown in further detail in Supplementary Figs. 1 and 2 (Additional File 2), in which the subjects were categorized by cAge group and gender. Figure 3a illustrates the relationship between the  $\Delta$ -cAge-DNAmAge values in 1986 and those in 2011. The Spearman's rank sum correlation coefficient between the  $\Delta$ -cAge-DNAmAge values was 0.515 ( $P = 5.2 \times 10^{-9}$ ) for women ( $N = 113$ ), 0.567 ( $P = 3.1 \times 10^{-7}$ ) for men ( $N = 70$ ) and 0.535 ( $P = 6.1 \times 10^{-15}$ ) for all participants ( $N = 183$ ).

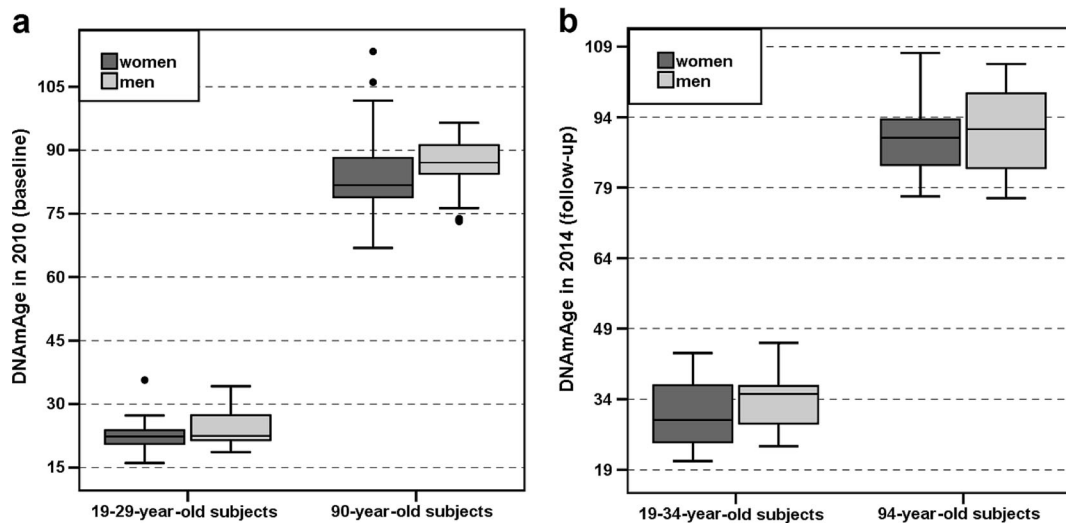
$\Delta$ -cAge-DNAmAge in the V90 participants

In the V90, the median  $\Delta$ -cAge-DNAmAge values in 2010 (baseline) were  $6.27 \pm 6.99$  among all 90-year-old subjects ( $N = 122$ ) and  $-0.363 \pm 2.83$  among all 19–29-year-old subjects ( $N = 21$ ). In 2014 (follow-up), the median  $\Delta$ -cAge-DNAmAge values were  $3.63 \pm 8.74$  among all 94-year-old subjects ( $N = 41$ ) and  $-3.87 \pm 3.34$  among all 19–34-year-old subjects ( $N = 40$ ). The  $\Delta$ -cAge-DNAmAge values are shown in further detail in Supplementary Figs. 3 and 4 (Additional File 2), in which the subjects were categorized by cAge group and gender. The relationship between the  $\Delta$ -cAge-DNAmAge values in 2010 and those in 2014 ( $N = 48$ ) is illustrated in Fig. 3b; the corresponding Spearman's rank sum correlation coefficient between the  $\Delta$ -cAge-DNAmAge values was 0.895 ( $P = 9.2 \times 10^{-18}$ ) for all subjects in the V90. Due to limited sample sizes, further statistical analyses in subgroups of small sample sizes were not performed.

Associations of DNAmAge with cell counts

The cell counts were determined using either genome-wide DNA methylation profiles together with a specific cell count estimation algorithm (Jaffe and Irizarry 2014) (YFS) or fluorescence-activated cell sorting (FACS) analysis (V90). Then, the associations between the cell counts and DNAmAge were determined using Spearman rank sum correlation analysis. First, the correlation of proportions of blood cell types at each time point with the corresponding  $\Delta$ -cAge-DNAmAge values were examined (Tables 2 and 3). In the V90, the most significant peripheral blood mononuclear cell (PBMC) subtype correlate with  $\Delta$ -cAge-DNAmAge was CD4+CD28-



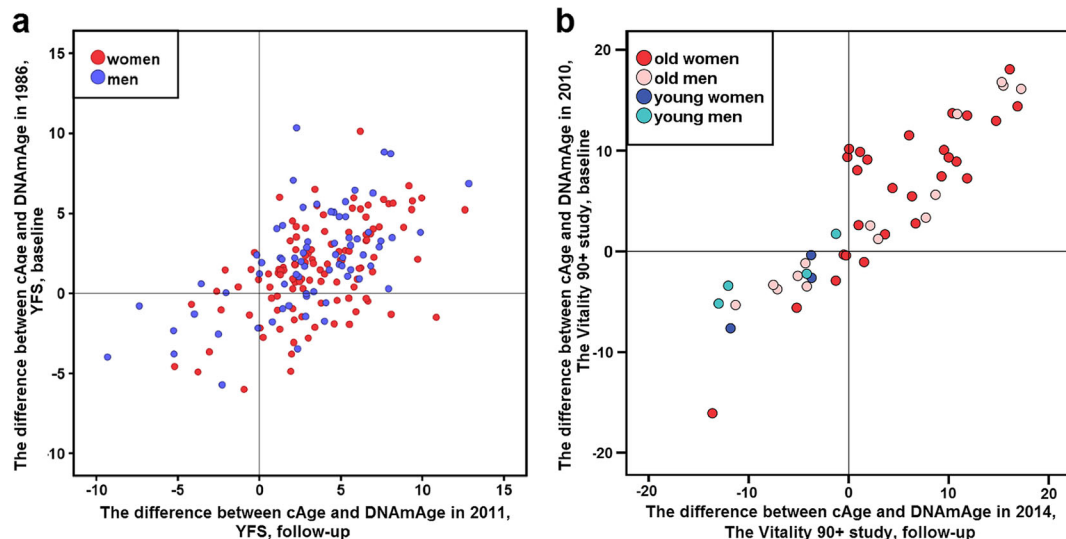


**Fig. 2 a, b** The DNAmAge values in the V90 samples are visualized as *boxplots*. The V90\_2010 sample ( $N = 143$ ) is shown in panel **a**, and the V90\_2014 sample ( $N = 81$ ) is shown in panel **b**. The samples were categorized by cAge group (on the *x-axis*) and gender (*box colours*). A general trend of gender difference was observed in the DNAmAges of PBMCs in the nonagenarians and the young controls at both time points of the V90. The difference in the median DNAmAge values between nonagenarian men ( $N = 33$ ) and women ( $N = 89$ ) was significant (Mann-Whitney  $U$

test,  $P = 0.006$ ) in the V90\_2010 sample (panel **a**). However, in the categories with smaller sample sizes, in the younger cAge groups of the V90\_2010 (panel **a**) and V90\_2014 (panel **b**) samples and in the 94-year-old group of the V90\_2014 (panel **b**) sample, the differences in the median DNAmAge values were not statistically significant between men and women (Mann-Whitney  $U$  test,  $P > 0.05$ ). More detailed information can be found in Additional Files 1 and 2

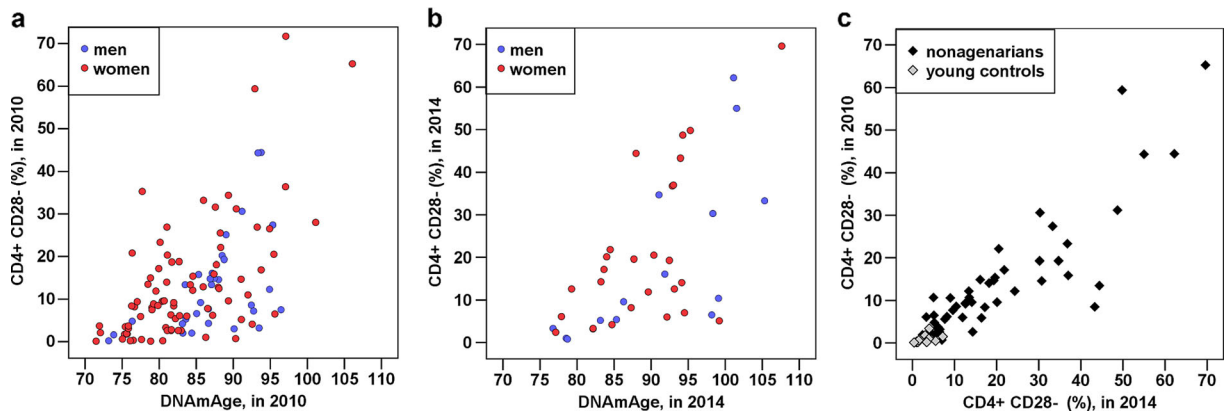
cells among nonagenarians: an increased proportion of these cells was associated with older DNAmAge. Other

significant correlates in nonagenarians were CD4+ CD3+, CD8+ of CD3+, CD8+CD28- cells and the



**Fig. 3 a, b** The correlations between the  $\Delta$ -cAge-DNAmAge values at different time points are shown in *scatterplots*. **a** Subjects participating in the YFS in 1986 and 2011. The *dots* corresponding to women ( $N = 113$ ) are coloured *red*, and the *dots* corresponding to men ( $N = 70$ ) are coloured *blue*. **b** Subjects participating in the V90 in 2010 and 2014 ( $N = 48$ ). The colours of the *dots* for each

subject are explained in the *graph legend*. The Spearman's rank sum correlation coefficient between  $\Delta$ -cAge-DNAmAge values was 0.535 ( $P = 6.1 \times 10^{-15}$ ) for the YFS sample ( $N = 183$ , panel **a**) and 0.895 ( $P = 9.2 \times 10^{-18}$ ) for all subjects in the V90 ( $N = 48$ , panel **b**). More detailed information can be found in Additional Files 1 and 2. (Colour figure online)



**Fig. 4** a–c The CD28–CD4+ cell counts of the participants examined in 2010 and 2014 correlate with their corresponding DNAmAge values and with one another. The associations are visualized as *scatterplots*. Each *dot* corresponds to one subject; the colours of the *dots* are explained in the *graph legends*. **a** The CD28–CD4+ proportions of 90-year-old subjects in V90\_2010 correlated with their DNAmAge in 2010 ( $r = 0.540$ ,  $P = 4.75 \times 10^{-10}$ ,  $N = 115$ ). **b** The CD28×CD4+ proportions of

94-year-old subjects in V90\_2014 correlated with their DNAmAge in 2014 ( $r = 0.626$ ,  $P = 1.20 \times 10^{-5}$ ,  $N = 41$ ). **c** The CD28–CD4+ proportions of all participants in V90\_2010 correlated with their CD28–CD4+ proportions in V90\_2014 (among all:  $r = 0.901$ ,  $P = 9.48 \times 10^{-24}$ ,  $N = 63$ ; among elderly:  $r = 0.864$ ,  $P = 3.33 \times 10^{-16}$ ,  $N = 51$ ; among young controls:  $r = 0.684$ ,  $P = 0.014$ ,  $N = 12$ ). More detailed information can be found in Table 2 and Table 4 and Supporting Material in Additional File 1

CD4+/CD8+ ratio (Table 2). In the young controls, the significant correlates were the CD4+/CD8+ ratio, CD3+ of live-gated, CD4+CD3+, CD4+CD28– and CD8+ CD28– cells (Table 2). Among the YFS individuals aged 15–24 years, the significant WBL subtype correlate with  $\Delta$ -cAge-DNAmAge was CD4+ T cells (Table 3). However, after 25 years, the significant correlates for the same individuals were the CD4+/CD8+ ratio and NK cells (Table 3).

Next, the fluctuations in cell subtype proportions in blood over the follow-up period were evaluated using correlation analysis. The proportion of each cell type and CD4+/CD8+ ratio at baseline were correlated against the corresponding variable at the follow-up time-point (Tables 4 and 5). During the 4-year follow-up period in the V90 participants, the cell proportions were stationary (Table 4, Fig. 4c). The most stable cell types or ratios during the 25-year-follow-up period in the YFS individuals were NK and B cells and the CD4+/CD8+ ratio; the correlation coefficients for these variables were between 0.45 and 0.55 (Table 5).

## Discussion

The human epigenetic clock, assessed as the DNAmAge estimate or DNAmAge acceleration (the difference between cAge and DNAmAge estimate), has recently gained much attention as a marker of human biological

age. However, the utility and the behaviour of this marker in the course of the lifespan are incompletely understood. Here, we present a longitudinal evaluation of the characteristics of DNAmAge estimate (Horvath 2013) during adulthood and at advanced ages. The most notable finding in our study is that the estimate of DNA methylome age acceleration is remarkably stable throughout adulthood and that the rate at which this trajectory proceeds appears to be set before adulthood. The data from the YFS demonstrated that already at the age of 15 years, the slope of DNA methylome ageing was already on an individual-specific level, such that over the 25-year follow-up period, very little deviation from this line was observed. A clear correlation in the scatterplot in Fig. 3a illustrates this finding. At old ages, the DNA methylome ageing rate was likewise steady (Fig. 3b). Previous findings provide support to our results. The creator of the DNAmAge calculator (Horvath 2013) reported first in his study using cross-sectional DNAmAge data that the epigenetic clock ticking rate is accelerated before adulthood and thereafter the rate slows down. In addition, the DNA methylomes of twins are identical at birth and begin to diverge thereafter, and by the late adulthood, the methylomes of twin pairs display only 39 % similarity (at median age of 63 years, all women) (Horvath 2013). Most importantly, a recent follow-up study (Simpkin et al. 2016) on children (baseline at birth, follow-ups at ages of 7 and 17 years) showed that epigenetic age acceleration (the difference

**Table 2** The associations of PBMC subtype proportions with DNAmAge values at baseline and at follow-up among nonagenarians (born in 1920) and young controls (aged 19–34 years) in the V90 sample

Cell type proportion (%)	Nonagenarians				Young controls			
	$\Delta$ -cAge-DNAmAge at baseline in 2010, $N = 115$		$\Delta$ -cAge-DNAmAge at follow-up in 2014, $N = 41$		$\Delta$ -cAge-DNAmAge at baseline in 2010, $N = 21$		$\Delta$ -cAge-DNAmAge at follow-up in 2014, $N = 40$	
	$r$	$P$	$r$	$P$	$r$	$P$	$r$	$P$
CD3+ of live-gated cells	0.106	0.261	-0.251	0.114	-0.065	0.780	0.404*	0.010
CD4+ of CD3+	0.316*	0.001	0.311*	0.048	0.505*	0.019	0.382*	0.015
CD8+ of CD3+	-0.310*	0.001	-0.326*	0.038	-0.233	0.309	-0.232	0.150
CD4+/CD8+ ratio	0.309*	0.001	0.346*	0.027	0.396	0.075	0.332*	0.036
CD4+CD28-	-0.540*	4.75e-10	-0.626*	1.20e-5	-0.222	0.333	-0.344*	0.030
CD8+CD28-	-0.269*	0.004	-0.327*	0.037	-0.369	0.100	-0.348*	0.028
CD14+ of live-gated cells	-0.112	0.235	0.033	0.836	-0.139	0.549	0.077	0.635
CD19+ of live-gated cells	na	na	0.246	0.122	na	na	-0.025	0.879

The cell counts were determined using FACS. The associations were determined using Spearman rank-sum correlation analysis; larger and smaller  $\Delta$ -cAge-DNAmAge values are referred to as 'younger' and 'older' DNA methylome age, respectively. Data not available are denoted as 'na'. Associations with a  $P$  value less than 0.05 are denoted by asterisks. Additional young control subjects ( $N = 33$ ) were recruited at follow-up in 2014

between epigenetic and chronological age) is only minor at the time of birth when compared to later stages at childhood: the divergence in epigenetic age acceleration on those children was increased in association with chronological age. Simpkin et al. (2016) concluded also that within-subject correlation between epigenetic ages at different time-points increases with increasing chronological age. Thus, these results are well in line with our conclusion that the primary trajectory of epigenetic age is set before adulthood. Interestingly, Simpkin et al.

(2016) found several phenotypes that associated with epigenetic age acceleration in the children: e.g. maternal alcohol consumption and smoking during pregnancy among others. Future long-term follow-up studies with full-scale phenotypic information from the beginning of life, through early childhood to later adulthood (e.g. middle-age) would provide more exact knowledge of epigenetic clock timing.

Several environmental factors, such as diet, physical and cognitive characteristics and toxin exposure, may

**Table 3** The associations of WBL subtype proportions with  $\Delta$ -cAge-DNAmAge values at baseline in 1986 (cAge 15–24 years) and at follow-up in 2011 (cAge 40–49 years) of the subjects in the YFS

Cell type proportion (%)	YFS, $\Delta$ -cAge-DNAmAge in 1986, $N = 183$		YFS, $\Delta$ -cAge-DNAmAge in 2011, $N = 183$	
	$r$	$P$	$r$	$P$
CD8+ T cells	-0.085	0.255	-0.089	0.231
CD4+ T cells	0.150*	0.044	0.093	0.211
CD4+/CD8+ ratio	0.106	0.160	0.224*	0.003
NK cells	-0.024	0.747	-0.192*	0.009
B cells	0.106	0.152	0.115	0.120
Monocytes	-0.028	0.711	-0.015	0.840
Granulocytes	0.029	0.700	0.054	0.465

The cell counts were determined using genome-wide DNA methylation profiles and a specific cell count estimation algorithm (Jaffe and Irizarry 2014). Associations were determined using Spearman rank-sum correlation analysis; larger and smaller  $\Delta$ -cAge-DNAmAge values are referred to as 'younger' and 'older' DNA methylome age, respectively. Associations with a  $P$  value less than 0.05 are denoted by asterisks

**Table 4** Correlations of the PBMC subtype proportions of subjects ( $N = 63$ ) in the V90 between 2010 and 2014

Cell type proportion (%)	$r$	$P$
CD3+ of live-gated cells	0.628*	2.26e-06
CD4+ of CD3+	0.830*	5.34e-13
CD8+ of CD3+	0.855*	2.13e-14
CD4+/CD8+ ratio	0.811*	7.93e-16
CD4+ CD28-	0.901*	9.48e-24
CD8+CD28-	0.838*	2.06e-13
CD14+ of live-gated cells	0.317*	3.01e-02

The cell counts of the samples were determined using FACS. Associations were determined using Spearman rank-sum correlation analysis. Associations with a  $P$  value less than 0.05 are denoted by asterisks. Correlation analysis was performed on the total sample without further categorization due to a small sample size

influence the ticking rate of the epigenetic clock (Zampieri et al. 2015; Simpkin et al. 2016). Thus, in the current study, these and various other factors may underlie the observation that some of the individuals deviated from the diagonal line in the scatterplots of their  $\Delta$ -cAge-DNAme values (Fig. 3a, b). In addition, some of the dispersion may be due to technical reasons, i.e. inaccuracies in the DNAme measurement. Although numerous factors may influence the DNAme estimate, surprisingly high correlation coefficients between inter-individual  $\Delta$ -cAge-DNAme values at the two time-points were observed in both cohorts. Therefore, our results indicate that in the majority of individuals, ageing of the DNA methylome proceeds surprisingly steadily (Fig. 3a, b).

Another important finding in our study is that the DNAme estimate significantly paralleled the ageing-related shifts in leukocyte proportions (Table 2). The most significant PBMC subtype that correlated with  $\Delta$ -cAge-DNAme in the V90 sample was CD4+CD28- cells; an increased frequency of these cells was associated with older DNA methylome age in the nonagenarians at both follow-up time points (Table 2, Fig. 4a, b). Moreover, CD4+CD28- cells were similarly associated with DNAme in the young controls (Table 2). The observation that increased proportions of immunosenescence-representing cell types (CD4+CD28- and CD8+CD28- cells) and a reduced CD4+/CD8+ cell ratio emerged as highly significant correlates of greater DNAme is of specific importance to ageing and age-related phenotypes. That is, the CD4+CD28- cell count can be a potential confounding factor in analyses of the associations between DNAme and ageing-related phenotypes. Interestingly, the association between DNAme and immune cell distribution is not restricted to ageing. It has been demonstrated that more rapid DNA methylome ageing estimate in HIV-1 patients is accompanied by increased frequencies of NK and CD28-CD34RA-CD8+ T cells and decreased frequencies of CD4+ granulocytes and naive CD8+ T cells (Horvath and Levine 2015). Notably, HIV-1 patients are likely to prematurely develop ageing-related pathologies (Pathai et al. 2014). Based on the correlation between the DNAme estimate and the NK cell count in the YFS cohort follow-up when the participants were 40–49 years of age (Table 3), it seems that the NK cell count and DNAme estimate are associated only in later adulthood and not in adolescence and early adulthood. However, identifying the basis of this observation and elucidating

**Table 5** Correlations of WBL subtype proportions for subjects in the YFS sample between 1986 and 2011

Cell type proportion (%)	YFS, all $N = 183$		YFS, women $N = 113$		YFS, men $N = 70$	
	$r$	$P$	$r$	$P$	$r$	$P$
CD8+ T cells	0.317*	1.20e-05	0.356*	1.35e-04	0.261*	2.90e-02
CD4+ T cells	0.312*	1.57e-05	0.349*	1.86e-04	0.360*	2.19e-03
CD4+/CD8+ ratio	0.453*	1.97e-10	0.464*	4.12e-07	0.511*	1.00e-05
NK cells	0.398*	2.18e-08	0.430*	2.79e-06	0.312*	8.55e-03
B cells	0.545*	1.33e-15	0.484*	8.69e-08	0.589*	8.39e-08
Monocytes	0.547*	9.73e-16	0.561*	1.79e-10	0.521*	3.71e-06
Granulocytes	0.206*	4.92e-03	0.268*	4.60e-03	0.080	0.509

The cell counts were determined using genome-wide DNA methylation profiles and a specific cell count estimation algorithm (Jaffe and Irizarry 2014). Associations were determined using Spearman rank-sum correlation analysis. Associations with a  $P$  value less than 0.05 are denoted by asterisks

the potential ageing- and immunosenescence-related methylomic changes in NK cells warrants further studies in purified NK cells.

As shown in our previous study (Marttila et al. 2015), CD4+CD28<sup>-</sup> cells appear to be an obvious confounding factor in analyses of age-related DNA methylation changes at genome-wide level in PBMCs; the proportion of CD4+CD28<sup>-</sup> cells explained a large part of the variation in DNA methylation levels of ageing-associated CpG sites. Although the mechanistic basis for the role of CD4+CD28<sup>-</sup> cells in ageing and immunosenescence is not completely clear, we speculate that certain intrinsic features of these cells, i.e. their cytokine secretion profile, shortened telomeres and resistance to apoptosis (Weiskopf et al. 2009; Arnold et al. 2011), might also be related to alterations in DNA methylation patterns in the ageing-associated sites as well as in the epigenetic clock sites. In addition to ageing, expansion of these cells has been reported in certain immune-related diseases, such as multiple sclerosis, rheumatoid arthritis and acute coronary syndromes (Broux et al. 2012). Hence, our findings may have implications on the methylomic analyses of these disorders as well.

Previous studies have shown that DNA methylomes of men age more rapidly than those of women (Hannum et al. 2013; Weidner et al. 2014). Thus, the results were interpreted according to gender. Although our subsample sizes of men and women were quite small for statistical analysis, a general trend of gender difference was observed in the DNAmAges of PBMCs in the nonagenarians and the young controls at both time points of the V90: men had older DNA methylomes than women (Fig. 2a, b); the difference was statistically significant in the V90\_2010 cohort where the subsample sizes were sufficient. In the WBL samples from the YFS participants, no significant gender differences in the DNA methylome ages were observed in any cAge category (Table 1 and Figs. 1a, b). A reason for this negative result may have been the small sample size of the different cAge groups or variability of the blood sample types. However, in the YFS\_1986 cohort, there was a slight trend in which 15–24-year-old men and women displayed unique, but constant trajectories of DNA methylome ageing estimate as a function of cAge (i.e. the slope of the trend line of this change was different between genders); women aged faster than men, as shown in Fig. 1a. Nevertheless, the overall ageing rate of DNA methylomes (i.e. how many years the DNA

methylome had aged over the 4- or 25-year follow-up period) did not differ between women and men in either the V90 or the YFS. Further studies using larger samples are needed to clarify these observations.

Analogous to previous findings (Horvath 2013; Marioni et al. 2015a), the DNAmAge values correlated with cAge, showing high correlation coefficients (Spearman's rho >0.5) and high statistical significance ( $P < 4 \times 10^{-6}$ ) in both of our follow-up studies. Interestingly, as shown in Table 1 and Supplementary Table 1 (Additional File 2) as well as in Figs. 1 and 2, the median DNAmAges were often smaller than the corresponding median cAges; thus, the  $\Delta$ -cAge-DNAmAge values were positive in the majority of samples, as shown in Figs. 3a, b and Supplementary Figs. 1–4 (Additional File 2). This result may be interpreted as an indicator of a 'young' DNA methylome age of blood cells (on both PBMC and WBL). Nevertheless, some of the young controls in the V90 exhibited older DNA methylomes in their PBMCs (Supplementary Fig. 4 in Additional File 2, Table 1). Analysis of the follow-up data revealed that the DNA methylomes in WBLs of the YFS subjects had aged by an average of ~23 years during the 25-year follow-up period; this rate corresponds to 3.7 years of methylome ageing in 4 calendar years. During the 4-year-follow-up in the V90 cohort, the DNA methylomes in PBMCs of the individuals aged 19–34 years appeared to age faster than those of nonagenarians (~7.4 years compared to ~4.5 years). These results may refer to the slightly varying rates of DNA methylome ageing in PBMCs between different stages of the human lifespan. However, our results also indicate that the DNA methylome ageing rate is rather stable after young adulthood, even at extreme ages (Figs. 3a, b).

Previous studies have shown that there are non-pathologic tissue-specific differences in DNAmAge. For example, the DNA methylome of the cerebellum ages more slowly (Horvath et al. 2015b) and the DNA methylome of sperm is considered much younger (Horvath 2013) than that in other parts of the human body. Studies of the within-individual differences in  $\Delta$ -cAge-DNAmAge values between leukocytes and other healthy tissues using reasonable sample sizes are lacking. Few reports have assessed the estimated blood cell subtypes in their DNAmAge analyses (Horvath 2013; Marioni et al. 2015a) and concluded that DNAmAge is relatively independent of the fluctuation in the predominant blood cell subtypes. Nevertheless, these studies

have assessed only the major blood cell categories, and most importantly, these studies have not examined very old individuals.

Our results demonstrate in both of our follow-up cohorts that (i) the DNAmAge is associated with the cell subtype distributions (Tables 2 and 3); (ii) there are highly significant intra-individual correlations of the blood cell subtype distributions during several years of follow-ups (Tables 4 and 5, Fig. 4c) and (iii) the difference between DNAmAge estimate and cAge is largely unchanged during the follow-ups (Fig. 3a, b). Our data do not permit us to elucidate why the proportions of the specific blood cell types are so strongly associated with the DNAmAge estimate, but the fact that the blood cell composition changes with age is also evident in our results (Table 1). This issue may be one of the major reasons for average DNA methylation differences in PBMCs over age. However, our results do not indicate a causal relationship between the ageing-associated changes in cell composition and the actual DNA methylation changes in specific loci per cell passage.

Our study contains evident limitations due to limited sample sizes. In addition, we did not obtain flow cytometry data for the proportions of minor cell subtypes aside from CD4+CD28<sup>-</sup> and CD8+CD28<sup>-</sup> cells. Furthermore, the cell counts of the YFS samples were determined using a prediction algorithm that relies on genome-wide cell-type-specific methylation profiles (Jaffe and Irizarry 2014). Moreover, the DNA samples were of slightly different origin: YFS sample DNA was extracted from WBLs, and V90 DNA was extracted from PBMCs. Nevertheless, we detected analogous results from the two independent follow-up cohorts examined in this study.

## Conclusions

In summary, our results demonstrate that most of the differences in estimated DNAmAge arise before adulthood and that the estimated DNAmAge changing rate is rather stable thereafter. We also provide evidence that DNAmAge estimate is tightly associated with blood cell subtype frequencies, especially with the immunosenescence-reflecting markers, the CD4+CD28<sup>-</sup> and CD8+CD28<sup>-</sup> cells and the CD4+/CD8+ ratio. However, it remains to be determined how and exactly when before adulthood the inter-individual differences in the DNAmAge arise. Moreover, other very

important questions remain to be answered: to what extent is it possible to alter the ticking rate of the epigenetic clock by implementing lifestyle changes (diet, physical activity, etc.) during adulthood? Future studies are needed to address these questions.

## Methods

### Study populations

#### *The Young Finns study*

The YFS consisted of a series of cohorts representing the general population born in 1962, 1965, 1968, 1971, 1974 and 1977 from five cities with university hospitals in Finland (Helsinki, Kuopio, Oulu, Tampere and Turku) (Raitakari et al. 2008). A subsample of 183 individuals was randomly selected for this study from a follow-up study performed in 1986 (denoted as the YFS\_1986 sample), and the same individuals were analysed again from a follow-up study performed 25 years later in 2011 (denoted as the YFS\_2011 sample). The sample collections in 1986 and 2011 are described in further detail elsewhere (Akerblom et al. 1989; Nuotio et al. 2014). At baseline in 1986, the sample examined in the current study contained four age groups: 15-, 18-, 21- and 24-year-old subjects. The respective group sizes were 49, 44, 55 and 35, and the respective percentage of women was 59.2, 68.2, 56.4 and 60.0 %. All of the participants were of western European descent. This study followed the guidelines of the Declaration of Helsinki and was approved by the Ethical Review Committee of Turku University Hospital, and all participants provided informed consent.

#### *The Vitality 90+ study*

The Vitality 90+ study (V90) is an on-going prospective population-based study that includes both home-dwelling and institutionalized subjects aged 90 years or older who live in the city of Tampere, Finland. The recruitment and characterization of the subjects were performed as reported for previous V90 cohorts (Goebeler et al. 2003). In the DNA methylation analysis, the study population at baseline in 2010 (denoted as V90\_2010) consisted of 122 subjects born in 1920 (women  $N = 89$ , men  $N = 33$ ) and 21 young control subjects (19–28-year-old women  $N = 14$ , 20–29-year-

old men  $N = 7$ ). Of these subjects, 48 (control subjects  $N = 7$  and 94-year-old surviving subjects  $N = 41$ ; denoted as the V90\_2014 sample) participated in a 4-year follow-up examination. In addition, 33 new control subjects (22–33-year-old men  $N = 9$ , 19–34-year-old women  $N = 24$ ) were recruited together with the 4-year follow-up participants; therefore, the full sample size of the V90\_2014 sample was 81. In total, 223 DNA samples from 176 subjects were used in the DNAmAge analysis. In addition, 136 subjects in the V90\_2010 sample (95 % of the total sample size; young  $N = 21$  and old  $N = 115$ ) and 81 subjects in the V90\_2014 sample (100 % of the total sample size; young  $N = 40$  and old  $N = 41$ ) were examined for PBMC subtype composition characterization via flow cytometry (FACS); PBMC samples from 63 subjects (51 nonagenarians and 12 young controls) were characterized using FACS at both time points, at baseline in 2010 and at follow-up in 2014. Further, of these subjects, 48 (the same as those mentioned above) were used for DNAmAge measurement at both time points. All study subjects were of Western European descent and had not suffered from any infections or received any vaccinations in the 30 days prior to blood sample collection. The study participants provided their written informed consent. This study has been conducted according to the principles expressed in the Declaration of Helsinki, and the study protocol was approved by the ethics committee of the city of Tampere (1592/403/1996; 765/13.03.01/2008, PSHP 7/2014, ETL R14002).

### Sample preparation

WBL DNA of the YFS cohort at baseline and at the 25-year follow-up was obtained from blood samples stored in EDTA using a Wizard<sup>®</sup> Genomic DNA Purification Kit (Promega Corporation, Madison, WI, USA) according to the manufacturer's instructions. The blood samples of the subjects in the V90 at baseline and at follow-up were collected into EDTA-containing tubes during a home visit. The samples were directly subjected to leucocyte separation on a Ficoll-Paque density gradient (Ficoll-Paque<sup>™</sup> Premium, cat. no. 17-5442-03, GE Healthcare Bio-Sciences AB, Uppsala, Sweden). The PBMC layer was collected and suspended in 1 ml of a freezing solution (5/8 FBS, 2/8 RPMI-1640 medium, 1/8 DMSO) (FBS: cat. no. F7524, Sigma-Aldrich, St. Louis, MO, USA; RPMI: cat. no. R0883, Sigma-Aldrich, St. Louis, MO, USA; DMSO: cat. no.

1.02931.0500, VWR, Espoo, Finland) and stored in liquid nitrogen. DNA was extracted from PBMCs using the QIAamp DNA Mini Kit (Qiagen, CA, USA) according to the centrifugation protocol provided in the manufacturer's instructions. The DNA was eluted in 60  $\mu$ l of AE elution buffer and stored at  $-20$  °C. The concentration and the quality of the DNA were assessed via the Qubit dsDNA HS Assay (Invitrogen, Eugene, OR, USA).

### DNAmAge quantification

The genome-wide DNA methylation levels in WBLs from the YFS\_2011 individuals and in PBMCs from the V90 participants were obtained using Illumina Infinium HumanMethylation450 BeadChips (Illumina, San Diego, CA, USA) (Bibikova et al. 2006; Bibikova et al. 2009; Bibikova et al. 2011) according to the manufacturer's protocol at the Core Facility of the Institute of Molecular Medicine Finland (FIMM), University of Helsinki. The corresponding DNA methylation profiles of the YFS\_1986 samples were measured using the same methodology at Helmholtz Zentrum, München, Germany. Samples were applied to the arrays in a randomized order. Aliquots of 1  $\mu$ g of DNA were subjected to bisulphite conversion using the EZ-96 DNA Methylation Kit (Zymo Research, Irvine, CA, USA) according to the manufacturer's instructions. A 4- $\mu$ l aliquot of bisulphite-converted DNA was subjected to whole-genome amplification, followed by enzymatic fragmentation and hybridization to an Infinium HumanMethylation450 BeadChip. The BeadChips were scanned with the iScan reader (Illumina), and the measured probe intensities were transformed to  $\beta$  values using a standard equation in which  $\beta$  is the ratio of the methylated probe ( $m$ ) intensities to the overall intensities ( $m + u + \alpha$ , where  $\alpha$  is the constant offset, 100, and  $u$  is the unmethylated probe intensity). Thus, in the heterogenic sample, the resulting  $\beta$  values ranged linearly from 0 (completely unmethylated, 0 %) to 1 (completely methylated, 100 %). The quality of the DNA samples and the methylation data was carefully ensured using standard examinations, such as principal component analysis (PCA), gender prediction and visualizations. The methylation intensity values and the PCA results were viewed using density plots, boxplots and dotplots. The gender

prediction was based on the average levels of DNA methylation at CpG sites located in X chromosomes. Finally, the absolute methylation values ( $\beta$  values) of the selected probes ( $N = 28,587$ ) were extracted from all data sets, and those values were used as the input for calculation of the DNAmAge (<https://dnamage.genetics.ucla.edu/home>) (Horvath 2013). The methylation data were inputted into the calculator in a blinded manner, i.e. without any preliminary background information concerning sample type, cAge or gender. Batch effects were normalized using the BMIQ function implemented in the DNAmAge algorithm.

### FACS analysis of the V90 samples

The proportions of different leukocyte subtypes were determined with randomized sample orders using fluorescence-activated cell sorting analysis (FACS; BD FACSCanto II). In 2010, the results were analysed with BD FACSDiva, version 6.1.3 (BD Biosciences, Franklin Lakes, NJ, USA), and in 2014, the results were analysed with FlowJo software (Tree Star Inc., Ashland, OR, USA). The antibodies used were FITC-CD14 (cat. no. 11-0149), PerCP-Cy5.5-CD3 (45-0037), APC-CD28 (17-0289), PE-CD19 (12-0199) (eBioscience, San Diego, CA, USA), PE-Cy<sup>TM</sup>7-CD4 (cat. no. 557852) and APC-Cy<sup>TM</sup>7-CD8 (557834) (BD Biosciences). In 2010, CD19 antibody was not used. Staining was performed in phosphate-buffered saline (PBS) containing 1 % foetal bovine serum (FBS) after an incubation step with Fc Receptor Binding Inhibitor (cat. no 16-9161, eBioscience) to minimize non-specific staining of the cells.

### Cell count estimates of the YFS samples

The leukocyte cell type distributions in the YFS samples were determined using R software ( $R \geq 2.15.3$ ) and the estimation algorithm implemented in the `estimateCellCounts` function of the `minfi` Bioconductor package (Jaffe and Irizarry 2014). This algorithm utilizes a subset of 600 control probes in the 450 BeadChip, which represent specific DNA methylation signatures of CD8<sup>+</sup> and CD4<sup>+</sup> T cells, monocytes, granulocytes, NK cells and B cells. The reference data used in the estimation algorithm are available in the `FlowSorted.Blood.450K` Bioconductor package.

### Statistical analyses

Descriptive statistical analyses of the DNAmAges included calculations of medians and MADs. Between-group comparisons were conducted using non-parametric Mann-Whitney  $U$  tests (two-sided Wilcoxon's test), and the correlation analyses were performed using non-parametric Spearman's rank sum tests. The nominal  $P$  value threshold was set to 0.05. Analyses and visualizations were performed using R software ( $R \geq 2.15.3$ ) and IBM SPSS Statistics v.22 (IBM Corporation, Armonk, NY, USA).

**Acknowledgments** The YFS has been financially supported by the Academy of Finland (grants 286284 (T.L.), 132704 (M.H.), 134309 (Eye), 126925, 121584, 124282, 129378 (Salve), 117787 (Gendi), 41071 (Skidi) and 250602 (Coctel to M.J.); the Social Insurance Institution of Finland; Kuopio, Tampere and Turku University Hospital Medical Funds (grant X51001 for T.L.); the Juho Vainio Foundation; the Paavo Nurmi Foundation; the Finnish Foundation of Cardiovascular Research (T.L.); the Tampere Tuberculosis Foundation (T.L., M.H., I.J.); the Emil Aaltonen Foundation (T.L., L.Ku.); the Finnish Cultural Foundation, Pirkanmaa Regional fund (N.M); and the Yrjö Jahnsson Foundation (T.L., M.H.) This work was also supported by grants from the Competitive Research Fund of Pirkanmaa Hospital District (9M017, 9N013 to M.H. and 9P002 to M.J.), Sigrid Juselius Foundation (I.J.), Finnish Medical Association (I.J.) and Competitive Research Fund of Fimlab Laboratories (X51409, I.J.). We thank Nina Peltonen, Sinikka Repo-Koskinen, Eeva Timonen, Janette Hinkka and Kristiina Lehtinen for their excellent technical assistance.

**Authors' contributions** LK processed the data, performed the analyses and was responsible for writing the manuscript. LK, SM, JJ, LKu, NM and TN were responsible for performing the experiments. MH, MK, IJ and TL provided reagents and materials. LK, SM, TL, JJ and MH contributed to the design of the study. MH, MJ, AH, OTR, MK and TL were responsible for recruiting the subjects for the study. All authors contributed to the writing of the manuscript and read and approved the final manuscript.

### Compliance with ethical standards

**Competing interests** The authors declare that they have no competing interests.

### References

- Akerblom HK, Viikari J, Rasanen L, Kuusela V, Uhari M, Lautala P (1989) Cardiovascular risk in young Finns, results from the second follow-up study. *Ann Med* 21:223–225
- Arnold CR, Wolf J, Brunner S, Herndler-Brandstetter D, Grubeck-Loebenstien B (2011) Gain and loss of T cell subsets in old



- age-age-related reshaping of the T cell repertoire. *J Clin Immunol* 31:137–146. doi:10.1007/s10875-010-9499-x
- Benetos A, Okuda K, Lajemi M, Kimura M, Thomas F, Skurnick J, Labat C, Bean K, Aviv A (2001) Telomere length as an indicator of biological aging: the gender effect and relation with pulse pressure and pulse wave velocity. *Hypertension* 37:381–385
- Bibikova M, Barnes B, Tsan C, Ho V, Klotzle B, Le JM, Delano D, Zhang L, Schroth GP, Gunderson KL, et al. (2011) High density DNA methylation array with single CpG site resolution. *Genomics* 98:288–295. doi:10.1016/j.ygeno.2011.07.007
- Bibikova M, Le J, Barnes B, Saedinia-Melnyk S, Zhou L, Shen R, Gunderson KL (2009) Genome-wide DNA methylation profiling using Infinium(R) assay. *Epigenomics* 1:177–200. doi:10.2217/epi.09.14
- Bibikova M, Lin Z, Zhou L, Chudin E, Garcia EW, Wu B, Doucet D, Thomas NJ, Wang Y, Vollmer E, et al. (2006) High-throughput DNA methylation profiling using universal bead arrays. *Genome Res* 16:383–393
- Bocklandt S, Lin W, Sehl ME, Sanchez FJ, Sinsheimer JS, Horvath S, Vilain E (2011) Epigenetic predictor of age. *PLoS One* 6:e14821. doi:10.1371/journal.pone.0014821
- Boks MP, van Mierlo HC, Rutten BP, Radstake TR, De Witte L, Geuze E, Horvath S, Schalkwyk LC, Vinkers CH, Broen JC, et al. (2015) Longitudinal changes of telomere length and epigenetic age related to traumatic stress and post-traumatic stress disorder. *Psychoneuroendocrinology* 51:506–512. doi:10.1016/j.psyneuen.2014.07.011
- Broux B, Markovic-Plese S, Stinissen P, Hellings N (2012) Pathogenic features of CD4+CD28- T cells in immune disorders. *Trends Mol Med* 18:446–453. doi:10.1016/j.molmed.2012.06.003
- Florath I, Butterbach K, Muller H, Bewerunge-Hudler M, Brenner H (2014) Cross-sectional and longitudinal changes in DNA methylation with age: an epigenome-wide analysis revealing over 60 novel age-associated CpG sites. *Hum Mol Genet* 23:1186–1201. doi:10.1093/hmg/ddt531
- Garagnani P, Bacalini MG, Pirazzini C, Gori D, Giuliani C, Mari D, Di Blasio AM, Gentilini D, Vitale G, Collino S, et al. (2012) Methylation of ELOVL2 gene as a new epigenetic marker of age. *Aging Cell* 11:1132–1134. doi:10.1111/accel.12005
- Goebeler S, Jylha M, Hervonen A (2003) Medical history, cognitive status and mobility at the age of 90. A population-based study in Tampere, Finland. *Aging Clin Exp Res* 15:154–161
- Hannum G, Guinney J, Zhao L, Zhang L, Hughes G, Sadda S, Klotzle B, Bibikova M, Fan J, Gao Y, et al. (2013) Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol Cell* 49:359–367. doi:10.1016/j.molcel.2012.10.016
- Horvath S (2013) DNA methylation age of human tissues and cell types. *Genome Biol* 14:R115
- Horvath S, Levine AJ (2015) HIV-1 infection accelerates age according to the epigenetic clock. *J Infect Dis* 212:1563–1573
- Horvath S, Erhart W, Brosch M, Ammerpohl O, von Schonfels W, Ahrens M, Heits N, Bell JT, Tsai PC, Spector TD, et al. (2014) Obesity accelerates epigenetic aging of human liver. *Proc Natl Acad Sci U S A* 111:15538–15543. doi:10.1073/pnas.1412759111
- Horvath S, Garagnani P, Bacalini MG, Pirazzini C, Salvioli S, Gentilini D, Di Blasio AM, Giuliani C, Tung S, Vinters HV, et al. (2015a) Accelerated epigenetic aging in Down syndrome. *Aging Cell* 14:491–495. doi:10.1111/accel.12325
- Horvath S, Mah V, Lu AT, Woo JS, Choi OW, Jasinska AJ, Riancho JA, Tung S, Coles NS, Braun J, et al. (2015b) The cerebellum ages slowly according to the epigenetic clock. *Aging (Albany NY)* 7:294–306
- Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, Wiencke JK, Kelsey KT (2012) DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinf* 13:86. doi:10.1186/1471-2105-13-86
- Jaffe AE, Irizarry RA (2014) Accounting for cellular heterogeneity is critical in epigenome-wide association studies. *Genome Biol* 15:R31. doi:10.1186/gb-2014-15-2-r31
- Kananen L, Marttila S, Nevalainen T, Jylhava J, Mononen N, Kahonen M, Raitakari OT, Lehtimaki T, Hurme M (2016) Aging-associated DNA methylation changes in middle-aged individuals: the Young Finns study. *BMC Genomics* 17:103. doi:10.1186/s12864-016-2421-z
- Lam LL, Emberly E, Fraser HB, Neumann SM, Chen E, Miller GE, Kobor MS (2012) Factors underlying variable DNA methylation in a human community cohort. *Proc Natl Acad Sci U S A* 109(Suppl 2):17253–17260. doi:10.1073/pnas.112124910
- Marioni RE, Shah S, McRae AF, Chen BH, Colicino E, Harris SE, Gibson J, Henders AK, Redmond P, Cox SR, et al. (2015a) DNA methylation age of blood predicts all-cause mortality in later life. *Genome Biol* 16:25. doi:10.1186/s13059-015-0584-6
- Marioni RE, Shah S, McRae AF, Ritchie SJ, Muniz-Terrera G, Harris SE, Gibson J, Redmond P, Cox SR, Pattie A, et al. (2015b) The epigenetic clock is correlated with physical and cognitive fitness in the Lothian Birth Cohort 1936. *Int J Epidemiol* 44:1388–1396
- Marttila S, Kananen L, Hayrynen S, Jylhava J, Nevalainen T, Hervonen A, Jylha M, Nykter M, Hurme M (2015) Ageing-associated changes in the human DNA methylome: genomic locations and effects on gene expression. *BMC Genomics* 16. doi:10.1186/s12864-015-1381-z
- Mather KA, Jorm AF, Parslow RA, Christensen H (2011) Is telomere length a biomarker of aging? A review. *J Gerontol A Biol Sci Med Sci* 66:202–213. doi:10.1093/gerona/gql180
- Nuotio J, Oikonen M, Magnussen CG, Jokinen E, Laitinen T, Hutri-Kahonen N, Kahonen M, Lehtimaki T, Taittonen L, Tossavainen P, et al. (2014) Cardiovascular risk factors in 2011 and secular trends since 2007: the Cardiovascular Risk in Young Finns Study. *Scand J Public Health* 42:563–571. doi:10.1177/1403494814541597
- Pathai S, Bajillan H, Landay AL, High KP (2014) Is HIV a model of accelerated or accentuated aging? *J Gerontol Ser A Biol Med Sci* 69:833–842. doi:10.1093/gerona/glt168
- Pawelec G, Larbi A, Derhovanessian E (2010) Senescence of the human immune system. *J Comp Pathol* 142(Suppl 1):S39–S44. doi:10.1016/j.jcpa.2009.09.005
- Raitakari OT, Juonala M, Ronnema T, Keltikangas-Jarvinen L, Rasanen L, Pietikainen M, Hutri-Kahonen N, Taittonen L, Jokinen E, Marniemi J, et al. (2008) Cohort profile: the cardiovascular risk in Young Finns Study. *Int J Epidemiol* 37:1220–1226. doi:10.1093/ije/dym225
- Simpkin AJ, Hemani G, Suderman M, Gaunt TR, Lyttleton O, Mcardle WL, Ring SM, Sharp GC, Tilling K, Horvath S, et al. (2016) Prenatal and early life influences on epigenetic age in

- children: a study of mother-offspring pairs from two cohort studies. *Hum Mol Genet* 25:191–201. doi:[10.1093/hmg/ddv456](https://doi.org/10.1093/hmg/ddv456)
- Steegenga WT, Boekschoten MV, Lute C, Hooiveld GJ, de Groot PJ, Morris TJ, Teschendorff AE, Butcher LM, Beck S, Muller M (2014) Genome-wide age-related changes in DNA methylation and gene expression in human PBMCs. *Age (Dordr)* 36:9648. doi:[10.1007/s11357-014-9648-x](https://doi.org/10.1007/s11357-014-9648-x)
- Weidner CI, Lin Q, Koch CM, Eisele L, Beier F, Ziegler P, Bauerschlag DO, Jockel KH, Erbel R, Muhleisen TW, et al. (2014) Aging of blood can be tracked by DNA methylation changes at just three CpG sites. *Genome Biol* 15:R24. doi:[10.1186/gb-2014-15-2-r24](https://doi.org/10.1186/gb-2014-15-2-r24)
- Weiskopf D, Weinberger B, Grubeck-Loebenstien B (2009) The aging of the immune system. *Transpl Int* 22:1041–1050. doi:[10.1111/j.1432-2277.2009.00927.x](https://doi.org/10.1111/j.1432-2277.2009.00927.x)
- Zampieri M, Ciccarone F, Calabrese R, Franceschi C, Burkle A, Caiafa P (2015) Reconfiguration of DNA methylation in aging. *Mech Ageing Dev* 151:60–70
- Zilbauer M, Rayner TF, Clark C, Coffey AJ, Joyce CJ, Palta P, Palotie A, Lyons PA, Smith KG (2013) Genome-wide methylation analyses of primary human leukocyte subsets identifies functionally important cell-type-specific hypomethylated regions. *Blood* 122:e52–e60. doi:[10.1182/blood-2013-05-503201](https://doi.org/10.1182/blood-2013-05-503201)

# Methylomic predictors demonstrate the role of NF- $\kappa$ B in old-age mortality and are unrelated to the aging-associated epigenetic drift

Juulia Jylhävä<sup>1,2</sup>, Laura Kananen<sup>1,2</sup>, Jani Raitanen<sup>3,4</sup>, Saara Marttila<sup>1,2</sup>, Tapio Nevalainen<sup>1,2</sup>, Antti Hervonen<sup>2,3</sup>, Marja Jylhä<sup>2,3</sup> and Mikko Hurme<sup>1,2,5</sup>

<sup>1</sup> Department of Microbiology and Immunology, School of Medicine, University of Tampere, Tampere, Finland

<sup>2</sup> Gerontology Research Center, University of Tampere, Tampere, Finland

<sup>3</sup> School of Health Sciences, University of Tampere, Tampere, Finland

<sup>4</sup> UKK Institute for Health Promotion Research, Tampere, Finland

<sup>5</sup> Fimlab Laboratories, Tampere, Finland

**Correspondence to:** Juulia Jylhävä, **email:** juulia.jylhava@uta.fi

**Keywords:** methylation, mortality, aging, longevity, Cox model, Gerotarget

**Received:** November 06, 2015

**Accepted:** March 10, 2016

**Published:** March 22, 2016

## ABSTRACT

**Changes in the DNA methylation (DNAm) landscape have been implicated in aging and cellular senescence. To unravel the role of specific DNAm patterns in late-life survival, we performed genome-wide methylation profiling in nonagenarians ( $n=111$ ) and determined the performance of the methylomic predictors and conventional risk markers in a longitudinal setting. The survival model containing only the methylomic markers was superior in terms of predictive accuracy compared with the model containing only the conventional predictors or the model containing conventional predictors combined with the methylomic markers. At the 2.55-year follow-up, we identified 19 mortality-associated (false-discovery rate  $<0.5$ ) CpG sites that mapped to genes functionally clustering around the nuclear factor kappa B (NF- $\kappa$ B) complex. Interestingly, none of the mortality-associated CpG sites overlapped with the established aging-associated DNAm sites. Our results are in line with previous findings on the role of NF- $\kappa$ B in controlling animal life spans and demonstrate the role of this complex in human longevity.**

## INTRODUCTION

The influential role of genomic factors, such as DNA methylation (DNAm) in the course of development, aging and age-related pathologies is well established. Several studies have also reproducibly demonstrated that the level of methylation at specific CpG sites changes as a function of age [1-5], hence providing a marker of chronological and, potentially, biological age. An intriguing characteristic of age-related DNAm signatures is that many of the age-associated DNAm changes have been observed to be common in several different tissues, such as whole blood, brain, lung and cervix [1, 3, 6]. These observations suggest that a global mechanism(s) might be responsible for age-associated modifications in the epigenetic landscape. Nevertheless, studies with monozygotic twins have demonstrated that the rate of

divergence in methylomic patterns increases with age [7, 8], suggesting that the age-related modifications in DNAm are also subject to various environmental, stochastic and life style-related effects.

However, the consequences of the aging-accompanied DNAm alterations for late-life health and functional abilities are largely unknown. A recent epigenome-wide association study (EWAS) demonstrated that the association between age-related DNAm changes and healthy aging phenotypes in individuals 32-80 years of age is negligible [8]. The results of this study also reveal that the DNAm regions associated with aging phenotypes are distinct from those associated with chronological age. These findings suggest that the CpG sites involved in health-related outcomes in later life are largely regulated by sites other than the established age-related DNAm regions [8]. In addition, using an EWAS

approach, we have recently demonstrated that the CpG sites that are associated with aging-related inflammation, i.e., inflammaging [9] are largely different from the sites associated with age [5]. This phenomenon is also observable in regard to gene expression profiles and old age mortality. We have previously demonstrated that the genes exhibiting aging-related changes in expression levels are predominantly different from those that predict mortality in late life [10]. These findings underscore the complexity and unknown nature of the genomic factors that control the human health span and late-life events.

Nevertheless, the mortality-predicting genes in our previous study were found to be functionally connected to the nuclear factor kappa B (NF- $\kappa$ B) complex, which is a central mediator in immunoinflammatory responses and has been advocated as the culprit in aging and cellular senescence (reviewed in [11]). Aberrant activation of NF- $\kappa$ B has been reported in various age-associated conditions, such as neurodegeneration, immunosenescence, inflammaging, sarcopenia and osteoporosis (reviewed in [12-14]), whereas studies involving mouse models have observed that NF- $\kappa$ B activation is a key determinant of accelerated aging and longevity [15, 16]. In the mouse models, it was demonstrated that the hypothalamic activation of NF- $\kappa$ B is a driving force of systemic aging through immune-endocrine connections [16].

Life span regulation in humans is a multifactorial process, and very little is known about the genomic determinants that control late-life mortality after the ages of the common killers, i.e., cardiovascular events and cancer, have passed. In this study, we sought to explore how the human genome-wide methylome is associated with old-age survival within a shorter (2.55 years) and a longer (4 years) follow-up time. A large panel of traditional (bio)markers and mortality risk factors was assessed alongside the methylomic markers to elucidate the relationship between the aging-related biophysiological changes and epigenetics.

## RESULTS

The characteristics of the study population and distribution of the variables in the population with methylation data available ( $n = 111$ ) are presented in Table 1. The variables (i.e., the conventional markers) exhibiting significant ( $p < 0.05$ ) univariate and multivariate associations at the 2.55 follow-up are presented in Supplementary Table 1. The predictors remaining in the multivariate model, body mass index (BMI) and Mini-Mental State Examination (MMSE) test score, were used as the model factors in the assessment of the predictive accuracy of modeling (see Methods). The measure of “epigenetic clock” [17], the DNA methylation age was not predictive of mortality in our cohort ( $p = 0.733$ ).

In the Cox univariate assessment, 19,621 and 15,505

CpG sites were associated with mortality ( $p < 0.05$ ) in the 2.55-year and 4-year follow-up data, respectively (Supplementary Tables 2 and 3). After B-H correction (FDR  $< 0.5$ ), 19 CpG sites remained significant for the 2.55-year follow-up and 7 CpG sites for the 4-year follow-up data (Supplementary Tables 2 and 3). The Ingenuity Pathway Analysis (IPA)-generated network from the 16 known genes harboring the 19 significant CpG sites at the 2.55-year follow-up is presented in Figure 1a. This network displayed NF- $\kappa$ B as a central node and involved 10 of 16 of the genes mapped to the 19 mortality-associated CpG sites (FDR  $< 0.5$ ). We also ran the IPA network and pathway analyses from the genes harboring the 250 top-ranking CpG sites according to the 2.55-year follow-up data (sites presented in Supplementary Table 2). The highest-ranking network in this analysis also placed NF- $\kappa$ B as a central complex (Figure 1b). The significant B-H-corrected canonical pathways from this data set are presented in Table 2. At the 4-year follow-up, the functional implications of the methylomic predictors were attenuated as no significant B-H-corrected canonical pathways were identified in IPA from the genes harboring the 250 highest-ranking CpG sites and no significant network enrichment was observed among the genes harboring the 7 CpG sites (FDR  $< 0.5$ ).

Assessment of the predictive accuracy of the tested models revealed that the Ridge regression containing only the methylomic markers (Ridge1) performed better than the other models; i.e., a model containing only the conventional predictors, a Ridge regression model containing both the conventional predictors (Ridge2) and the methylomic markers and a model containing only the methylomic markers selected on the basis of their significance level in Cox univariate assessment. Specifically, the methylomic markers alone exhibited the smallest median deviance from the null model (Supplementary Figure 1), and were thus used in assessing the final mortality-predicting signature in the Cox multivariate model for 2.55-year follow-up data. The deviances of the conventional markers exhibited clearly the smallest variation but their median was nevertheless higher than that of the methylomic markers in Ridge1.

The Ridge regression-organized 19 methylomic markers entered to the Cox multivariate model are presented in Supplementary Table 4. Inclusion of the methylomic markers in the final model was based on selection of the model with the best goodness of fit (Akaike Information criterion, AIC), which for the selected model was 239.0. The final Cox multivariate model is presented in Table 3 and the distributions of the beta values for the seven CpG sites (batch effect-corrected) included this mortality-predicting signature are presented in Supplementary Figure 2.

The discriminative power (Harrell's  $C$ ) for this model was 89.9%. The proportionality assumption in the

**Table 1: Characteristics of the study population (n = 111).** Distributions of the variables are presented according to the data at the 2.55-years mortality follow-up.

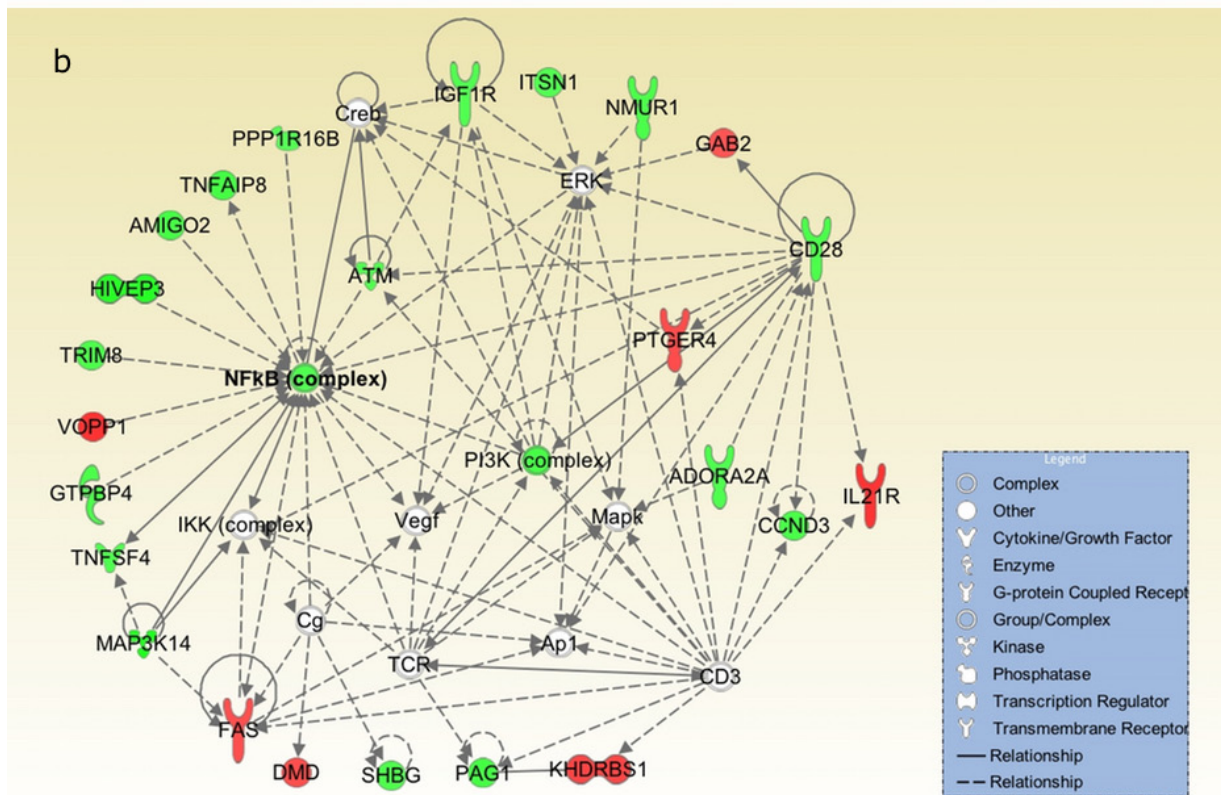
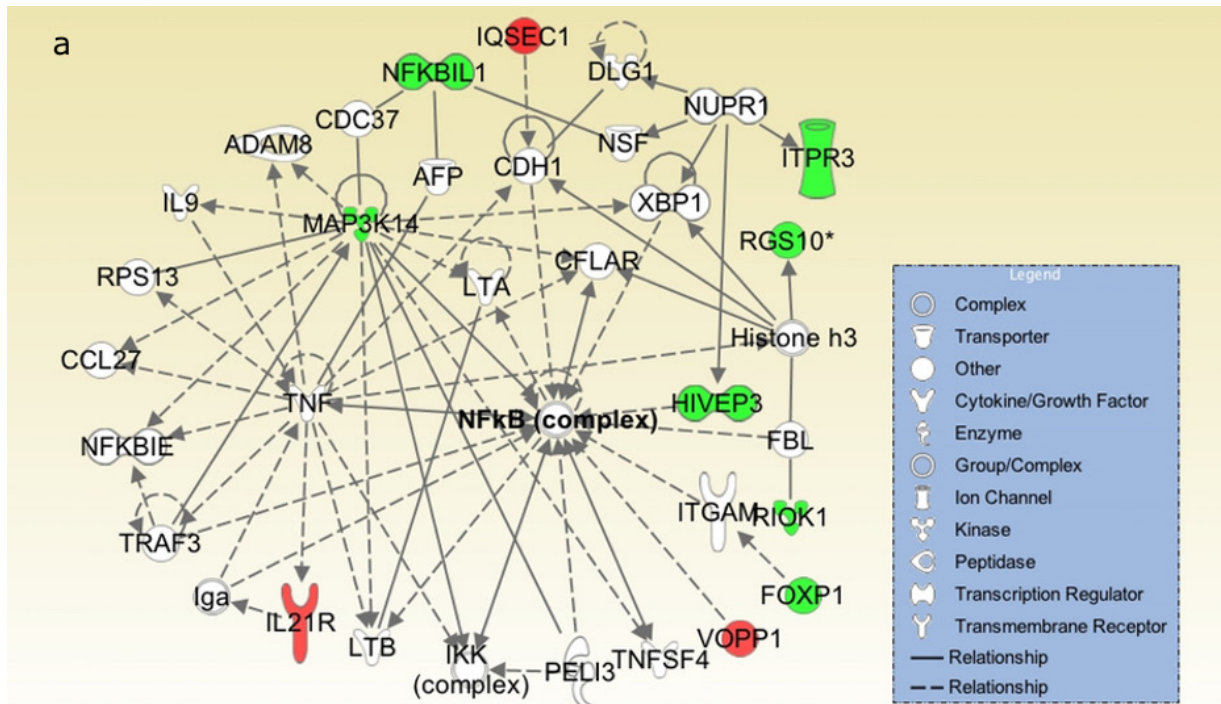
Variable	Non-survivors		Survivors	
	Mean/Med	SEM/IQR/%	Mean/Med	SEM/IQR/%
Women (n/%)	27	75.0	54	72.0
Age (months)	1079.5	0.61	1080.2	0.37
Systolic blood pressure (mmHg)	145	4.6	149	3.4
Diastolic blood pressure (mmHg)*	71.5	13.5	74.0	19.0
Weight (kg)	61.9	2.2	70.6	1.6
BMI (kg/m <sup>2</sup> )	24.3	0.75	27.5	0.54
Waist circumference (cm)	89.6	2.1	95.5	1.4
Hip circumference (cm)*	98	10.0	102	12.0
MMSE*	23.5	8.0	26.0	4.0
Barthel index*	95.0	20.0	95	5.0
Handgrip (kg)*	18.0	11.0	20.0	7.0
Able to perform chair-rise test (n = yes/%)	19	57.6	59	78.7
Able to perform chair-stand test (n = yes/%)	22	71.0	62	82.7
Frailty index (n/%)				
Non-frail	3	8.3	26	34.7
Pre-frail	22	61.1	37	49.3
Frail	11	30.6	12	16.0
CRP level (ng/ml)*	1.8	3.3	1.9	3.5
IL-1 $\beta$ level (pg/ml)*	14.2	27.6	19.0	34.0
IL-6 level (pg/ml)*	4.5	3.3	3.8	3.8
IL-7 level (pg/ml)*	7.8	5.3	6.4	5.2
IL-10 level (pg/ml)*	1.8	1.5	1.5	2.6
cf-DNA level ( $\mu$ g/ml)*	0.93	0.19	0.87	0.16
Unmethylated cf-DNA level ( $\mu$ g/ml)*	0.75	0.20	0.67	0.15
Plasma mtDNA (copy number)*	4.30E <sup>8</sup>	2.37E <sup>8</sup>	3.75E <sup>8</sup>	2.09E <sup>8</sup>
<i>Alu</i> repeat cf-DNA (GE)*	74.4	50.4	66.8	38.3
DHEAS ( $\mu$ g/ml)*	0.25	0.48	0.25	0.31
Cortisol (ng/ml)*	133	54.3	117	68.0
IDO activity (Kyn/Trp)*	44.3	25.5	51.8	25.3
Anti-CMV antibody titer*	19.000	8.000	19.000	9000
Anti-EBV antibody titer*	405	315	410	410
DNAm age	76.1	1.04	76.1	0.64
CD3+ cells (%)* <sup>a</sup>	62.0	15.8	57.9	12.0
CD4+ cells (%)* <sup>b</sup>	62.9	2.5	63.8	1.6
CD8+ cells (%)* <sup>b</sup>	30.6	2.3	28.9	1.5
CD4+/CD8+ cells (ratio)*	2.4	2.3	2.3	2.4
CD4+CD28- cells (%)*	9.2	16.2	9.2	13.0
CD8+CD28- cells (%)	63.3	2.8	63.3	2.1
CD14+ cells (%)* <sup>a</sup>	8.3	5.9	8.1	6.3

\*median values and IQR presented

<sup>a</sup>percentage of live-gated cells; <sup>b</sup>percentage of total T lymphocytes (CD3+ cells);

<sup>c</sup>percentage of CD4+ cells; <sup>d</sup>percentage of CD8+ cells

Abbreviations: BMI, body mass index; CD, cluster of differentiation; CMV, cytomegalovirus; CRP, C-reactive protein; cf-DNA, cell-free DNA; DHEAS, dehydroepiandrosterone sulfate; DNAm, DNA methylation; EBV, Epstein-Barr virus; GE, genomic equivalent; IDO, indoleamine 2,3-dioxygenase; IL, interleukin; Kyn, kynurenine; MMSE, Mini-Mental State Examination; mtDNA, mitochondrial DNA; Trp, tryptophan



© 2000-2015 QIAGEN. All rights reserved.

**Figure 1: The highest-ranking networks from the 16 known genes harboring the top 19 significant (FDR < 0.5) CpG sites a.** and from the genes harboring the top 250 CpG sites **b.** in the 2.55-year follow-up ( $n = 111$ ). Both networks displayed NF- $\kappa$ B as a central node and were enriched for the common term *Hematological System Development and Function*. The green color of the molecule indicates that hypomethylation of a CpG site in the gene was associated with increased mortality, and the red color indicates that hypermethylation of a CpG site in the gene was associated with increased mortality. The networks were generated through the use of QIAGEN's Ingenuity Pathway Analysis (IPA<sup>®</sup>, QIAGEN Redwood City, [www.qiagen.com/ingenuity](http://www.qiagen.com/ingenuity)).

**Table 2: Canonical pathways constructed from the genes harboring the top 250 CpG sites associated with mortality at the 2.55-year follow-up.**

<b>Ingenuity Canonical Pathways</b>	<b>-log(p)*</b>	<b>Ratio</b>	<b>Molecules</b>
Chronic Myeloid Leukemia Signaling	1.91	7.61E-02	<i>TGFBR2, GAB2, HDAC4, SMAD3, PIK3R2, NFKB1, ATM</i>
Germ Cell-Sertoli Cell Junction Signaling	1.58	5.13E-02	<i>TGFBR2, MAP3K14, MAP3K10, ACTA2, KEAP1, ITGA2, PIK3R2, ATM</i>
Role of NFAT in Cardiac Hypertrophy	1.58	4.55E-02	<i>IL6ST, TGFBR2, HDAC4, ITPR3, IGF1R, SLC8A3, PIK3R2, ATM</i>
Cell Cycle: G1/S Checkpoint Regulation	1.58	7,94E-02	<i>CCND2, HDAC4, CCND3, SMAD3, ATM</i>
Regulation of the Epithelial-Mesenchymal Transition Pathway	1.58	4.40E-02	<i>MAML1, TGFBR2, FZD3, SMAD3, PIK3R2, NFKB1, SMURF1, ATM</i>
iCOS-iCOSL Signaling in T Helper Cells	1.58	5.83E-02	<i>GAB2, CD28, ITPR3, PIK3R2, NFKB1, ATM</i>
Rac Signaling	1.58	5,83E-02	<i>CYFIP2, ITGA2, PIK3R2, NFKB1, ATM, ANK1</i>
NF-κB Activation by Viruses	1.58	6.85E-02	<i>MAP3K14, ITGA2, PIK3R2, NFKB1, ATM</i>
Hepatic Fibrosis/Hepatic Stellate Cell Activation	1.58	4.08E-02	<i>TGFBR2, TNFSF4, ACTA2, MYH14, SMAD3, IGF1R, NFKB1, FAS</i>
GADD45 Signaling	1.58	1.58E-01	<i>CCND2, CCND3, ATM</i>
PKCθ Signaling in T Lymphocytes	1.57	5.31E-02	<i>MAP3K14, MAP3K10, CD28, PIK3R2, NFKB1, ATM</i>
Molecular Mechanisms of Cancer	1.57	3.06E-02	<i>TGFBR2, GAB2, CCND2, CCND3, FZD3, SMAD3, ITGA2, PIK3R2, NFKB1, FAS, ATM</i>
CNTF Signaling	1.46	8.16E-02	<i>IL6ST, CNTF, PIK3R2, ATM</i>
B Cell Receptor Signaling	1.44	4.09E-02	<i>GAB2, MAP3K14, MAP3K10, PAG1, PIK3R2, NFKB1, ATM</i>
RANK Signaling in Osteoclasts	1.44	5.81E-02	<i>MAP3K14, MAP3K10, PIK3R2, NFKB1, ATM</i>
Virus Entry via Endocytic Pathways	1.44	5.62E-02	<i>ITSN1, ACTA2, ITGA2, PIK3R2, ATM</i>
Crosstalk between Dendritic Cells and Natural Killer Cells	1.44	5.62E-02	<i>CD28, ACTA2, KLRD1, NFKB1, FAS</i>
Lymphotoxin β Receptor Signaling	1.44	7.41E-02	<i>MAP3K14, PIK3R2, NFKB1, ATM</i>
Death Receptor Signaling	1.44	5.49E-02	<i>MAP3K14, ACTA2, PARP12, NFKB1, FAS</i>
Colorectal Cancer Metastasis Signaling	1.43	3.46E-02	<i>IL6ST, TGFBR2, FZD3, SMAD3, PIK3R2, NFKB1, PTGER4, ATM</i>
Myc Mediated Apoptosis Signaling	1.40	6.90E-02	<i>IGF1R, PIK3R2, FAS, ATM</i>
T Cell Receptor Signaling	1.40	5,21E-02	<i>CD28, PAG1, PIK3R2, NFKB1, ATM</i>
Estrogen-Dependent Breast Cancer Signaling	1.30	6.45E-02	<i>IGF1R, PIK3R2, NFKB1, ATM</i>
CD40 Signaling	1.30	6.25E-02	<i>MAP3K14, PIK3R2, NFKB1, ATM</i>

HGF Signaling	1.30	4.81E-02	<i>MAP3K14, MAP3K10, ITGA2, PIK3R2, ATM</i>
Pancreatic Adenocarcinoma Signaling	1.30	4.72E-02	<i>TGFBR2, SMAD3, PIK3R2, NFKB1, ATM</i>
NGF Signaling	1.30	4.72E-02	<i>MAP3K14, MAP3K10, PIK3R2, NFKB1, ATM</i>
T Helper Cell Differentiation	1.30	5.97E-02	<i>IL6ST, TGFBR2, CD28, IL21R</i>
IL-9 Signaling	1.30	8.82E-02	<i>PIK3R2, NFKB1, ATM</i>

\*Benjamini-Hochberg-corrected *p*-value

**Table 3: The final mortality-predicting signature at the 2.55-year follow-up assessed from the Ridge regression -organized methylomic markers.**

	HR (95% CI)	S.E.	Z	p
cg08421934 (NA)	0.41 (0.26-0.64)	0.10	-3.84	<0.001
cg15770702 ( <i>MAP3K14</i> )	0.40 (0.27-0.61)	0.08	-4.38	<0.001
cg08596308 ( <i>ATP6V1G2; NFKB1L1</i> )	0.50 (0.34-0.73)	0.10	-3.60	<0.001
cg23282964 ( <i>RIOK1</i> )	0.56 (0.37-0.84)	0.12	-2.82	0.005
cg16720947 ( <i>PLEC1</i> )	0.52 (0.34-0.80)	0.13	-2.94	0.003
cg27027151 ( <i>IL21R</i> )	2.09 (1.44-3.02)	0.39	3.90	<0.001
cg26843567 (NA)	0.68 (0.46-0.99)	0.13	-2.01	0.045

Abbreviations: CI, confidence interval; HR, hazard ratio; NA, not available; S.E., standard error

Cox Regression model was tested using the global test by calculating the scaled Schoenfeld residuals for each of the independent predictors in the final Cox model. Statistically significant dependence of mortality on time was not observed ( $p = 0.280$ ) indicating that the proportionality assumption was not violated.

Due to the small number of mortality-associated CpG sites in the methylomic data at the 4-year follow-up, no comparison of the prediction accuracies of the different modeling options or assessment of the final mortality-predicting signature was performed for the 4-year mortality data.

Correlation analysis between the methylation levels in the mortality-associated CpG sites and the corresponding gene product(s) revealed a significant correlation between three CpG site/transcript pairs. Inverse correlations were observed between the cg03348466 (*CRTC3*) and *CRTC3* mRNA level and between cg04182483 (*RGS10*) and the *RGS10* mRNA level. A direct correlation was observed between cg22794214 (*HIVEP3*) and *HIVEP3* mRNA level. All the correlations are presented in Supplementary Table 5.

Analysis of the genomic locations of the top 19 CpG sites (FDR < 0.5, in Supplementary Table 2) for transcription factor (TF) binding sites and other genomic regulatory elements revealed that a majority of the sites were located on active *cis*-regulatory regions; they either harbored TF binding sites, DNase I hypersensitivity regions, and/or were identified as “Predicted promoter region including transcription start site (TSS)”, “Predicted

enhancer (E)” or “Predicted weak enhancer or open chromatin (WE)”. In addition, six CpG sites demonstrated functional significance as they were annotated for “Predicted transcribed region (T)”. The most abundant TFs were *POLR2A* and *RELA* which both had binding sites on four CpG site loci. Full data of this assessment are presented in Table 4.

## DISCUSSION

We have previously demonstrated, using genome-wide gene expression data, that the NF- $\kappa$ B complex is centrally involved in controlling human old-age mortality [10]. In the present study, we expanded the examination of the genomic factors regulating late-life survival by analyzing the predictive ability of genome-wide methylomic data at the 2.55-year follow-up. The results of this study corroborate the role of NF- $\kappa$ B in all-cause elderly mortality; the molecular network constructed from the genes harboring the mortality-associated CpG sites displayed the NF- $\kappa$ B complex as a central mediator (Figure 1). The genes *nuclear factor of kappa light polypeptide gene enhancer in B-cells 1 (NFKB1)* and *ataxia telangiectasia mutated (ATM)* were also identified in the network. Intriguingly, both *NFKB1* and *ATM* have previously been linked with accelerated aging and cellular senescence in studies with genetically engineered mice [15, 18, 19]. These studies advocated that *NFKB1* and *ATM*-regulated aberrant NF- $\kappa$ B activation and the ensuing chronic systemic inflammatory state are the ultimate



**Table 4: Assessment of the 19 mortality associated CpG sites (FDR<0.5) in the 2.55-year follow-up (*n* = 111) for transcription factor binding sites and other functional genomic elements using ENCODE data in the UCSC genome browser.**

CpG site (gene)	GRCh37/hg19 coordinate	Transcription Factors	Genome status	DNase Hypersensitivity Cluster
cg24859528 ( <i>IQSECI</i> )	chr3:12941421		T	NO
cg03348466 ( <i>CRTC3</i> )	chr15:91104770	CEBPB	T	YES
cg02395768 ( <i>ATP5SL</i> )	chr19:41945578	SIN3AK20, POLR2A, SP2, SP1, CHD2, NFYB, PBX3, MAZ, NFIC, GTF2F1, MTA3, TAF1, TBL1XR1, JUND, KDM5B, STAT5A, HDAC1, SAP30, FOS, YY1, PHF8, FOXM1, TBP, CEBPB, REST, TCF12, IRF1, TEAD4, ZBTB7A, GABPA, MEF2A, PML, RELA	TSS	YES
cg15770702 ( <i>MAP3K14</i> )	chr17:43384845	PML, STAT5A, NFATC1, CEBPB, BCL3, TCF12, EBF1, FOXM1, EP300, RELA, STAT3, NFIC, TBL1XR1, JUND, MEF2A, PAX5, BHLHE40, MEF2C, ATF2, SP1, BATF, RUNX3, IRF4, BCL11A	TSS/T	YES
cg16720947 ( <i>PLEC1</i> )	chr8:145048137		n.a.	YES
cg22794214 ( <i>HIVEP3</i> )	chr1:42123463	CTCF	WE/R	YES
cg08596308 ( <i>ATP6V1G2</i> ; <i>NFKBIL1</i> )	chr6:31516045	CHD1, RBBP5, ZNF274, POLR2A, E2F6, E2F4, KDM5B, MYC, MAX, MAZ	TSS	YES
cg23282964 ( <i>RIOK1</i> )	chr6:7417780		T	NO
cg21200667 (NA)	chr2:30628085		R	YES
cg08421934 (NA)	chr6:33942413		R	NO
cg08486432 ( <i>ITPR3</i> )	chr6:33598003		T/R	YES
cg08352439 ( <i>VOPI1</i> )	chr7:55637123	POLR2A, POU2F2	TSS	YES
cg25356639 ( <i>FOXP1</i> )	chr3:71349304		R	NO
cg04395703 ( <i>METAPI</i> )	chr4:99982762		T	YES
cg03171419 ( <i>GPR124</i> )	chr8:37700802	POLR2A	T	YES
cg26843567 (NA)	chr12:104846281		R	YES
cg00291478 ( <i>RGS10</i> )	chr10:121301041	RELA, RUNX3, RBBP5	TSS	YES
cg27027151 ( <i>IL21R</i> )	chr16:27461638	POLR2A, MTA3, NFATC1, RELA, BCLAF1, EBF1	E/R	YES
cg04182483 ( <i>RGS10</i> )	chr10:121259610		T	NO

Abbreviations: TSS, Predicted promoter region including transcription start site (TSS); T, Predicted transcribed region; WE, Predicted weak enhancer or open chromatin cis-regulatory element; E, Predicted enhancer; R, Predicted Repressed or Low Activity region

drivers of senescence and aging-associated deterioration [15, 18]. Although our data do not provide a mechanistic link between the hypomethylation of these CpG sites and the risk of mortality, we speculate that the mechanism involves an inflammatory component by which the genomic factors control late-life mortality.

Analysis of the 19 mortality-associated CpG sites (FDR < 0.5) for genomic regulatory elements revealed that a majority of the sites were located on active *cis*-regulatory regions (Table 4). That is, they harbored TF binding sites, located on DNase I hypersensitivity areas and/or displayed one of the following predicted genomic states:

promoter region including transcription start site, enhancer or weak enhancer/open chromatin. It is possible that the association between these sites and longevity is mediated through altered binding of TFs or methyl-binding domain proteins, of which the latter recruit chromatin-modifying proteins to achieve a repressive chromatin state. However, our data do not allow us to determine whether disrupted regulation of chromatin permissiveness underlies the increased mortality risk. Interestingly, RELA, which is a subunit of the NF- $\kappa$ B complex, was identified to have a binding site on four of the analyzed 19 CpG sites. This observation further supports the hypothesis of the functional role of NF- $\kappa$ B in old-age mortality.

Region of a predicted transcription start site was observed for cg02395768 (*ATP5SL*), cg15770702 (*MAP3K14*), cg08596308 (*ATP6V1G2*; *NFKBIL1*), cg08352439 (*VOPPI*) and cg00291478 (*RGS10*). However, the methylation levels in these sites were not correlated with gene expression (Supplementary Table 5). Instead, methylation levels of cg03348466 (*CRTC3*), cg22794214 (*HIVEP3*) and cg04182483 (*RGS10*) correlated with the corresponding transcript expression level. The observation that the correlations were overall modest is, however, in line with previous findings on minimal correlations between age-associated changes methylation and transcription [5, 20, 21]. Six sites, including cg03348466 (*CRTC3*) and cg04182483 (*RGS10*) resided in predicted transcribed area, and can hence also be considered functionally significant. The potential regulatory role of these sites (in the gene body region) may involve e.g. alternative splicing. However, the exact mechanism connecting the mortality-associated changes in methylation to alternative splicing requires further research.

The canonical pathways constructed from the genes harboring the top 250 mortality-associated CpG sites at the 2.55-year follow-up covered a wide variety of cellular signaling functions among which several inflammation and immunity-related processes were represented. Interestingly, pathways termed NF- $\kappa$ B Activation by Viruses, GADD45 Signaling and Cell Cycle: G1/S Checkpoint Regulation were also identified. The emergence of these pathways suggests that NF- $\kappa$ B might also be involved late-life control of cellular growth and survival, DNA repair and apoptosis, as these functions are ascribed to the induction of the NF- $\kappa$ B- GADD45 cascade [22]. Interestingly, in our previous paper on the transcriptomic mortality predictors, we observed that an increased expression of *GADD45B* was predictive of an increased risk of mortality in these nonagenarians [10].

However, as the number of mortality-associated CpG sites was markedly reduced from the 2.55-years follow-up to the 4-years follow-up, we speculate that the methylomic markers might exhibit a dynamic nature even in the extreme ages. That is, a substantial part of the genomic CpG sites might be constantly remodeled,

and during 4 years, their methylation levels are likely to change to an extent that their predictive ability in our population is reduced. The longer follow-up time also allows more time for stochastic mortality determinants, such as trauma, to operate, which may thus weaken the role of the genomic predictors.

Although the methylomic markers did not exhibit very strong statistical significances after FDR-correction and we used a liberal threshold for including them in the Ridge regression (FDR < 0.5), the methylomic data demonstrated good performance in terms of generalizability and discriminative power. Specifically, the methylomic data alone exhibited better predictive accuracy than the conventional markers alone or in combination with the methylomic markers, and the seven CpG sites in the final Cox model had a discriminative power of 89.9%. In this respect, the methylomic data also performed better than the transcriptomic mortality predictors in our previous study [10]. Nevertheless, we acknowledge that the major weaknesses of our study are a lack of a separate verification cohort and a rather small study population. Hence, the results must be considered as tentative and hypothesis-generating. The strength of our study, however, is the fact that all the study participants were 90 years of age at baseline. Therefore our results are not confounded by the effect of chronological age on DNAm.

A recent study by Moore et al. analyzed genome-wide methylomic mortality predictors in individuals with a wide age range (30-100 years at 9-year follow-up, mean mortality follow-up time 4.4 years) [23]. They identified 76 CpG sites where the rate of change in DNAm was associated with mortality and 88 markers where the year 9 level of DNAm was associated with mortality. Interestingly, their mortality-associated DNAm sites also included genes with immunoinflammatory functions and a link to NF- $\kappa$ B regulation. However, no overlap between individual mortality-associated CpG sites were found in our data sets. These differences may arise due to different population characteristics, such as age range and the causes of death. Hence, further studies are required to establish the potentially age and population-specific relationships between DNAm and mortality.

When we examined the seven final signature mortality-predicting CpG sites and their corresponding genes (Table 3) for overlap with the genes harboring the most commonly aging-associated CpG sites - *ELOVL2*, *FHL2* [2, 21, 24-26], *KLF14* [2, 21, 25, 26], *SST* [2, 25, 26], *OTUD7A* [2, 24, 26], *PENK* [2, 21, 24, 25] and *EDARADD* [2, 6, 24, 26] - we found no overlap between these markers. Moreover, none of our top 250 mortality-associated methylomic sites (Supplementary Table 2) were among the 525 common age-associated CpG sites that have been observed in more than one study (summarized in [21]). Moore et al. [23] have also observed a similar phenomenon in their population: very few (< 0,05%)

of the aging-associated CpG sites were also mortality-associated. These observations suggest that aging-associated epigenetic drift and the epigenetic control of the life span in old age might operate through different genomic mechanisms. This hypothesis is also in line with our previous findings on age-associated transcripts [27], which displayed very little similarity with mortality-predicting transcripts [10].

Despite the increasing body of data that suggests that several manifestations of organismal aging and development are of epigenetic origin, the associations reported thus far on DNAm and aging-phenotypes are scarce and/or the findings have been negative. Bell et al. (2012) examined the genome-wide associations between DNAm and 16 age-related phenotypes and found that two phenotypes - lung function and low-density lipoprotein levels - exhibited an association with one CpG site (cg16463460 in *WT1* and cg03001305 in *STAT5A*, respectively) and maternal longevity exhibited an association with two CpG sites (cg13870866 in *TBX20* and cg09259772 in *ARL44*) [8]. In another EWAS, Marioni et al. (2015) detected no individual CpG sites associated with physical or cognitive fitness in an elderly population [28]. However, they did find a cross-sectional association between a measure of DNAm age - the epigenetic clock based on the Horvath predictor [17] -, and physical and cognitive fitness yet the DNAm age was not predictive of a longitudinal change in the fitness measures [28]. The DNAm age has also been recently demonstrated to predict all-cause mortality in four different cohorts of elderly individuals [29] and in Danish twins [30]. However, the DNAm age was not predictive of mortality in our study. One reason for the negative finding might be that individuals in our cohort were all very old at baseline (90 years), and death at this age likely has different underpinnings than at younger old ages and when assessed in cohorts with wider age spectra.

In conclusion, the results of this study support the genomic-level role of NF- $\kappa$ B at the very end of the human life span. We hypothesize that our findings could relate to the recent observation of a programmatic role of hypothalamic NF- $\kappa$ B and I $\kappa$ B kinase- $\beta$  activation in the control of the life span in experimental mouse models [16]. Adhering to the conclusion of this mouse study that the decisive role of hypothalamic NF- $\kappa$ B is exerted systemically level through immune-neuroendocrine crosstalk [16], we suggest that our findings on immune cells might represent the peripheral correspondence of hypothalamic NF- $\kappa$ B activation. However, establishing the systemic-level events that connect NF- $\kappa$ B function to all cause-mortality in aged humans will require further research.

## MATERIALS AND METHODS

### Study population

The study population consisted of nonagenarian subjects participating to the Vitality 90+ study, which is an ongoing, prospective population-based study on individuals aged 90 years and above and who reside in the city of Tampere, Finland. The Vitality 90+ study was initiated in 1995, and since then several nonagenarian cohorts have been recruited and examined for biological, clinical, demographic and social measures. Mortality rates have been analyzed longitudinally using complete follow-ups. The recruitment protocol and characterization of the subjects in the current study has been previously described [10]. The data in this study concern individuals born in 1920 and recruited in 2010 for sample collection. Genome-wide methylation data and the full covariate data including cell type proportions were available for 111 subjects ( $n = 81$  women and  $n = 30$  men). The all-cause mortality data were collected from the Population Register Center. As we wanted to assess both shorter and longer-term survival predictors for this cohort, the mortality data was collected in two different time points. The first data collection was performed on 31<sup>st</sup> of January in 2013 corresponding to a 2.55-year median follow-up and the second one was on 31<sup>st</sup> of May in 2014 corresponding to a 4-year follow-up. The mortality rate at the 2.55-year follow-up was 32.4% (36/111) and 47.7% (53/111) at the 4-year follow-up. All the participants gave their written informed consent. The study was conducted following the guidelines of the Declaration of Helsinki, and the study protocol was approved by the ethics committee of the city of Tampere.

### Sample collection and processing

Venous blood samples were collected in EDTA-containing tubes by a trained home-visiting medical student between 8 am and 12 am. Plasma was separated and stored at -70°C. Genomic DNA and total RNA were extracted from PBMCs in which the blood samples were subjected to leucocyte separation using a Ficoll-Paque density gradient (Ficoll-Paque™ Premium, cat. no. 17-5442-03, GE Healthcare Bio-Sciences AB, Uppsala, Sweden). The PBMC layer was collected, and the cells allocated for RNA extraction were suspended in 150  $\mu$ l of RNeasy lysis solution (Qiagen, Crawley, UK). Cells that were allocated to FACS analysis and DNA extraction were suspended in 1 ml of a freezing solution (5/8 FBS, 2/8 RPMI-160 medium, 1/8 DMSO; FBS cat. no. F7524, Sigma-Aldrich, MO, USA; RPMI: cat. no. R0883, Sigma-Aldrich, MO, USA; DMSO: cat. no. 1.02931.0500, VWR, Espoo, Finland).

Characterization of the subjects for their anthropometric measures, functional performance, plasma biomarkers and blood cell distributions (all parameters presented in Table 1) has been previously described (please see [10] and the references therein). In addition to these variables, in the current study we also determined a measure of the “epigenetic clock” - the DNAm age in the PBMCs - using the methodology presented in the study by Horvath et al. (2013) [17], (algorithm available at <https://dnamage.genetics.ucla.edu/>). Initially, the age predictor was generated by Horvath with elastic net regression using 21,369 probes that are present in HumanMethylation450 as well as in HumanMethylation27 BeadChips. The predictor was trained with 8,000 samples of various tissue types in 82 Illumina DNA methylation array data sets. Based on the training results, the “epigenetic clock” i.e. the regression model was built with 353 CpG-sites whose methylation level explains most of the age variation.

### Illumina methylation array and preprocessing of the data

Genome-wide DNA methylation profiling was performed using the Infinium HumanMethylation450 BeadChip (Illumina, San Diego, CA, USA) according to the manufacturer’s protocol at the Institute for Molecular Medicine Finland (FIMM) Technology Centre of the University of Helsinki. For bisulfite conversion, 1 µg of DNA was used (EZ-96 DNA Methylation Kit, Zymo Research, Irvine, CA, USA) and 4 µl of the bisulfite-converted DNA was subjected to whole-genome amplification and enzymatic fragmentation. Hybridization was carried out according to the manufacturer’s protocol. Samples were run on the arrays in a randomized order and the chips were scanned with the iScan reader (Illumina).

The *methylumiset* object in the R software with the *wateRmelon* array-specific package from Bioconductor was used in preprocessing of the data. Probes mapping to sex chromosomes ( $n = 11,648$ ) were also removed. In addition, all polymorphic sites and sites exhibiting unspecific probe binding ( $n = 76,775$ ) were filtered out based on database information [31]. CpG target sites demonstrating technically poor quality were filtered out, including sites with a beadcount of  $< 3$  in 5% of the samples ( $n = 515$ ) and sites for which 1% of the samples had a detection  $p$ -value  $> 0.05$  ( $n = 698$ ). The annotation information for the CpG sites was retrieved using the GRCh37/hg19 genome assembly (released in February 2009). The *dasen* method was used for background correction and quantile normalization individually for the two applied chemistries in the Illumina platform (Infinium I and II) and for the intensities of methylation ( $m$ ) and un-methylation ( $u$ ). Following the *dasen* method, the  $u$  and  $m$  intensities were transformed to beta ( $\beta$ ) and  $M$  values, where  $\beta$  is the ratio of the methylated probe ( $m$ )

intensity in relation to the overall intensities ( $m + u + \alpha$ ), where  $\alpha$  is the constant offset, i.e., 100. Lastly, the batch effect of the different chemistries was corrected using the BMIQ method, which is based on beta mixture models and the EM algorithm [32]. The batch effect produced by two different run series was corrected using an Empirical Bayes-based algorithm implemented in the R package *Combat*. Because the proportions of the CD4+CD28-, CD8+CD28- and CD14+ cells and the CD4+ to CD8+ cell ratio were associated with the variation in methylation data in the principal component analysis [33], the data was regressed in the *variable dispersion beta regression* model from Ferrari and Cribari-Neto [34] with the explanatory variables of gender and the proportions of blood cell types after which the standardized weighted residuals were extracted and used in all further statistical analyses. The model utilizes beta density function with parameterizations:

$$\varphi(y, \mu, \phi) = \frac{\Gamma(\phi)}{\Gamma(\phi)^2 \Gamma(1-\mu)\phi} y^{\phi-1} (1-y)^{(1-\mu)\phi-1}$$

where  $\Gamma(\cdot)$  is the gamma function;  $y$  is the continuous response variable with a mean value of  $\mu$ , which is assumed to follow a beta distribution inside the interval  $y \in (0,1)$ ; and  $\phi < 0$  is the precision parameter. The variance of  $y$  is inherited from the binomial variance  $\mu(1-\mu)$ , and it can be written as  $\mu(1-\mu)/(1+\phi)$ . Beta regression utilizes maximum likelihood for estimating the parameters in the equation, and the mean value of  $y$  is connected to the linear equation with the canonical link function *logit*. The model is implemented in the R package *betareg* as a default setting. The methylation data are available in the GEO database (<http://www.ncbi.nlm.nih.gov/geo/>) under the accession number GSE68194.

All the CpG sites passing the quality control and preprocessing criteria described above as well as the conventional variables presented in Table 1 were first analyzed for their univariate association with mortality after which all the significant methylomic markers ( $p < 0.05$ ) were corrected for FDR with the B-H -method ( $FDR < 0.5$ ). The Cox regression models were performed using Stata software (version 13.0 for Windows, StataCorp LP, TX, USA), and the corrections for FDR were performed using R version 3.0.2.

### Ridge regression

Due to the high dimensionality and multicollinearity of the genome-wide data, the standard Cox regression method cannot be directly applied to yield parameter estimates. Hence, several different dimension reduction and feature selection procedures have been presented for such data. In this study, we made use of the Ridge regression [35] that is based on penalized partial likelihood, and provides a means to avoid overfitting and unstable predictors. It has also been shown to produce reproducible results in whole-genome data sets by others

[36] and us [10].

Ridge regression is a technique to analyze data when predictors are correlated with other predictors. In the presence of this multicollinearity, the variance of the regression coefficients is increased making them unstable. By adding a little bias (tuning parameter  $\lambda$ ) to the coefficients, the Ridge regression reduces the variance considerably. In the Ridge regression, the regression coefficients are regularized by imposing penalties on their size. Thus, the coefficients are shrunk toward zero and toward each other, and the tuning parameter  $\lambda$  controls for the amount of shrinkage. There is no definitive rule for choosing  $\lambda$ , but the objective is to produce only a small increase in the weighted sum of square errors [37]. To select an optimal value of  $\lambda$ , a  $k$ -fold cross-validation is often performed. For the Cox proportional hazards model, Verweij and van Houwelingen [38] introduced a cross-validated partial log-likelihood method. In  $k$ -fold cross-validation, the data set is split in  $k$  pieces, using  $k - 1$  of those used to build the model and from thereon validating on the  $k$ th, and *via* cycling through this assessment, validating on each of the  $k$  pieces sequentially, and then averaging or summing the  $k$  different deviances [39].

We estimated the optimal value of  $\lambda$  by maximizing the 10-fold cross-validated log partial likelihood. The optimal  $\lambda$  was then used to obtain parameter estimates for the different models, i.e., the conventional markers (MMSE and BMI) alone, the methylomic markers (the 19 CpG sites with an FDR < 0.5) alone and in combination with the above-mentioned conventional markers, and the methylomic markers ordered according to their statistical significance ( $p$ -value) in the univariate selection. The R package penalized was used in this assessment.

### Assessment of the predictive accuracy of modeling (generalizability) through cross-validation

We sought the most accurate mortality prediction model by assessing the differences in the deviances through cross-validation. The tested data sets were the above-mentioned three models i.e., the conventional markers alone, Ridge regressions containing the methylomic markers alone (Ridge1) and combined with the conventional markers (Ridge2) and the methylomic markers assessed through univariate selection. The procedure was performed following the guidelines presented by Bovelstad et al. (2011) [40]. In specific, the study population was randomly split 50 times into training and test sets (74 and 37 individuals, respectively). The difference in deviance between the fitted model and the null model containing no covariates is given by

$$\hat{\delta} = -2 \{l^{(test)}(\hat{\beta}_{train}) - l^{(test)}(\mathbf{0})\},$$

where  $l^{(test)}(\hat{\beta}_{train})$  and  $l^{(test)}(\mathbf{0})$  are the Cox log partial likelihoods for the test data evaluated at  $\hat{\beta}_{train}$  and  $\mathbf{0}$ , respectively. A small value of  $\hat{\delta}$  is indicative of good

performance.

### Assessment of the final mortality-predicting signature

The final signature predictive of mortality in the population with methylation data available ( $n = 111$ ) was assessed at the 2.55-year follow-up. The variables (19 CpG sites with an FDR-corrected  $p$ -value < 0.5) were collected from the model demonstrating the best accuracy of prediction (i.e., the Ridge regression containing only the methylation markers) and assessed in a stepwise Cox multivariate regression model. AIC was used to select the Cox regression model congaing the best set of predictors.

### Pathway analyses

IPA (QIAGEN Ingenuity Pathway Analysis (IPA<sup>®</sup>, QIAGEN Redwood City, www.qiagen.com/ingenuity) was used to identify canonical pathways and networks for the mortality-associated genes harboring the CpG sites (presented in Supplementary Table 2). If a CpG site was mapped to more than one gene, each of the genes were included in the network and pathway analyses. A description and principles of the pathway analysis have been previously provided in more detail [10]. B-H correction for FDR was used to assess the significance of the pathways; canonical pathways were considered significant at  $p < 0.05$  (corresponding to a  $-\log p < 1.3$ ).

### Correlations between the methylomic markers and gene expression

The genome-wide gene expression analysis was performed using HumanHT-12 v4 Expression BeadChip (Cat no. BD-103-0204, Illumina Inc., CA, USA) at the Core Facility of the Department of Biotechnology of the University of Tartu. Preprocessing and analysis of the data were performed as previously described [10]. Briefly, the lumi pipeline was used; the background was corrected with the bgAdjust.affy package, the data were log2-transformed and quantile-normalized. Poor-quality data and background noise were filtered out as follows: probes exhibiting expression levels of < 5 or > 100 in more than 5 (3.3%) samples per transcript were excluded. The gene expression data are available in the GEO database (<http://www.ncbi.nlm.nih.gov/geo/>) under the accession number GSE65218. The correlations between the transcript expression levels and CpG site methylation level (the standardized weighted residuals) were analyzed using Spearman's rho. In the analysis we included the top 19 mortality-associated GpC sites presented in Supplementary Table 2 and the corresponding transcripts with expression level above the selected threshold of 5 i.e.,

ATP5SL, FOXP1, HIVEP3, IQSEC1, ITPR3, MAP3K14, METAP1, RGS10, RIOK1 and VOPP1.

### Analysis of the mortality-associated CpG site loci for gene regulatory elements

To obtain further functional information about the mortality-predicting CpG sites, the single-base resolution locations of these sites were examined for gene regulatory elements using the Encyclopedia of DNA Elements (ENCODE) Consortium data [41] in the UCSC genome browser (<http://genome.ucsc.edu/>, accessed 02/2016). Specifically, we searched for TF binding sites (ChIP-seq data), genome states determined through combined genome segmentation data (ChromHMM and Segway programs) and DNase I hypersensitivity clusters indicative of genomic regulatory regions. Default settings were used in inspecting the elements. However, we considered data only from cell types of blood origin; that is, the DNase I hypersensitivity clusters and TFs were included in the results only if cells of blood origin were included in the cluster score, and for the analysis genomic states, data from GM12878 and K562 cells were accepted.

### ACKNOWLEDGMENTS

The authors wish to thank Sinikka Repo-Koskinen for her skillful technical assistance.

### CONFLICTS OF INTEREST

The authors declare that they have no competing interests.

### GRANT SUPPORT

This work was financially supported by The Academy of Finland (M.H.) and The Competitive Research Funding of the Tampere University Hospital (M.H. grant 9M0179). The funding bodies had no role in the design of the study, collection of the material, analysis and interpretation of data, writing of the manuscript or decision to submit the manuscript for publication.

### REFERENCES

1. Horvath S, Zhang Y, Langfelder P, Kahn RS, Boks MPM, van Eijk K, van den Berg LH and Ophoff RA. Aging effects on DNA methylation modules in human brain and blood tissue. *Genome biology*. 2012; 13:R97.
2. Hannum G, Guinney J, Zhao L, Zhang L, Hughes G, Sada S, Klotzle B, Bibikova M, Fan JB, Gao Y, Deconde R, Chen M, Rajapakse I, Friend S, Ideker T and Zhang K. Genome-wide methylation profiles reveal quantitative views of

- human aging rates. *Molecular cell*. 2013; 49:359-367.
3. Rakyan VK, Down TA, Maslau S, Andrew T, Yang T-P, Beyan H, Whittaker P, McCann OT, Finer S, Valdes AM, Leslie RD, Deloukas P and Spector TD. Human aging-associated DNA hypermethylation occurs preferentially at bivalent chromatin domains. *Genome research*. 2010; 20:434-439.
4. Jaffe AE and Irizarry RA. Accounting for cellular heterogeneity is critical in epigenome-wide association studies. *Genome biology*. 2014; 15:R31.
5. Marttila S, Kananen L, Hayrynen S, Jylhava J, Nevalainen T, Hervonen A, Jylha M, Nykter M and Hurme M. Ageing-associated changes in the human DNA methylome: genomic locations and effects on gene expression. *BMC Genomics*. 2015; 16:179.
6. Teschendorff AE, Menon U, Gentry-Maharaj A, Ramus SJ, Weisenberger DJ, Shen H, Campan M, Noushmehr H, Bell CG, Maxwell AP, Savage DA, Mueller-Holzner E, Marth C, Kocjan G, Gayther SA, Jones A, et al. Age-dependent DNA methylation of genes that are suppressed in stem cells is a hallmark of cancer. *Genome research*. 2010; 20:440-446.
7. Fraga MF, Ballestar E, Paz MF, Ropero S, Setien F, Ballestar ML, Heine-Suner D, Cigudosa JC, Urioste M, Benitez J, Boix-Chornet M, Sanchez-Aguilera A, Ling C, Carlsson E, Poulsen P, Vaag A, et al. Epigenetic differences arise during the lifetime of monozygotic twins. *Proc Natl Acad Sci U S A*. 2005; 102:10604-10609.
8. Bell JT, Tsai PC, Yang TP, Pidsley R, Nisbet J, Glass D, Mangino M, Zhai G, Zhang F, Valdes A, Shin SY, Dempster EL, Murray RM, Grundberg E, Hedman AK, Nica A, et al. Epigenome-wide scans identify differentially methylated regions for age and age-related phenotypes in a healthy ageing population. *PLoS Genet*. 2012; 8:e1002629.
9. Nevalainen T, Kananen L, Marttila S, Jylha M, Hervonen A, Hurme M and Jylhava J. Transcriptomic and epigenetic analyses reveal a gender difference in aging-associated inflammation: the Vitality 90+ study. *Age (Dordrecht, Netherlands)*. 2015; 37:9814.
10. Jylhava J, Raitanen J, Marttila S, Hervonen A, Jylha M and Hurme M. Identification of a prognostic signature for old-age mortality by integrating genome-wide transcriptomic data with the conventional predictors: the Vitality 90+ Study. *BMC medical genomics*. 2014; 7:54.
11. Salminen A and Kaarniranta K. Insulin/IGF-1 paradox of aging: regulation *via* AKT/IKK/NF-kappaB signaling. *Cellular signalling*. 2010; 22:573-577.
12. Tak PP and Firestein GS. NF-kappaB: a key role in inflammatory diseases. *The Journal of clinical investigation*. 2001; 107:7-11.
13. Le Saux S, Weyand CM and Goronzy JJ. Mechanisms of immunosenescence: lessons from models of accelerated immune aging. *Annals of the New York Academy of Sciences*. 2012; 1247:69-82.

14. Salminen A, Huuskonen J, Ojala J, Kauppinen A, Kaarniranta K and Suuronen T. Activation of innate immunity system during aging: NF- $\kappa$ B signaling is the molecular culprit of inflamm-aging. *Ageing research reviews*. 2008; 7:83-105.
15. Osorio FG, Barcena C, Soria-Valles C, Ramsay AJ, de Carlos F, Cobo J, Fueyo A, Freije JM and Lopez-Otin C. Nuclear lamina defects cause ATM-dependent NF- $\kappa$ B activation and link accelerated aging to a systemic inflammatory response. *Genes & development*. 2012; 26:2311-2324.
16. Zhang G, Li J, Purkayastha S, Tang Y, Zhang H, Yin Y, Li B, Liu G and Cai D. Hypothalamic programming of systemic ageing involving IKK- $\beta$ , NF- $\kappa$ B and GnRH. *Nature*. 2013; 497:211-216.
17. Horvath S. DNA methylation age of human tissues and cell types. *Genome biology*. 2013; 14:R115.
18. Jurk D, Wilson C, Passos JF, Oakley F, Correia-Melo C, Greaves L, Saretzki G, Fox C, Lawless C, Anderson R, Hewitt G, Pender SL, Fullard N, Nelson G, Mann J, van de Sluis B, et al. Chronic inflammation induces telomere dysfunction and accelerates ageing in mice. *Nature communications*. 2014; 2:4172.
19. Tilstra JS, Robinson AR, Wang J, Gregg SQ, Clauson CL, Reay DP, Nasto LA, St Croix CM, Usas A, Vo N, Huard J, Clemens PR, Stolz DB, Guttridge DC, Watkins SC, Garinis GA, et al. NF- $\kappa$ B inhibition delays DNA damage-induced senescence and aging in mice. *The Journal of clinical investigation*. 2012; 122:2601-2612.
20. Yuan T, Jiao Y, de Jong S, Ophoff RA, Beck S and Teschendorff AE. An integrative multi-scale analysis of the dynamic DNA methylation landscape in aging. *PLoS Genet*. 2015; 11:e1004996.
21. Steegenga WT, Boekschoten MV, Lute C, Hooiveld GJ, de Groot PJ, Morris TJ, Teschendorff AE, Butcher LM, Beck S and Muller M. Genome-wide age-related changes in DNA methylation and gene expression in human PBMCs. *Age*. 2014; 36:9648.
22. Yang Z, Song L and Huang C. Gadd45 proteins as critical signal transducers linking NF- $\kappa$ B to MAPK cascades. *Current cancer drug targets*. 2009; 9:915-930.
23. Moore AZ, Hernandez DG, Tanaka T, Pilling LC, Nalls MA, Bandinelli S, Singleton AB and Ferrucci L. Change in Epigenome-Wide DNA Methylation Over 9 Years and Subsequent Mortality: Results From the InCHIANTI Study. *The journals of gerontology Series A, Biological sciences and medical sciences*. 2015 Sep 9. pii: glv118. PMID: 26355017.
24. Florath I, Butterbach K, Muller H, Bewerunge-Hudler M and Brenner H. Cross-sectional and longitudinal changes in DNA methylation with age: an epigenome-wide analysis revealing over 60 novel age-associated CpG sites. *Hum Mol Genet*. 2014; 23:1186-1201.
25. Garagnani P, Bacalini MG, Pirazzini C, Gori D, Giuliani C, Mari D, Di Blasio AM, Gentilini D, Vitale G, Collino S, Rezzi S, Castellani G, Capri M, Salvioli S and Franceschi C. Methylation of ELOVL2 gene as a new epigenetic marker of age. *Aging Cell*. 2012; 11:1132-1134.
26. Heyn H, Li N, Ferreira HJ, Moran S, Pisano DG, Gomez A, Diez J, Sanchez-Mut JV, Setien F, Carmona FJ, Puca AA, Sayols S, Pujana MA, Serra-Musach J, Iglesias-Platas I, Formiga F, et al. Distinct DNA methylomes of newborns and centenarians. *Proc Natl Acad Sci U S A*. 2012; 109:10522-10527.
27. Marttila S, Jylhava J, Nevalainen T, Nykter M, Jylha M, Hervonen A, Tserel L, Peterson P and Hurme M. Transcriptional analysis reveals gender-specific changes in the aging of the human immune system. *PloS one*. 2013; 8:e66229.
28. Marioni RE, Shah S, McRae AF, Ritchie SJ, Muniz-Terrera G, Harris SE, Gibson J, Redmond P, Cox SR, Pattie A, Corley J, Taylor A, Murphy L, Starr JM, Horvath S, Visscher PM, et al. The epigenetic clock is correlated with physical and cognitive fitness in the Lothian Birth Cohort 1936. *International journal of epidemiology*. 2015; 44:1388-96.
29. Marioni RE, Shah S, McRae AF, Chen BH, Colicino E, Harris SE, Gibson J, Henders AK, Redmond P, Cox SR, Pattie A, Corley J, Murphy L, Martin NG, Montgomery GW, Feinberg AP, et al. DNA methylation age of blood predicts all-cause mortality in later life. *Genome biology*. 2015; 16:25.
30. Christiansen L, Lenart A, Tan Q, Vaupel JW, Aviv A, McGue M and Christensen K. DNA methylation age is associated with mortality in longitudinal Danish twin study. *Aging Cell*. 2016; 15:149-154.
31. Chen Y-a, Lemire M, Choufani S, Butcher DT, Grafodatskaya D, Zanke BW, Gallinger S, Hudson TJ and Weksberg R. Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. *Epigenetics*. 2013; 8:203-209.
32. Teschendorff AE, Marabita F, Lechner M, Bartlett T, Tegner J, Gomez-Cabrero D and Beck S. A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. *Bioinformatics*. 2013; 29:189-196.
33. Marttila S, Kananen L, Hayrynen S, Jylhava J, Nevalainen T, Hervonen A, Jylha M, Nykter M and Hurme M. Ageing-associated changes in the human DNA methylome: genomic locations and effects on gene expression. *BMC Genomics*. 2015; 16:1381.
34. Ferrari SLP and Cribari-Neto F. Beta regression for modelling rates and proportions. *J Appl Stat*. 2004; 31:799-815.
35. Hoerl AE and Kennard RW. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*. 2000; 42:80-86.

36. Bovelstad HM, Nygard S and Borgan O. Survival prediction from clinico-genomic models - a comparative study. *Bmc Bioinformatics*. 2009; 10.
37. Xue XN, Kim MMY and Shore RE. Cox regression analysis in presence of collinearity: an application to assessment of health risks associated with occupational radiation exposure. *Lifetime Data Anal*. 2007; 13:333-350.
38. Verweij PJ and Van Houwelingen HC. Penalized likelihood in Cox regression. *Stat Med*. 1994; 13:2427-2436.
39. Simon N, Friedman J, Hastie T and Tibshirani R. Regularization Paths for Cox's Proportional Hazards Model *via* Coordinate Descent. *J Stat Softw*. 2011; 39:1-13.
40. Bovelstad HM and Borgan O. Assessment of evaluation criteria for survival prediction from genomic data. *Biom J*. 2011; 53:202-216.
41. Rosenbloom KR, Sloan CA, Malladi VS, Dreszer TR, Learned K, Kirkup VM, Wong MC, Maddren M, Fang RH, Heitner SG, Lee BT, Barber GP, Harte RA, Diekhans M, Long JC, Wilder SP, et al. ENCODE Data in the UCSC Genome Browser: year 5 update. *Nucleic Acids Res*. 2013; 41:D56-D63.