(Genetic) Epidemiology of Aging

Linda Broer

The work presented in this thesis was conducted at the Genetic Epidemiology Unit, Department of Epidemiology, Erasmus University Medical Center, Rotterdam.

The ERF study was supported by grants from the European Commission FP6 STRP grant number 018947 (LSHG-CT-2006-01947), the European Community's Seventh Framework Programme (FP7/2007-2013)/grant agreement HEALTH-F4-2007-201413 by the European Commission under the programme "Quality of Life and Management of the Living Resources" of 5th Framework Programme (no. QLG2-CT-2002-01254). High-throughput analysis of the ERF data was supported by joint grant from Netherlands Organisation for Scientific Research and the Russian Foundation for Basic Research (NWO-RFBR 047.017.043).

The Rotterdam Study is funded by the Netherlands Organisation of Scientific Research NWO Investments (nr. 175.010.2005.011, 911-03-012, the Research Institute for Diseases in the Elderly (014-93-015; RIDE2), the Netherlands Genomics Initiative (NGI)/Netherlands Organisation for Scientific Research (NWO) project nr. 050-060-810, the Erasmus Medical Center and Erasmus University, Rotterdam, Netherlands Organization for the Health Research and Development (ZonMw), the Research Institute for Diseases in the Elderly (RIDE), the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the European Commission (DG XII), and the Municipality of Rotterdam. Exome sequencing in longevity research was funded by BBMRI-NL.

The work presented in this thesis is conducted as part of the Netherlands Consortium of Healthy Aging (NCHA). The work was further supported by the by the CHANCES project funded in the FP7 framework programme of DG-RESEARCH in the European Commission (grant no. 242244) and ENGAGE Consortium, grant agreement HEALTH-F4-2007- 201413.

Cover: Linda Broer
Lay-out: Legatron Electronic Publishing
Printing: IPSKAMP Drukkers

ISBN: 978-94-6191-837-6

(Genetic) Epidemiology of Aging

(Genetische) epidemiologie naar veroudering

Proefschrift

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de
rector magnificus

Prof.dr. H.G. Schmidt

en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op

woensdag 9 oktober 2013 om 11:30 uur

door

Linda Broer
geboren te Rotterdam

PROMOTOREN
Prof.dr.Ir. C.M. van Duijn
Prof.dr. B.A. Oostra

OVERIGE LEDEN
Prof.dr. P.E. Slagboom
Prof.dr. H.W. Tiemeier
Prof.dr. A.G. Uitterlinden

COPROMOTOR
Dr. N. Amin

PUBLICATIONS AND MANUSCRIPTS IN THIS THESIS

Chapter 2.1

Broer, L., Buchman, A. S., Deelen, J., Evans, D.S., Faul, J. D., Lunetta, K. L., Sebastiani, P., Smith, J. A., Smith, A. V., Tanaka, T., Yu, L., Arnold, A. M., Aspelund, T., Benjamin, E. J., De Jager, P. L., Eirkisdottir, G., Evans, D. A., Garcia, M. E., Hofman, A., Kaplan, R. C., Kardia, S. L. R., Kiel, D. P., Oostra, B. A., Orwoll, E. S., Parimi, N., Psaty, B. M., Rivadeneira, F., Rotter, J., Seshadri, S., Singleton, A., Tiemeier, H., Uitterlinden, A. G., Zhao, W., Bandinelli, S., Bennett, D. A., Ferrucci, L., Gudnason, V., Harris, T. B., Karasik, D., Launer, L. J., Newman, A. B., Perls, T. T., Slagboom, P. E., Tranah, G. J., Weir, D. R., van Duijn, C. M., Murabito, J.M. *GWAS of longevity in the CHARGE consortium confirms APOE and FOXO3* (In preparation)

Chapter 2.2

Broer, L., Amin, N., Uitterlinden, A. G., Oostra, B. A., Hofman, A., Tiemeier, H., van Duijn, C.M. *The genetic architecture of longevity in a population-based study, the Rotterdam Study* (In preparation)

Chapter 3.1

Broer, L., Ikram, M. A., Schuur, M., Destefano, A. L., Bis, J. C., Liu, F., Rivadeneira, F., Uitterlinden, A. G., Beiser, A. S., Longstreth, W. T., Hofman, A., Aulchenko, Y., Seshadri, S., Fitzpatrick, A. L., Oostra, B. A., Breteler, M. M. & Van Duijn, C. M. (2011) Association of hsp70 and its co-chaperones with alzheimer's disease. *J Alzheimers Dis,* 25(1)**,** 93-102.

Chapter 3.2

Broer, L., Koudstaal, P. J., Amin, N., Rivadeneira, F., Uitterlinden, A. G., Hofman, A., Oostra, B. A., Breteler, M. M., Ikram, M. A. & Van Duijn, C. M. (2011) Association of heat shock proteins with parkinson's disease. *Eur J Epidemiol,* 26(12)**,** 933-5.

Chapter 3.3

Broer, L., Demerath, E. W., Garcia, M. E., Homuth, G., Kaplan, R. C., Lunetta, K. L., Tanaka, T., Tranah, G. J., Walter, S., Arnold, A. M., Atzmon, G., Harris, T. B., Hoffmann, W., Karasik, D., Kiel, D. P., Kocher, T., Launer, L. J., Lohman, K. K., Rotter, J. I., Tiemeier, H., Uitterlinden, A. G., Wallaschofski, H., Bandinelli, S., Dorr, M., Ferrucci, L., Franceschini, N., Gudnason, V., Hofman, A., Liu, Y., Murabito, J. M., Newman, A. B., Oostra, B. A., Psaty, B. M., Smith, A. V. & Van Duijn, C. M. (2013) Association of heat shock proteins with all-cause mortality. Age (Dordr), 35(4), 1367-76.

Chapter 4.1

Broer, L., Codd, V., Nyholt, D. R., Deelen, J., Mangino, M., Willemsen, G., Albrecht, E., Amin, N., Beekman, M., De Geus, E. J., Henders, A., Nelson, C. P., Steves, C. J., Wright, M. J., De Craen, A. J., Isaacs, A., Matthews, M., Moayyeri, A., Montgomery, G. W., Oostra, B. A., Vink, J. M., Spector, T. D., Slagboom, P. E., Martin, N. G., Samani, N. J., Van Duijn, C. M. & Boomsma, D. I. (2013) Meta-analysis of telomere length in 19 713 subjects reveals high heritability, stronger maternal inheritance and a paternal age effect. *Eur J Hum Genet*.

Chapter 4.2

Broer, L., Amin, N., Codd, V., Oostra, B. A., Samani, N. J., van Duijn, C.M. *Telomere length variation reduces with age: evidence of survivor effect* (In preparation)

Chapter 4.3

Broer, L., Mueezzinler, A., Hageman, G. J., Amin, N., Albrecht, E., Codd, V., Giltay, E. J., Martino, A., Boffetta, P., Trichopoulou, A., Oostra, B. A., Zaineddin, A. K., Samani, N. J., Kromhout, D., Brenner, H., van Duijn, C. M. *Association of height and telomere length in the CHANCES consortium* (In preparation)

Chapter 4.4

Broer, L., Raschenberger, J., Deelen, J., Mangino, M., Codd, V., Pietiläinen, K. H., Albrecht, E., Amin, N., Beekman, M., de Craen, A. J. M., Gieger, C., Haun, M., Henneman, P., Herder, C., Hovatta, I., Kedenko, L., Koenig, W., Kollerits, B., Moilanen, E., Oostra, B. A., Paulweber, B., Quaye, L., Rissanen, A., Roden, M., Surakka, I., Valdes, A. M., Vuolteenaho, K., Thorand, B., van Dijk, K. W., Kaprio, J., Spector, T. D., Slagboom, P. E., Samani, N. J., Kronenberg, F., van Duijn, C. M., Ladwig, K-H. *Association of adiponectin and leptin with relative telomere length in 7 independent cohorts including 11,460 participants* (Submitted)

Chapter 4.5

Broer, L., Draisma, H. H. M., Pool, R., Albrecht, E., Beekman, M., Mangino, M., Raag, M., Nyholt, D. R., Dharuri, H. K., Codd, V., Amin, N., de Geus, E. J. C., Deelen, J., Demirkan, A., Erte, I., Fischer, K., Haller, T., Henders, A. K., Isaacs, A., Medland, S. E., Montgomery, G. W., Mooijaart, S. P., Oostra, B. A., Strauch, K., Suchiman, H. E. D., Vaarhorst, A. A. M., van Heemst, D., Wang-Sattler, R., Whitfield, J. B., Willemsen, G., Wright, M. J., Martin, N. G., Samani, N. J., Metspalu, A., Slagboom, P. E., Spector, T. D., Boomsma, D. I., van Duijn, C. M., Gieger, C. *Associations between telomere length and metabolite profiles provide new insight into the biology of aging* (In preparation)

Chapter 5.1

Broer, L., Lill, C. M., Schuur, M., Amin, N., Roehr, J. T., Bertram, L., Ioannidis, J. P. & Van Duijn, C. M. (2013) Distinguishing true from false positives in genomic studies: P values. *Eur J Epidemiol*, 28(2), 131-8.

Table of contents

Chapter 1 Introduction

General Introduction 9

Chapter 2 Genome Wide Association Studies of longevity

2.1 GWAS of longevity in the CHARGE consortium confirms *APOE* and *FOXO3* 27

2.2 The genetic architecture of longevity in a population-based study, the Rotterdam Study 39

Chapter 3 The role of Heat Shock Proteins in morbidiy and mortality

3.1 Association of HSP70 and its co-chaperones with Alzheimer's Disease 57

3.2 Association of Heat Shock Proteins with Parkinson's Disease 71

3.3 Association of Heat Shock Proteins with all-cause mortality 77

Chapter 4 The role of telomere length in morbidity and mortality

4.1 Meta-analysis of telomere length in 19,713 subjects reveals high heritability, stronger maternal inheritance and a paternal age effect 91

4.2 Telomere length variation reduces with age: evidence of survivor effect 105

4.3 Association of height and telomere length in the CHANCES consortium 117

4.4 Association of adiponectin and leptin with telomere length in 7 independent cohorts including 11,460 participants 131

4.5 Associations between telomere length and metabolite profiles provide new insight into the biology of aging 147

Chapter 5 General Discussion

5.1 Distinguishing true from false positives in genomic studies: p-values 161

5.2 Findings of this thesis and future research 175

Chapter 6 Summary and Samenvatting 189

Chapter 7 Appendices (supplementary material) 197

Chapter 8 Acknowledgements / Publications 261

Chapter 1

1

General Introduction

LONGEVITY

Longevity is usually defined as age at death or survival to an exceptional age, such as 90 years or older or even 100 years or older [1]. In the past century, most Western countries have experienced substantial increases in life expectancy. This has been mostly due to a marked reduction in early life mortality during the first half of the twentieth century, followed by an almost twofold reduction in mortality at ages above 70 years in the past 50 years [2] (**Figure 1**; source: CBS). Longevity is a complex phenotype to which both environmental factors such as lifestyle and genetic factors are known to contribute [3]. The genetic contribution to age at death has been estimated to range from 15 to 25%, and up to 40% for reaching longevity, suggesting a significant but relatively modest genetic contribution to the human lifespan [4-6]. However, the clustering of extreme ages in families and the increased relative risk of survival to old age amongst the siblings of nonagenarians [7] and of centenarians (100+) [8-13] suggest that the genetic factors play a much stronger role as the phenotype becomes more extreme [3]. While environmental factors affecting longevity include smoking, alcohol consumption, diastolic blood pressure and C-reactive protein [14], *APOE* [3, 15-19] and *FOXO3* [20, 21] remain the only genes consistently associated with longevity.

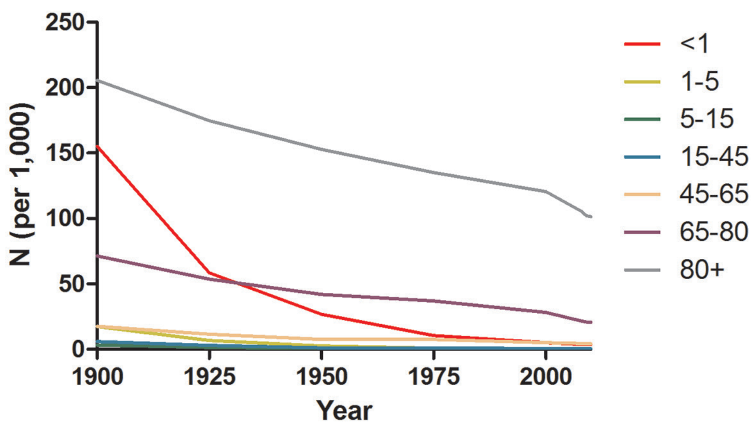


Figure 1 | Age at death in the previous century (Source: CBS 2013).

METHODS FOR GENE DISCOVERY IN COMPLEX TRAITS

Two basic gene finding methods are linkage analysis and association analysis. Linkage analysis searches for regions of the genome with a higher-than-expected number of shared alleles among affected individuals within a family [22]. These chromosomal regions are expected to harbor mutation(s) that cause the disease. As this method involves segregation, it can only be applied in a family based setting. The method is powerful to discover loci

harboring mutations that might be rare and family specific and have large effects on the disease. To identify variants that are common in populations and have small effects on the disease (as are expected in the case of a complex trait like longevity), association analysis is best performed. Association analysis tests whether a specific genetic variant is more frequent in cases than in controls in case of a disease. There are two ways in which association analysis can be performed; (1) direct association analysis and (2) indirect association analysis. In direct association analysis functional/causal variants are tested for association with the disease while in indirect association analysis common variant(s), that are expected to tag the unknown functional/causal variant, are tested for association with the disease. The direct association analysis can be performed in a candidate gene approach where causal variants in genes that are known to have a biological impact on the disease are directly genotyped; or on Next generation sequence (NGS) data, which captures most variation in the genome [23] including the rare and causal variants. Indirect association is generally performed as genome-wide association studies (GWAS) where studies use various single nucleotide polymorphisms (SNPs) genotyping arrays (e.g., http://www.illumina.com/applications/genotyping.ilmn) to perform genotyping of common SNPs ranging from a few hundred thousand to millions across the genome, assuming that these will tag the actual causal variant by means of proximity and correlation referred to as linkage disequilibrium (LD). Since GWAS involve testing of all SNPs genome-wide regardless of any prior hypothesis, unlike the candidate gene approach, GWAS are hypothesis free. As common variants are likely to have small effects on the disease, large cohorts are needed to achieve sufficient statistical power for their discovery. Therefore, in genome wide association studies (GWAS), the aim is to accumulate the highest achievable number of genotypes of as many subjects as possible. This is achieved by pooling of GWAS in a meta-analysis. Pooling requires studies to perform genetic imputations to a common reference. Previously, for most GWAS HapMap Central European (CEU) population was used as a reference, which enabled a meta-analysis of ~ 2.5 million SNPs. More recently the 1000 Genomes (http://www.1000genomes.org/) and the Genome of the Netherlands (GoNL) (http://www.bbmri.nl/en-gb/activities/projects/131-genome-of-the-netherlands) projects have provided reference populations enabling imputations of up to ~ 38 million genetic variants including single nucleotide variants (SNVs) and small insertions and deletions over the entire human genome. Thus the coverage of the study is increased not only of the common genetic variants that exert small effects on the disease but also of the rare variants that may have large effects on the disease. Discovery of both rare and common variants require large samples in a population based setting. However, families with a few affected individuals may provide a powerful base to detect large effect rare variants using linkage analysis (**Figure 2**).
Despite the growing popularity of GWAS at the cost of candidate gene research, the GWAS technology has also opened the opportunity to study candidate genes at a larger scale. Using GWAS arrays it is now possible to study families of candidate genes simultaneously.

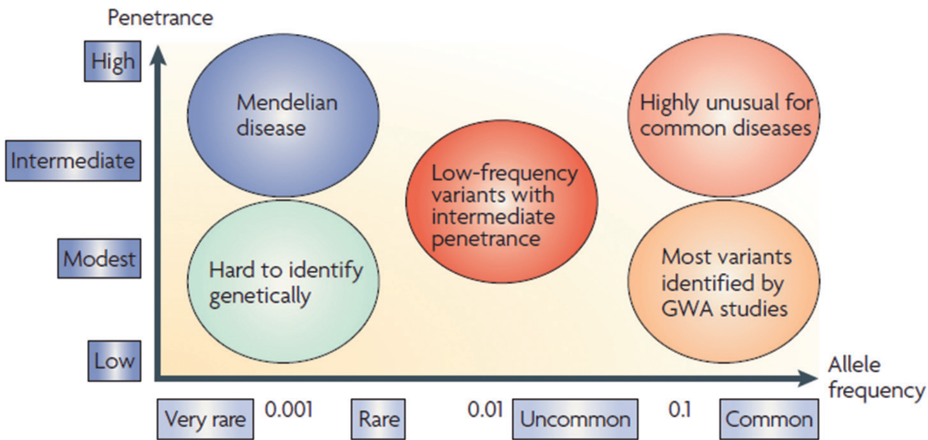


Figure 2 | Low-frequency variants and disease susceptibility (Source: McCarthy *et al*. 2008)

GENETIC EPIDEMIOLOGY OF LONGEVITY

For longevity, only two candidate genes have been consistently identified by meta-analysis, namely *APOE* and *FOXO3*. The *APOE* ε4 isoform has also been linked to elevated cholesterol, cardiovascular disease, age-related cognitive decline and dementia [16]. Thus *APOE* may influence longevity through premature onset of age-related diseases. *FOXO3* was first associated with longevity in a candidate gene study conducted in male centenarians and subsequently replicated in various samples of centenarians and long-lived individuals [20, 21]. Large scale GWAS have not been very successful in identifying genetic variants associated with longevity [1]. To date, six GWAS on longevity have been performed [3, 17, 19, 24-26]. The only consistent association in these GWAS is the *APOE* locus. Increasing the sample size may improve power to detect associations, but the lack of successful identification of genetic variants associated with longevity may also be due to the fact that GWAS (at least with the HapMap imputed ~ 2.5 million genetic variants) are better powered to detect only common variants that exert small effects on the trait. Although relatively rare variants that may exert large effects on the trait may be studied by using the imputed genetic variants from the 1000G or GoNL, the imputation of such variants is difficult and often unreliable. To date no GWAS on longevity has been performed using the imputed genetic variants from the 1000G or GoNL. NGS technology offers an expensive but a better alternate for studying rare genetic variants in relation to longevity.

As to linkage analysis in families three studies have been conducted, leading to eight chromosomal regions that are linked to longevity with LOD > 3 [27-29]. Only the region containing *APOE* has yielded the identification of genes associated with longevity.

MODEL ORGANISMS LEADING TO PATHWAYS IN LONGEVITY

Model organisms have been the major means of discovering the genetic pathways leading to longevity, providing candidate genes for further study in humans. The term “model organism” has been applied to those species, that, because of their small size and short generation times, facilitate experimental laboratory research [30]. Species such as *Caenorhabditis elegans*, *Drosophila melanogaster* and *Mus musculus* have become favorite in experimental research as they combine the advantages of a relatively short life span, low maintenance costs, and excellent background genetics [31]. Experiments in model organisms have demonstrated that changes in genes can dramatically increase their lifespan. In some cases, mean and maximum lifespan is extended up to fivefold [32]. The equivalent life-extending effect in humans would result in an average lifespan of 400 years. Such large effects are not expected in humans, however genes and pathways identified in model organisms can give hints about genetic processes involved in human longevity. As the majority of candidate genes of longevity have been identified by studies with mutants, they do not necessarily represent the effect of natural variation in the same genes that determine human longevity [32]. Natural variation in these genes likely convey more moderate effects on longevity. The most well-known and studied pathways associated with longevity in model organisms are the insulin/insulin-like growth factor signaling pathway, the stress response and senescence.

INSULIN/INSULIN-LIKE GROWTH FACTOR SIGNALING PATHWAY

The insulin/insulin-like growth factor (IGF) signalling (IIS) pathway was the first pathway discovered for longevity. The first gene in the IIS pathway linked to longevity was *daf-2* (worm insulin receptor) in *C. elegans* [33, 34]. Long-lived *daf-2* mutant worms show resistance to oxidative stress and increased thermo-tolerance [35, 36]. The increase in stress resistance of long-lived IIS mutant worms requires the presence of *daf-16*, although resistance to heat stress has been uncoupled from the IIS pathway in most model organisms [37]. Stress resistance will be described in more detail below.
The first IIS mutation in fruit flies (*D. melanogaster*) that extends lifespan was identified in 2001 [38]. $chico^{1}$ flies have a mutation in CHICO, the *Drosophila* orthologue of the insulin receptor substrate [39]. The homozygous mutant flies have low IGF-1 levels, are less than half the size of wild-type flies and have an increased median lifespan of up to 48% [38, 39]. A mutation in the *Drosophila* insulin receptor (dINR) also increases lifespan. Inr^{p5545}/Inr^{E19} dwarf females live 85% longer than controls and the dwarf males show reduced late-life, age-specific mortality [40].
All known mutations in the IIS pathway in *C. elegans* and *D. melanogaster* lead to reduced levels of IGF-1. The association between low IGF-1 levels and longevity has also been seen in mammals. Two separate dwarf mouse (*M. musculus*) mutants, the Snell and Ames

mutants, are deficient in growth hormone (GH), thyroid hormone and prolactin due to underdeveloped pituitary cells and live 40 to 65% longer than non-mutant littermates [41, 42]. Additionally, a GH receptor knock-out mouse model, whose only deficiency is a GH resistance and subsequently lower IGF-1 levels, also has an increased lifespan [43].
Thanks to intensive selective breeding in the last 300 or 400 years an enormous diversification of dog breeds has occurred. This diversification is largely due to human pressures leading to unparalleled variations in body mass and lifespan within a single species. Body mass can vary from 1 kg in the Chihuahua to over 100 kg in the Great Dane. Lifespan also varies tremendously from 8 to 15+ years in pet dogs. This diversity poses a unique opportunity to evaluate factors that contribute to lifespan. Lifespan in dogs has been shown to be inversely related to body size in multiple studies [44-47]. The finding that small dog breeds, that live longer, also have lower IGF-1 levels supports a role for the IIS pathway in this association [48].
In humans the IIS pathway (**Figure 3**) has been studied extensively for association with longevity. The IIS pathway is involved in cell survival, cell growth and proliferation in humans. The only replicated genetic variant associated with longevity that falls in the IIS pathway in humans is *FOXO3* [20, 21]. *FOXO3* is the human homologue of *daf-16*, one of the known lifespan increasing genes in *C. elegans*. Additionally there is some evidence that long-lived individuals have decreased plasma IGF-1 levels but replication are lacking [49]. Mutations in the *IGF-IR* gene have been associated with low levels of plasma IGF-1 and longevity previously [50], but this gene did not surface in GWAS of longevity. The most compelling evidence is perhaps that a composite score of IIS genes (*GHRHR*, *GH1*, *IGF1*, *INS*, *IRS1*) was found to be significantly associated with smaller body size (p-value = 0.004) and longevity (p-value = 0.047) [51]. Despite the modest p-value, this gene based analysis parallels the finding in animals, the increase in lifespan in most European countries is paralleled by an increase in height, which makes it unlikely that this pathway drives human longevity.
Within the Rotterdam Study the IIS pathway has been studied extensively by Vaessen *et al* [52-54] and will therefore not be covered in this thesis.

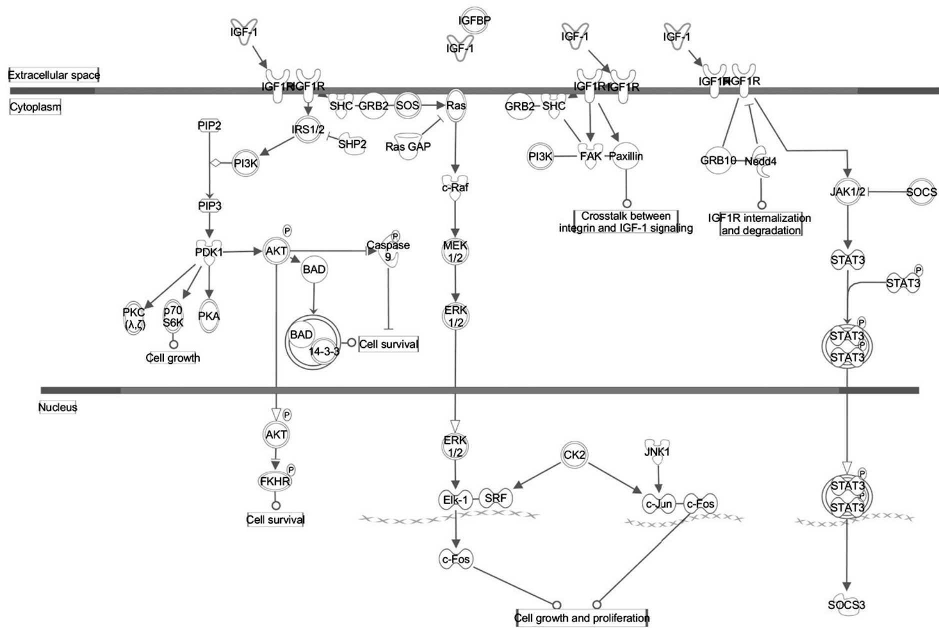


Figure 3 | Graphical representation of the human insulin/insulin-like growth factor signalling pathway. Genes are represented in their location within the pathway. In square fields additional pathways which are controlled downstream from IGF-1 signalling are depicted.

STRESS RESPONSE

The second pathway involved in longevity consistently in animal models is the response to stress. Stress, or stressors, is used here to describe experiences that are challenging emotionally and/or physiologically. Well known stressors include heat, reactive oxygen species, ischemia, inflammation and hormonal imbalance [55]. Stressors cause modifications to the intracellular milieu that may damage deoxy ribonucleic acid (DNA) and protein products. This can cause proteins to misfold, which disrupts the proteostasis. Proteostasis refers to controlling the concentration, conformation, binding interactions and location of individual proteins by readapting the innate biology of the cell [56]. Wrongly folded proteins have been associated to various age-related diseases, including Creutzfeldt-Jakob disease (prion protein), Alzheimer's disease (AD; β-amyloid) and Parkinson's disease (PD; α-Synuclein) among others [57].

Proteostasis is tightly regulated by a dedicated family of proteins known as Heat Shock Proteins (HSPs), also named stress proteins or molecular chaperones. HSPs ensure survival under stressful conditions that, if left unchecked, would lead to irreversible cell damage and ultimately to cell death [56, 58]. Abnormal expression of these proteins has been widely observed in a number of disease states, including oxidant injury, ischemia, cardiac

hypertrophy, fever, inflammation, metabolic diseases, tissue trauma, neuronal injury and cancer, as well as in experimental models of aging [59].

Thus variation in genes encoding HSPs are plausible risk factors for association with longevity, as well as genetic causes of protein folding disorders like AD and PD, yet they were not systematically investigated primarily due to the great complexity and number of HSPs, which has only recently been resolved with a new and comprehensive nomenclature [60].

SENESCENCE

A pathway involved in longevity is senescence that can now be translated to large-scale population studies through telomere metabolism [61]. Telomeres are tandem repeats of long hexamers (TTAGGG) at the end of chromosomes [62, 63]. Their predominant function is the protection of chromosome extremities from fusion and degradation [64]. The necessity of telomeres was first recognized in the late 1930s and it was only in the early 1960s that it was observed that normal human cells have a limited replicative capacity [65, 66]. The concept of the 'end replication problem', stemming from the inability of DNA polymerase to completely replicate the DNA, provided the basis for the mitotic clock hypothesis which connects the natural shortening of telomeres with age to the limited replicative capacity of a cell [67-69]. This limit corresponds to a state of mitotic senescence triggered by telomeres that have reached the shortest length compatible with stability [61]. Thus telomere length is often seen as a marker for biological age, rather than chronological age. Though telomere length can be restored by telomerase, this enzyme is not expressed at sufficient levels to maintain telomere length in humans and most other long-lived mammals [70, 71]. Therefore, telomere length shortening has been recognized as not only a marker of biological age but also a mechanism of aging [72, 73].

Convincing evidence for a causative association between telomere length and longevity stems from premature aging syndromes. Mutations in *DKC1*, a component of telomerase, are responsible for the X-linked form of Dyskeratosis congenita (DC), which has been associated with signs of premature aging [74-76]. Mutations in *TERC*, the RNA template of telomerase, are responsible for the autosomal dominant form of DC [77]. DC patients have shortened telomeres and show reduced *in vitro* telomerase activity thus pointing to telomere-dependent replicative senescence as the driving force of the disease [78-80].

Genetic research on telomere length is perhaps more interesting than longevity itself as the heritability is higher, ranging from 40 to 80% [81-83], compared to that of longevity (40%). Despite the relatively small GWAS ($N < 3500$), three loci were identified, including 18q12.2, *OBFC1* and *TERC* [84-86]. A recent GWAS including 37,684 participants identified seven loci for telomere length [87]. These include *TERC*, *TERT*, *NAF1*, *OBFC1*, *ZNF208*, *RTEL1* and *ACYP2*, confirming two of the previously identified loci. Additionally, in a gene set analysis

it was shown that genetic variation in the telomere maintenance pathway is associated to longevity [88]. It is highly likely that further genetic research will discover more loci, which could also be potentially relevant to longevity. Further genetic research with larger samples will likely discover more loci, which may serve as plausible candidates for association with longevity.

METABOLOMICS

Pathways associated with longevity in animal models, in particular the IIS pathway, are clearly linked to metabolism. Additionally, aging is often accompanied by disabilities and diseases such as cardiovascular diseases, AD, chronic joint symptoms, arthritis and diabetes, which are also known to have a strong metabolic component [89, 90]. In the Leiden Longevity Study, LDL and triglycerides were associated with longevity [91]. Metabolomics is a new technology that focuses on obtaining a more complete depiction of the metabolic status of an organism instead of single compounds [92]. Numerous small molecules are measured that represent a whole range of intermediate metabolic pathways and may serve as biomarkers [93]. Previously, metabolic profiles have been associated with age and aging-related diseases, such as type 2 diabetes, atherosclerosis, cancer and Alzheimer's Disease [94-99]. Metabolic profiles and biological age have been linked in mice [100], in humans an age-related metabolic profile of 22 metabolites explained 5% of the total variance in telomere length [96].

AIM OF THIS THESIS

The aim of this thesis is to find genetic markers associated with longevity in GWAS and candidate gene studies, as well as to gain a better understanding of telomere length metabolism.

Though in recent years great successes have been seen in complex trait genetics, for longevity no major successes have been made. In **Chapter 2** we performed a GWAS on the longevity phenotype. Longevity cases were defined as individuals of 90 years and over, while controls were defined as individuals who have died between 55 and 80 years of age (**Chapter 2.1**). In **Chapter 2.2** we aimed to unravel the genetic architecture of longevity, that is, whether rare variants with large effects or rather multiple common variants across the entire genome each exerting a small effect on longevity are predominantly determining longevity in the general population.

Chapter 3 aims to elucidate the association of candidate genes from the stress response pathway (HSPs) with longevity, here defined as all-cause mortality, as well as the association of HSPs with two age-related protein-folding disorders AD and PD. In contrast to the classical indirect candidate gene studies we targeted complete gene families based on pathways

and animal models. In **Chapter 3.1** we investigated the association of all SNPs genotyped or imputed in the HSP70 genes and its main co-chaperones with AD. HSP70 has been shown to co-localize with Aβ plaques and participates in the neuroprotective response to these plaques [101-104]. Additionally we performed pathway analysis using the SNP Ratio Test tool to identify if any HSP sub-family had more significant SNPs than would be expected by chance [105]. In **Chapter 3.2** we turned our focus to PD. This time we investigated previously identified candidates HSP10, HSP60, HSP70, HSP90 and Heat Shock Factors (HSF). Finally, in **Chapter 3.3**, we investigated the association of HSP70, small HSP and HSF families with all-cause mortality.

In **Chapter 4** we investigate telomere length as a measure of senescence. We first elucidated the heritability as well as major familial effects in telomere length in **Chapter 4.1**. In **Chapter 4.2** we investigated a previously described survivor effect in telomere length reducing the variance over age and investigated whether there is evidence for selection against short or long telomeres using a three-generation family-based approach. In **Chapter 4.3** we investigated the association between telomere length and height as observed previously in two separate animal species bridging the IIS theory of body size and the senescence hypotheses. Finally, in **Chapter 4.4** and **4.5** we investigated metabolites for their association with telomere length. First we investigated adipokines, known to be associated with oxidative stress, and their association with telomere length (**Chapter 4.4**). Next we investigated a wide range of serum metabolites using targeted metabolite profiling by electrospray ionization (ESI) tandem mass-spectrometry (BIOCRATES platform) (**Chapter 4.5**). The metabolites investigated ideally represent a wide range of biological processes therefore providing a new opportunity to better understand the telomere length metabolism.

Chapter 5 includes the general discussion of the results of this thesis. **Chapter 5.1** focuses on the difficulties of distinguishing true from false positives, a problem encountered in **Chapters 3** and **4.5**. **Chapter 5.2** discusses findings of several chapters in a broader context and suggests future research. **Chapter 6** includes a summary of the thesis. The supplementary material for the publications in this thesis (chapters 2 to 5) can be found in **Chapter 7**. Finally, the acknowledgements and a publication list can be found in C**hapter 8**.

REFERENCES

1. Murabito, J.M., R. Yuan, and K.L. Lunetta, The search for longevity and healthy aging genes: insights from epidemiological studies and samples of long-lived individuals. J Gerontol A Biol Sci Med Sci, 2012. 67(5): p. 470-9.
2. Vaupel, J.W., et al., Biodemographic trajectories of longevity. Science, 1998. 280(5365): p. 855-60.
3. Sebastiani, P., et al., Genetic signatures of exceptional longevity in humans. PLoS One, 2012. 7(1): p. e29848.
4. Herskind, A.M., et al., The heritability of human longevity: a population-based study of 2872 Danish twin pairs born 1870-1900. Hum Genet, 1996. 97(3): p. 319-23.
5. McGue, M., et al., Longevity is moderately heritable in a sample of Danish twins born 1870-1880. J Gerontol, 1993. 48(6): p. B237-44.
6. Mitchell, B.D., et al., Heritability of life span in the Old Order Amish. Am J Med Genet, 2001. 102(4): p. 346-52.
7. Westendorp, R.G., et al., Nonagenarian siblings and their offspring display lower risk of mortality and morbidity than sporadic nonagenarians: The Leiden Longevity Study. J Am Geriatr Soc, 2009. 57(9): p. 1634-7.
8. Gudmundsson, H., et al., Inheritance of human longevity in Iceland. Eur J Hum Genet, 2000. 8(10): p. 743-9.
9. Kerber, R.A., et al., Familial excess longevity in Utah genealogies. J Gerontol A Biol Sci Med Sci, 2001. 56(3): p. B130-9.
10. Perls, T., et al., Survival of parents and siblings of supercentenarians. J Gerontol A Biol Sci Med Sci, 2007. 62(9): p. 1028-34.
11. Perls, T.T., et al., Siblings of centenarians live longer. Lancet, 1998. 351(9115): p. 1560.
12. Perls, T.T., et al., Life-long sustained mortality advantage of siblings of centenarians. Proc Natl Acad Sci U S A, 2002. 99(12): p. 8442-7.
13. Schoenmaker, M., et al., Evidence of genetic enrichment for exceptional survival using a family approach: the Leiden Longevity Study. Eur J Hum Genet, 2006. 14(1): p. 79-84.
14. Newson, R.S., et al., Predicting survival and morbidity-free survival to very old age. Age (Dordr), 2010. 32(4): p. 521-34.
15. Bathum, L., et al., Apolipoprotein e genotypes: relationship to cognitive functioning, cognitive decline, and survival in nonagenarians. J Am Geriatr Soc, 2006. 54(4): p. 654-8.
16. Christensen, K., T.E. Johnson, and J.W. Vaupel, The quest for genetic determinants of human longevity: challenges and insights. Nat Rev Genet, 2006. 7(6): p. 436-48.
17. Deelen, J., et al., Genome-wide association study identifies a single major locus contributing to survival into old age; the APOE locus revisited. Aging Cell, 2011. 10(4): p. 686-98.
18. Gerdes, L.U., et al., Estimation of apolipoprotein E genotype-specific relative mortality risks from the distribution of genotypes in centenarians and middle-aged men: apolipoprotein E gene is a "frailty gene," not a "longevity gene". Genet Epidemiol, 2000. 19(3): p. 202-10.
19. Nebel, A., et al., A genome-wide association study confirms APOE as the major gene influencing survival in long-lived individuals. Mech Ageing Dev, 2011. 132(6-7): p. 324-30.
20. Soerensen, M., et al., Replication of an association of variation in the FOXO3A gene with human longevity using both case-control and longitudinal data. Aging Cell, 2010. 9(6): p. 1010-7.
21. Willcox, B.J., et al., FOXO3A genotype is strongly associated with human longevity. Proc Natl Acad Sci U S A, 2008. 105(37): p. 13987-92.
22. Carlson, C.S., et al., Mapping complex disease loci in whole-genome association studies. Nature, 2004. 429(6990): p. 446-52.
23. Cirulli, E.T. and D.B. Goldstein, Uncovering the roles of rare variants in common disease through whole-genome sequencing. Nat Rev Genet, 2010. 11(6): p. 415-25.

24. Malovini, A., et al., Association study on long-living individuals from Southern Italy identifies rs10491334 in the CAMKIV gene that regulates survival proteins. Rejuvenation Res, 2011. 14(3): p. 283-91.

25. Newman, A.B., et al., A meta-analysis of four genome-wide association studies of survival to age 90 years or older: the Cohorts for Heart and Aging Research in Genomic Epidemiology Consortium. J Gerontol A Biol Sci Med Sci, 2010. 65(5): p. 478-87.

26. Walter, S., et al., A genome-wide association study of aging. Neurobiol Aging, 2011. 32(11): p. 2109 e15-28.

27. Beekman, M., et al., Genome-wide linkage analysis for human longevity: Genetics of Healthy Aging Study. Aging Cell, 2013. 12(2): p. 184-93.

28. Kerber, R.A., et al., A genome-wide study replicates linkage of 3p22-24 to extreme longevity in humans and identifies possible additional loci. PLoS One, 2012. 7(4): p. e34746.

29. Puca, A.A., et al., A genome-wide scan for linkage to human exceptional longevity identifies a locus on chromosome 4. Proc Natl Acad Sci U S A, 2001. 98(18): p. 10505-8.

30. Hedges, S.B., The origin and evolution of model organisms. Nat Rev Genet, 2002. 3(11): p. 838-49.

31. Kirkwood, T.B., Evolution of ageing. Mech Ageing Dev, 2002. 123(7): p. 737-45.

32. Kuningas, M., et al., Genes encoding longevity: from model organisms to humans. Aging Cell, 2008. 7(2): p. 270-80.

33. Kenyon, C., et al., A C. elegans mutant that lives twice as long as wild type. Nature, 1993. 366(6454): p. 461-4.

34. Kimura, K.D., et al., daf-2, an insulin receptor-like gene that regulates longevity and diapause in Caenorhabditis elegans. Science, 1997. 277(5328): p. 942-6.

35. Larsen, P.L., Aging and resistance to oxidative damage in Caenorhabditis elegans. Proc Natl Acad Sci U S A, 1993. 90(19): p. 8905-9.

36. Lithgow, G.J., et al., Thermotolerance of a long-lived mutant of Caenorhabditis elegans. J Gerontol, 1994. 49(6): p. B270-6.

37. Libina, N., J.R. Berman, and C. Kenyon, Tissue-specific activities of C. elegans DAF-16 in the regulation of lifespan. Cell, 2003. 115(4): p. 489-502.

38. Clancy, D.J., et al., Extension of life-span by loss of CHICO, a Drosophila insulin receptor substrate protein. Science, 2001. 292(5514): p. 104-6.

39. Bohni, R., et al., Autonomous control of cell and organ size by CHICO, a Drosophila homolog of vertebrate IRS1-4. Cell, 1999. 97(7): p. 865-75.

40. Tatar, M., et al., A mutant Drosophila insulin receptor homolog that extends life-span and impairs neuroendocrine function. Science, 2001. 292(5514): p. 107-10.

41. Brown-Borg, H.M., et al., Dwarf mice and the ageing process. Nature, 1996. 384(6604): p. 33.

42. Miller, R.A., Kleemeier award lecture: are there genes for aging? J Gerontol A Biol Sci Med Sci, 1999. 54(7): p. B297-307.

43. Coschigano, K.T., et al., Assessment of growth parameters and life span of GHR/BP gene-disrupted mice. Endocrinology, 2000. 141(7): p. 2608-13.

44. Egenvall, A., et al., Age pattern of mortality in eight breeds of insured dogs in Sweden. Prev Vet Med, 2000. 46(1): p. 1-14.

45. Greer, K.A., S.C. Canterberry, and K.E. Murphy, Statistical analysis regarding the effects of height and weight on life span of the domestic dog. Res Vet Sci, 2007. 82(2): p. 208-14.

46. Li, Y., et al., Cellular proliferative capacity and life span in small and large dogs. J Gerontol A Biol Sci Med Sci, 1996. 51(6): p. B403-8.

47. Patronek, G.J., D.J. Waters, and L.T. Glickman, Comparative longevity of pet dogs and humans: implications for gerontology research. J Gerontol A Biol Sci Med Sci, 1997. 52(3): p. B171-8.

48. Shire, J.G., Growth hormone and premature ageing. Nature, 1973. 245(5422): p. 215-6.

49. Arai, Y., et al., Serum insulin-like growth factor-1 in centenarians: implications of IGF-1 as a rapid turnover protein. J Gerontol A Biol Sci Med Sci, 2001. 56(2): p. M79-82.

50. Bonafe, M., et al., Polymorphic variants of insulin-like growth factor I (IGF-I) receptor and phosphoinositide 3-kinase genes affect IGF-I plasma levels and human longevity: cues for an evolutionarily conserved mechanism of life span control. J Clin Endocrinol Metab, 2003. 88(7): p. 3299-304.

51. van Heemst, D., et al., Reduced insulin/IGF-1 signalling and human longevity. Aging Cell, 2005. 4(2): p. 79-85.

52. Rivadeneira, F., et al., Association between an insulin-like growth factor I gene promoter polymorphism and bone mineral density in the elderly: the Rotterdam Study. J Clin Endocrinol Metab, 2003. 88(8): p. 3878-84.

53. Vaessen, N., et al., A polymorphism in the gene for IGF-I: functional properties and risk for type 2 diabetes and myocardial infarction. Diabetes, 2001. 50(3): p. 637-42.

54. Vaessen, N., et al., Association between genetic variation in the gene for insulin-like growth factor-I and low birthweight. Lancet, 2002. 359(9311): p. 1036-7.

55. Macario, A.J. and E. Conway de Macario, Sick chaperones, cellular stress, and disease. N Engl J Med, 2005. 353(14): p. 1489-501.

56. Balch, W.E., et al., Adapting proteostasis for disease intervention. Science, 2008. 319(5865): p. 916-9.

57. Selkoe, D.J., Folding proteins in fatal ways. Nature, 2003. 426(6968): p. 900-4.

58. Gething, M.J. and J. Sambrook, Protein folding in the cell. Nature, 1992. 355(6355): p. 33-45.

59. Morimoto, R.I., Tissieres, A. and Georgopoulos, C., in Stress Proteins in Biology and Medicine. 1990, Cold Spring harbor Laboratory Press: Cold Spring Harbor, NY. p. 1-36.

60. Kampinga, H.H., et al., Guidelines for the nomenclature of the human heat shock proteins. Cell Stress Chaperones, 2009. 14(1): p. 105-11.

61. Shay, J.W. and W.E. Wright, Senescence and immortalization: role of telomeres and telomerase. Carcinogenesis, 2005. 26(5): p. 867-74.

62. Blackburn, E.H. and J.G. Gall, A tandemly repeated sequence at the termini of the extrachromosomal ribosomal RNA genes in Tetrahymena. J Mol Biol, 1978. 120(1): p. 33-53.

63. Blackburn, E.H., C.W. Greider, and J.W. Szostak, Telomeres and telomerase: the path from maize, Tetrahymena and yeast to human cancer and aging. Nat Med, 2006. 12(10): p. 1133-8.

64. de Lange, T., Shelterin: the protein complex that shapes and safeguards human telomeres. Genes Dev, 2005. 19(18): p. 2100-10.

65. Hayflick, L., The Limited in Vitro Lifetime of Human Diploid Cell Strains. Exp Cell Res, 1965. 37: p. 614-36.

66. McClintock, B., The Stability of Broken Ends of Chromosomes in Zea Mays. Genetics, 1941. 26(2): p. 234-82.

67. Allsopp, R.C., et al., Telomere shortening is associated with cell division in vitro and in vivo. Exp Cell Res, 1995. 220(1): p. 194-200.

68. Harley, C.B., Telomere loss: mitotic clock or genetic time bomb? Mutat Res, 1991. 256(2-6): p. 271-82.

69. Watson, J.D., Origin of concatemeric T7 DNA. Nat New Biol, 1972. 239(94): p. 197-201.

70. Gomes, N.M., et al., Comparative biology of mammalian telomeres: hypotheses on ancestral states and the roles of telomeres in longevity determination. Aging Cell, 2011. 10(5): p. 761-8.

71. Greider, C.W. and E.H. Blackburn, A telomeric sequence in the RNA of Tetrahymena telomerase required for telomere repeat synthesis. Nature, 1989. 337(6205): p. 331-7.

72. Mather, K.A., et al., Is telomere length a biomarker of aging? A review. J Gerontol A Biol Sci Med Sci, 2011. 66(2): p. 202-13.

73. von Zglinicki, T. and C.M. Martin-Ruiz, Telomeres as biomarkers for ageing and age-related diseases. Curr Mol Med, 2005. 5(2): p. 197-203.

74. Bessler, M., D.B. Wilson, and P.J. Mason, Dyskeratosis congenita and telomerase. Curr Opin Pediatr, 2004. 16(1): p. 23-8.
75. Cohen, S.B., et al., Protein composition of catalytically active human telomerase from immortal cells. Science, 2007. 315(5820): p. 1850-3.
76. Heiss, N.S., et al., X-linked dyskeratosis congenita is caused by mutations in a highly conserved gene with putative nucleolar functions. Nat Genet, 1998. 19(1): p. 32-8.
77. Vulliamy, T., et al., The RNA component of telomerase is mutated in autosomal dominant dyskeratosis congenita. Nature, 2001. 413(6854): p. 432-5.
78. Marrone, A., et al., Heterozygous telomerase RNA mutations found in dyskeratosis congenita and aplastic anemia reduce telomerase activity via haploinsufficiency. Blood, 2004. 104(13): p. 3936-42.
79. Shay, J.W. and W.E. Wright, Telomeres in dyskeratosis congenita. Nat Genet, 2004. 36(5): p. 437-8.
80. Vulliamy, T.J., et al., Very short telomeres in the peripheral blood of patients with X-linked and autosomal dyskeratosis congenita. Blood Cells Mol Dis, 2001. 27(2): p. 353-7.
81. Andrew, T., et al., Mapping genetic loci that determine leukocyte telomere length in a large sample of unselected female sibling pairs. Am J Hum Genet, 2006. 78(3): p. 480-6.
82. Bischoff, C., et al., The heritability of telomere length among the elderly and oldest-old. Twin Res Hum Genet, 2005. 8(5): p. 433-9.
83. Vasa-Nicotera, M., et al., Mapping of a major locus that determines telomere length in humans. Am J Hum Genet, 2005. 76(1): p. 147-51.
84. Codd, V., et al., Common variants near TERC are associated with mean telomere length. Nat Genet, 2010. 42(3): p. 197-9.
85. Levy, D., et al., Genome-wide association identifies OBFC1 as a locus involved in human leukocyte telomere biology. Proc Natl Acad Sci U S A, 2010. 107(20): p. 9293-8.
86. Mangino, M., et al., A genome-wide association study identifies a novel locus on chromosome 18q12.2 influencing white cell telomere length. J Med Genet, 2009. 46(7): p. 451-4.
87. Codd, V., et al., Identification of seven loci affecting mean telomere length and their association with disease. Nat Genet, 2013. 45(4): p. 422-7, 427e1-2.
88. Deelen, J., et al., Gene set analysis of GWAS data for human longevity highlights the relevance of the insulin/IGF-1 signaling and telomere maintenance pathways. Age (Dordr), 2013. 35(1): p. 235-49.
89. Butler, R.N., et al., New model of health promotion and disease prevention for the 21st century. BMJ, 2008. 337: p. a399.
90. Wijsman, C.A., et al., Familial longevity is marked by enhanced insulin sensitivity. Aging Cell, 2011. 10(1): p. 114-21.
91. Vaarhorst, A.A., et al., Lipid metabolism in long-lived families: the Leiden Longevity Study. Age (Dordr), 2011. 33(2): p. 219-27.
92. Psychogios, N., et al., The human serum metabolome. PLoS One, 2011. 6(2): p. e16957.
93. He, Y., et al., Schizophrenia shows a unique metabolomics signature in plasma. Transl Psychiatry, 2012. 2: p. e149.
94. Cao, M., et al., NMR-based metabolomic analysis of human bladder cancer. Anal Sci, 2012. 28(5): p. 451-6.
95. Matsumoto, T., T. Kobayashi, and K. Kamata, Role of lysophosphatidylcholine (LPC) in atherosclerosis. Curr Med Chem, 2007. 14(30): p. 3209-20.
96. Menni, C., et al., Metabolomic markers reveal novel pathways of ageing and early development in human populations. Int J Epidemiol, 2013.
97. Oresic, M., et al., Metabolome in progression to Alzheimer's disease. Transl Psychiatry, 2011. 1: p. e57.
98. Suhre, K., et al., Metabolic footprint of diabetes: a multiplatform metabolomics study in an epidemiological setting. PLoS One, 2010. 5(11): p. e13953.

99. Yu, Z., et al., Human serum metabolic profiles are age dependent. Aging Cell, 2012. 11(6): p. 960-7.
100. Tomas-Loba, A., et al., A metabolic signature predicts biological age in mice. Aging Cell, 2013. 12(1): p. 93-101.
101. Muchowski, P.J. and J.L. Wacker, Modulation of neurodegeneration by molecular chaperones. Nat Rev Neurosci, 2005. 6(1): p. 11-22.
102. Kakimura, J., et al., Microglial activation and amyloid-beta clearance induced by exogenous heat-shock proteins. FASEB J, 2002. 16(6): p. 601-3.
103. Magrane, J., et al., Heat shock protein 70 participates in the neuroprotective response to intracellularly expressed beta-amyloid in neurons. J Neurosci, 2004. 24(7): p. 1700-6.
104. Sherman, M.Y. and A.L. Goldberg, Cellular defenses against unfolded proteins: a cell biologist thinks about neurodegenerative diseases. Neuron, 2001. 29(1): p. 15-32.
105. O'Dushlaine, C., et al., The SNP ratio test: pathway analysis of genome-wide association datasets. Bioinformatics, 2009. 25(20): p. 2762-3.

Chapter 2

Genome Wide Association Studies of longevity

Chapter 2.1

GWAS of longevity in the CHARGE consortium confirms *APOE* and *FOXO3*

Linda Broer, Aron S Buchman , Joris Deelen, Daniel S Evans, Jessica D Faul, Kathryn L Lunetta, Paola Sebastiani, Jennifer A Smith, Albert V Smith, Toshiko Tanaka, Lei Yu, Alice M Arnold, Thor Aspelund, Emelia J Benjamin, Phil L De Jager, Gudny Eirkisdottir, Denis A Evans, Melissa E Garcia, Albert Hofman, Robert C Kaplan, Sharon LR Kardia, Douglas P Kiel, Ben A Oostra, Eric S Orwoll, Neeta Parimi, Bruce M Psaty, Fernando Rivadeneira, Jerome Rotter, Sudha Seshadri, Andrew Singleton, Henning Tiemeier, André G Uitterlinden, Wei Zhao, Stefania Bandinelli, David A Bennett, Luigi Ferrucci, Vilmundur Gudnason, Tamara B Harris, David Karasik, Lenore J Launer, Anne B Newman, Thomas T Perls, P. Eline Slagboom, Gregory J Tranah, David R Weir, Cornelia M van Duijn, Joanne M Murabito

ABSTRACT

The genetic contribution to longevity in humans has been estimated to range from 15 to 25%. Only two genes, *APOE* and *FOXO3*, have shown consistent results across studies.
In this genome-wide association study we included 6,036 longevity cases that reached age 90 years or older and 3,757 controls that died between ages 55 and 80 years. We additionally attempted to replicate previously identified SNP associations with longevity.
We found suggestive evidence for the association of SNPs near *CADM2* and *GRIK2* and longevity. We further confirmed the association of *APOE* and *FOXO3* with longevity, but did not replicate other previously identified associations.
An important conclusion of this study is the lack of finding new associations with longevity to age 90+. There are several potential reasons why we did not observe definitive new loci. Longevity may represent multiple complex traits with heterogeneous genetic underpinnings. Alternatively, it could be that longevity is regulated by rare variants rather than common variants.

INTRODUCTION

There is ample evidence that genetic factors are involved in extreme longevity both in humans and in other organisms. In model organisms, ranging from *C. elegans* to mice, mutations in the insulin/IGF-1 signaling (IIS) pathway have been shown to substantially increase lifespan [1-6]. Other suggested mechanisms involve stress resistance as mediated by heat shock proteins [7, 8], cellular senescence as mediated by telomere length [9] and inflammation/immune function [10]. The genetic contribution to lifespan in humans has been estimated to range from 15 to 25% [11-13].

Though many human candidate longevity genes have been investigated, only two genes have shown consistent results across studies. The first consistent association found was *APOE* [14]. It was subsequently replicated in both candidate gene and genome-wide association studies (GWAS) with genome-wide significant evidence [15-20]. The second longevity gene identified was *FOXO3* in male centenarians and replicated in other candidate gene studies that include both men and women across ethnicities, though *FOXO3* was not identified in GWAS [21-25].

Early GWAS have failed to identify new longevity genes as reviewed by Murabito *et al* [26]. In 2010 the CHARGE consortium published a meta-analysis of 1,836 longevity (90+) cases and a comparison group of 1,955 individuals whom died between 55 and 80 years [27]. A total of 24 independent SNPs were identified, though none reached genome-wide significance.

The key to success to GWAS has proven to be large sample size and statistical power. Since the original CHARGE GWAS [27] additional members of each cohort have reached or exceeded age 90 and additional studies have joined the consortium permitting an expanded meta-analysis of longevity. Using the same study design as the original CHARGE longevity study, the sample size of the current study is 6,036 longevity cases and 3,757 controls.

METHODS

Participants

The participants of this study are of European ancestry and include cohorts of the CHARGE Consortium [28]. All cohorts periodically assess the vital status of their participants. Although some of the cohorts include multiple ethnic groups, only data from self-reported Caucasians were used. Informed written consent was obtained from all participants and appropriate institutional approvals were obtained. A brief description of each population is given in the **Supplementary Material.**

Longevity Phenotype

In the present study longevity was defined as reaching age 90 years or older. Genotyped participants who died between the ages of 55 and 80 years were used as the control group. The control group was limited to deceased participants to ensure that no one in the control

group could subsequently achieve longevity. The minimum age at death was set to match the minimum age at enrollment in the Rotterdam Study (RS). The maximal age at death was set arbitrarily at age 80 years to include the majority of deaths and to exclude those participants who survived far beyond the average life expectancy for their respective birth cohort. Across the cohorts there were 6,036 participants who achieved longevity and the control group had 3,757 participants.

Genotyping and imputation

As different genotyping platforms were used across studies, we imputed to 2.5 million SNPs using the HapMap 22 CEU (Build 36) genotyped samples as a reference. Details on the study-specific quality control procedures for genotyping and imputation can be found in **Supplementary Table 1.**

Statistical analysis

We used logistic regression to test each SNP for association with longevity using an additive model adjusting for sex. Fixed-effects inverse-variance meta-analysis was performed using METAL. P-values were corrected for genomic control. Associations with a p-value $< 5*10^{-8}$ were considered to be significant, while associations with a p-value $< 1*10^{-5}$ were considered to be suggestive and proceeded to replication.

Detailed analysis of replication cohorts and additional analyses can be found in the **Supplementary Material**. Conditional analysis of the *FOXO3* locus was performed using the GCTA tool without individual level data [29, 30]. We further attempted to replicate previously identified SNP associations with longevity. Finally, we used gene networks available at www.genenetworks.nl as a bioinformatics resource to further investigate significant findings.

RESULTS

General characteristics of discovery and replication cohorts can be found in **Table 1** and **Table 2** respectively. Between 0.0% and 70.3% of those achieving longevity were still alive at the time that longevity status was ascertained. Among those who had died, the distribution of causes of death differed between longevity cases and the comparison group. Whereas 1.0% to 17.3% of those achieving longevity died of cancer, 10.1% to 50.3% of deaths in the comparison group could be attributed to cancer.

Table 1 | General characteristics of discovery cohorts.

		N	Women, %	Alive, %	Cause of death, %*			
					CVD	Cancer	Other	Unknown
RS1	90+ cases	899	79.0	23.6	31.4	7.9	60.7	0.0
	comparison	1192	41.3	0.0	32.2	40.1	27.7	0.0
RS2	90+ cases	69	60.9	69.6	33.3	14.3	52.4	0.0
	comparison	161	37.3	0.0	23.6	50.3	26.1	0.0
SOF**	90+ cases	1720	100.0	54.4	16.8	3.6	21.5	0.0
	comparison	124	100.0	0.0	37.9	28.2	33.9	0.0
CHS	90+ cases	791	62.3	41.7	51.5	17.3	30.6	0.6
	comparison	560	53.2	0.0	33.6	39.3	27.0	0.2
MrOS***	90+ cases	670	0.0	43.6	19.9	6.9	21.8	3.1
	comparison	502	0.0	0.0	29.3	38.8	31.7	0.2
FHS	90+ cases	320	66.9	0.0	24.1	11.6	51.9	12.5
	comparison	484	34.3	0.0	24.6	43.8	25.2	6.4
HRS	90+ cases	384	67.5	68.5	3.0	15.0	12.6	69.5
	comparison	401	45.6	0.0	12.1	10.1	16.2	61.6
AGES	90+ cases	541	61.2	65.1	27.3	8.4	16.2	48.1
	comparison	145	49.0	0.0	28.2	35.5	18.4	18.0
RADC	90+ cases	468	75.2	40.0	0.0	0.0	0.0	100.0
	comparison	78	48.7	0.0	0.0	0.0	0.0	100.0
InCHIANTI	90+ cases	101	72.3	70.3	8.9	3.0	7.9	80.2
	comparison	75	36.0	0.0	6.7	20.0	4.0	69.3
BLSA	90+ cases	128	49.2	50.0	4.7	1.0	3.0	91.4
	comparison	42	26.2	0.0	9.5	19.0	2.3	69.1

RS1: Rotterdam Study 1; RS2: Rotterdam Study 2; SOF: Study of Osteoporotic Fractures; CHS: Cardiovascular Health Study; MrOS: Osteoporotic Fractures in Men Study; FHS: Framingham Heart Study; HRS: Health and Retirement Study; AGES: Age, Gene/Environment Susceptibility -Reykjavik Study; RADC: Religious Orders Study and Rush Memory and Aging Project; InCHIANTI: Invecchiare nel Chianti; BLSA: Baltimore Longitudinal Study of Ageing

* In 90+ cases the % of deaths is reported

** In SOF the vital status is not known for all individuals

*** In MrOS the vital status is not known for all individuals

2

Table 2 | General characteristics of replication cohorts.

	NECS		LLFS
	cases (100+)	comparison	survival analysis
N	801	914	4567
Age at DNA draw (median (range))	104 (95-119)	73 (54-90)	67 (25-110)
Follow-up (median (range))	NA		3 (0-7)
Women, %	72.0	56.0	55.0
Alive, %	3.0	0.0	83.0
Cause of death, %*			
CVD	0.0	0.0	0.0
Cancer	0.0	0.0	0.0
Other	0.0	0.0	0.0
Unknown	100.0	100.0	100.0

LLFS: Long Life Family Study; No study specific information for the EU_longevity consortium was available.

* In 100+ cases the % of deaths is reported

None of the SNP-longevity associations reached the genome-wide significance threshold of $5*10^{-8}$ in the discovery phase (**Supplementary Figure 1**), although the strongest association was for rs1416280 located 369kb from *GRIK2* (OR = 1.24; p-value = $5.09*10^{-8}$). In total, 7 loci passed the threshold for suggestive association (< $1*10^{-5}$) and were included in the replication phase (**Table 3**). For replication, we also included the most significant SNP in the current study from the known longevity gene, *FOXO3*. Forest plots of all 8 SNPs can be found in **Supplementary Figure 2**. Though none of the SNPs reached Bonferroni corrected significance in any of the replication cohorts, consistent results considering direction of effect were found for *FOXO3* across discovery and replication cohorts (**Table 3**). Association results for *CADM2* and, to a lesser degree, *GRIK2* were consistent across cohorts. In pathway analysis using the gene networks tool, both *CADM2* and *GRIK2* genes were found to be involved in neuronal pathways (**Supplementary Table 2**).

Of the 24 SNPs identified in the original CHARGE longevity GWAS [27] the p-value for only one of these (*GRIK2*; rs954551) improved in the current study (**Supplementary Table 3**). One further region (*RGS7*) appears among the strongest associations in the current study, but the association was not strengthened in the replication phase of the study.

Next we aimed to replicate the GWAS of Sebastiani *et al* conducted on centenarians (**Supplementary Table 4**) [20]. 281 SNPs were reported using four different genetic models [20]. 6 out of the 281 SNPs were reported to have a p-value < $1*10^{-5}$ in the additive model. We attempted to replicate these 6 SNPs, and only the SNP at the *APOE* locus replicated at the Bonferroni corrected significance level (OR = 1.20; p-value = $4.8*10^{-4}$). Similarly, of the 4 regions identified with linkage analysis of longevity sib-pairs (90+) [31] we also replicated only the *APOE* locus (**Supplementary Table 5**).

Table 3 | Association results for longevity: discovery and replication.

SNP	Chr	Gene	Distance (kb)	EA	EAF	Discovery p-value	Discovery OR	Discovery 95% CI	Discovery direction	EU_longevity effect	EU_longevity p-value	Centenarian (NECS) SNP	Centenarian (NECS) r2*	Centenarian (NECS) OR	Centenarian (NECS) p-value	Survival (LLFS) SNP	Survival (LLFS) r2*	Survival (LLFS) 1/HR	Survival (LLFS) p-value
rs1416280	6	GRIK2	369	C	0.75	$5.09*10^{-8}$	1.24	1.15-1.34	+-+++++++-+	+	0.276	rs9377361	1.00	0.98	0.833	rs9377361	1.00	1.00	0.981
rs9841144	3	CADM2	-236	A	0.79	$9.66*10^{-7}$	0.81	0.74-0.88	---------+-	+	0.825	rs9822731	0.96	1.03	0.726	rs9822731	0.96	0.85	0.053
rs4611001	1	RGS7	-28	A	0.97	$1.84*10^{-6}$	1.79	1.41-2.27	+?-++++++?+	-	0.207	rs7536260	0.68	0.65	**0.013**	rs75361849	1.00	1.07	0.728
rs11023737	11	SOX6	-28	A	0.32	$3.64*10^{-6}$	0.83	0.77-0.90	-----------	+	0.684	rs2196961	0.27	1.03	0.668	rs11023744	0.96	1.06	0.397
rs11753077	6	MBOAT1	-76	T	0.64	$7.51*10^{-6}$	1.17	1.09-1.26	+++++++-+-+	+	0.095	rs7763815	0.87	1.04	0.572	same	1.00	1.03	0.633
rs10875746	12	PFKM	intron	A	0.76	$7.83*10^{-6}$	1.20	1.11-1.30	+++++++++-+	+	0.240	rs2228500	0.74	1.03	0.796	rs2228500	0.74	0.93	0.349
rs10007810	4	LIMCH1	intron	A	0.23	$8.80*10^{-6}$	1.20	1.11-1.30	+-+++++++++	-	0.979	same	1.00	0.98	0.841	same	1.00	0.95	0.521
rs10457180	6	FOXO3	intron	A	0.70	$8.56*10^{-5}$	0.87	0.81-0.93	--+-----+--	-	**0.023**	rs2153960	0.89	0.98	0.822	rs2153960	0.89	0.84	**0.021**

EU_longevity: European consortium investigating the genetics of longevity with similar phenotype to discovery (cases: 90+; controls: <65); Centenarians (NECS): centenarian cases compared to younger controls; Survival (LLFS): all-cause mortality Cox-regression analysis, 1/HR shown for direct comparison of the direction of the effect. In NECS and LLFS no imputations were available.

*linkage disequilibrium with discovery SNP

Study order for direction column: RS1, RS2, SOF, CHS, MrOS, FHS, HRS, AGES, RADC, InCHIANTI, BLSA

EA: effect allele; EAF: effect allele frequency; '+': effect allele overrepresented in cases; '-': effect allele underrepresented in cases; '?': not tested in this study.

P-value < 0.05 in replication are bolded

2

Finally, we examined the original SNP reported for the *FOXO3* gene (rs2802292) in the current study and added our current study (each cohort individually) to the 4 published studies reporting on the same *FOXO3* SNP (**Supplementary Table 6**) [21, 23-25]. In the current study, rs2802292 reached a p-value of 0.012 (OR = 1.09). When including the published studies in the meta-analysis rs2802292 passed the threshold for genome-wide significance (OR = 1.18; p-value = $1.44*10^{-8}$). The Linkage Disequilibrium (LD) as measured by r^2 between this SNP and the top *FOXO3* SNP in the current study (rs10457180) was 0.64. In conditional analyses including both SNPs only rs10457180 remained significant (OR = 0.94; p-value = $5.53*10^{-4}$).

DISCUSSION

In this investigation of 6,036 longevity cases and 3,757 controls we find suggestive evidence for the involvement of SNPs near *CADM2* and *GRIK2*. We further confirmed the associations of *APOE* and *FOXO3* with longevity.

FOXO3 is a known candidate gene for longevity and part of the well characterized IIS pathway [6]. Though previously identified in candidate gene studies it did not previously reach genome-wide significance [21-25]. In our study the strongest signal in the *FOXO3* gene was found for rs10457180 which is in LD (r^2 = 0.64) with the known SNP rs2802292. Pooling our data with that of independent studies from literature yielded a genome-wide significant p-value of $1.44*10^{-8}$. In a conditional analysis only rs10457180 remained significant. This suggests that rs10457180 may be a better tagging SNP for the true causal variant.

Though replication for both *CADM2* and *GRIK2* was not consistent, these genes are interesting candidate genes for longevity. Both genes are involved in neuronal pathways, in particular in neuron cell-cell adhesion and regulation of synaptic transmission, glutamatergic (www.genenetworks.nl). Neuronal pathways have been previously implicated in all-cause mortality in a pathway analysis [32].

We did not replicate other previously reported associations with longevity, except for *FOXO3* and *APOE*. Though many attempts to unravel the genetic contribution to the longevity phenotype have been undertaken, none of the previously suggestive longevity genes identified by GWAS have ever been replicated [26]. GWAS has proven to be able to unravel the genetics of many complex traits with great success, even when the heritability of the trait is modest, as is the case in longevity. Yet the longevity phenotype remains resistant to the efforts to uncover new genetic associations. There are several potential explanations for this. First, the longevity phenotype could simply be too complex and heterogeneous for successful identification of longevity genes. It may require a great number of 'protective' genes all with small effects to have a genetic advantage to achieve longevity [20]. Though in highly heritable traits like height these genes have been uncovered, in a trait like longevity it may be nearly impossible with the current availability of genotyped individuals reaching longevity. Using biomarkers might be a more fruitful pursuit for finding associations to

longevity. Unfortunately no good biomarkers of aging currently exist, though many have been proposed [33]. Telomere length, a marker of senescence, could be an interesting biomarker of aging and GWAS has already proven successful in identifying genes for this trait [34, 35]. An alternative approach could be the development of novel phenotypes like the healthy aging index, which incorporates information across physiologic systems and predicts mortality better than age itself [36].

The second potential reason for not finding any genome-wide significant associations with the longevity phenotype is that GWAS targets common variants [37]. As longevity in the general population is quite rare, but common in specific long-lived families, it could be that rare variants, rather than common variants are involved in the longevity phenotype [38]. Such variants are not likely to be picked up in GWAS, but may be uncovered with next generation sequencing technology [39].

Finally, another reason for not finding significant associations with longevity could be found in epigenetics, which could influence the expression of genetic variations. Epigenetic modifications such as DNA methylation and histone modification are essential for development and differentiation, but can also arise later in life [40]. Age related changes in DNA methylation have been implicated in senescence and longevity [41]. Such changes could mask genetic associations.

In conclusion, we confirm the association of *FOXO3* and *APOE* with longevity, but do not find any new longevity associations, aside from the suggestive evidence for *CADM2* and *GRIK2*.

REFERENCES

1. Bohni, R., et al., Autonomous control of cell and organ size by CHICO, a Drosophila homolog of vertebrate IRS1-4. Cell, 1999. 97(7): p. 865-75.
2. Clancy, D.J., et al., Extension of life-span by loss of CHICO, a Drosophila insulin receptor substrate protein. Science, 2001. 292(5514): p. 104-6.
3. Coschigano, K.T., et al., Assessment of growth parameters and life span of GHR/BP gene-disrupted mice. Endocrinology, 2000. 141(7): p. 2608-13.
4. Kenyon, C., et al., A C. elegans mutant that lives twice as long as wild type. Nature, 1993. 366(6454): p. 461-4.
5. Kimura, K.D., et al., daf-2, an insulin receptor-like gene that regulates longevity and diapause in Caenorhabditis elegans. Science, 1997. 277(5328): p. 942-6.
6. Libina, N., J.R. Berman, and C. Kenyon, Tissue-specific activities of C. elegans DAF-16 in the regulation of lifespan. Cell, 2003. 115(4): p. 489-502.
7. Balch, W.E., et al., Adapting proteostasis for disease intervention. Science, 2008. 319(5865): p. 916-9.
8. Gething, M.J. and J. Sambrook, Protein folding in the cell. Nature, 1992. 355(6355): p. 33-45.
9. Shay, J.W. and W.E. Wright, Senescence and immortalization: role of telomeres and telomerase. Carcinogenesis, 2005. 26(5): p. 867-74.
10. Jeck, W.R., A.P. Siebold, and N.E. Sharpless, Review: a meta-analysis of GWAS and age-associated diseases. Aging Cell, 2012. 11(5): p. 727-31.
11. Herskind, A.M., et al., The heritability of human longevity: a population-based study of 2872 Danish twin pairs born 1870-1900. Hum Genet, 1996. 97(3): p. 319-23.
12. McGue, M., et al., Longevity is moderately heritable in a sample of Danish twins born 1870-1880. J Gerontol, 1993. 48(6): p. B237-44.
13. Mitchell, B.D., et al., Heritability of life span in the Old Order Amish. Am J Med Genet, 2001. 102(4): p. 346-52.
14. Schachter, F., et al., Genetic associations with human longevity at the APOE and ACE loci. Nat Genet, 1994. 6(1): p. 29-32.
15. Bathum, L., et al., Apolipoprotein e genotypes: relationship to cognitive functioning, cognitive decline, and survival in nonagenarians. J Am Geriatr Soc, 2006. 54(4): p. 654-8.
16. Christensen, K., T.E. Johnson, and J.W. Vaupel, The quest for genetic determinants of human longevity: challenges and insights. Nat Rev Genet, 2006. 7(6): p. 436-48.
17. Deelen, J., et al., Genome-wide association study identifies a single major locus contributing to survival into old age; the APOE locus revisited. Aging Cell, 2011. 10(4): p. 686-98.
18. Gerdes, L.U., et al., Estimation of apolipoprotein E genotype-specific relative mortality risks from the distribution of genotypes in centenarians and middle-aged men: apolipoprotein E gene is a "frailty gene," not a "longevity gene". Genet Epidemiol, 2000. 19(3): p. 202-10.
19. Nebel, A., et al., A genome-wide association study confirms APOE as the major gene influencing survival in long-lived individuals. Mech Ageing Dev, 2011. 132(6-7): p. 324-30.
20. Sebastiani, P., et al., Genetic signatures of exceptional longevity in humans. PLoS One, 2012. 7(1): p. e29848.
21. Anselmi, C.V., et al., Association of the FOXO3A locus with extreme longevity in a southern Italian centenarian study. Rejuvenation Res, 2009. 12(2): p. 95-104.
22. Flachsbart, F., et al., Association of FOXO3A variation with human longevity confirmed in German centenarians. Proc Natl Acad Sci U S A, 2009. 106(8): p. 2700-5.
23. Li, Y., et al., Genetic association of FOXO1A and FOXO3A with longevity trait in Han Chinese populations. Hum Mol Genet, 2009. 18(24): p. 4897-904.

24. Soerensen, M., et al., Replication of an association of variation in the FOXO3A gene with human longevity using both case-control and longitudinal data. Aging Cell, 2010. 9(6): p. 1010-7.
25. Willcox, B.J., et al., FOXO3A genotype is strongly associated with human longevity. Proc Natl Acad Sci U S A, 2008. 105(37): p. 13987-92.
26. Murabito, J.M., R. Yuan, and K.L. Lunetta, The search for longevity and healthy aging genes: insights from epidemiological studies and samples of long-lived individuals. J Gerontol A Biol Sci Med Sci, 2012. 67(5): p. 470-9.
27. Newman, A.B., et al., A meta-analysis of four genome-wide association studies of survival to age 90 years or older: the Cohorts for Heart and Aging Research in Genomic Epidemiology Consortium. J Gerontol A Biol Sci Med Sci, 2010. 65(5): p. 478-87.
28. Psaty, B.M., et al., Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium: Design of prospective meta-analyses of genome-wide association studies from 5 cohorts. Circ Cardiovasc Genet, 2009. 2(1): p. 73-80.
29. Yang, J., et al., Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. Nat Genet, 2012. 44(4): p. 369-75, S1-3.
30. Yang, J., et al., GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet, 2011. 88(1): p. 76-82.
31. Beekman, M., et al., Genome-wide linkage analysis for human longevity: Genetics of Healthy Aging Study. Aging Cell, 2013. 12(2): p. 184-93.
32. Walter, S., et al., A genome-wide association study of aging. Neurobiol Aging, 2011. 32(11): p. 2109 e15-28.
33. Johnson, T.E., Recent results: biomarkers of aging. Exp Gerontol, 2006. 41(12): p. 1243-6.
34. Codd, V., et al., Identification of seven loci affecting mean telomere length and their association with disease. Nat Genet, 2013. 45(4): p. 422-7.
35. von Zglinicki, T. and C.M. Martin-Ruiz, Telomeres as biomarkers for ageing and age-related diseases. Curr Mol Med, 2005. 5(2): p. 197-203.
36. Newman, A.B., et al., A physiologic index of comorbidity: relationship to mortality and disability. J Gerontol A Biol Sci Med Sci, 2008. 63(6): p. 603-9.
37. Maher, B., Personal genomes: The case of the missing heritability. Nature, 2008. 456(7218): p. 18-21.
38. Westendorp, R.G., et al., Nonagenarian siblings and their offspring display lower risk of mortality and morbidity than sporadic nonagenarians: The Leiden Longevity Study. J Am Geriatr Soc, 2009. 57(9): p. 1634-7.
39. Cirulli, E.T. and D.B. Goldstein, Uncovering the roles of rare variants in common disease through whole-genome sequencing. Nat Rev Genet, 2010. 11(6): p. 415-25.
40. Gravina, S. and J. Vijg, Epigenetic factors in aging and longevity. Pflugers Arch, 2010. 459(2): p. 247-58.
41. Bell, J.T. et al., Epigenome-wide scans identify differentially methylated regions fora ge and age-related phenotypes in a healthy ageing population. PLoS Genet, 2012. 8(4): p.e1002629.

Chapter 2.2

The genetic architecture of longevity in a population-based study, the Rotterdam Study

Linda Broer, Najaf Amin, André G. Uitterlinden, Ben A. Oostra, Albert Hofman, Henning Tiemeier, Cornelia M. van Duijn

ABSTRACT

Genome-wide association studies (GWAS) of longevity have failed to identify any new genes beyond the known candidate genes *FOXO3* and *APOE*. It has been postulated that longevity is caused by rare variants rather than common variants. An alternative hypothesis is that a very large number of common variants together determine longevity.

In this study we investigated the association of rare variants, as detected by exome sequencing, with longevity in a population-based study, the Rotterdam Study (RS-I; n = 202 longevity cases and 267 controls) and replicated using 1000G imputations in the remaining individuals of the study (802 longevity cases and 864 controls). Additionally, we determined the extent to which the heritability of longevity can be explained by common variants using GCTA using the genome-wide array data.

In single variant analysis none of the exonic variants approached genome-wide significance. Pathway analysis of the single variants using PANTHER showed the involvement of cell cycle control mechanisms in longevity. Burden test analysis showed borderline genome-wide significant association (p-value = $2.87*10^{-6}$) with the gene *WSCD2* with longevity. *OR52L1* is the only gene emerging in both the single variant and burden test analyses, but findings are far from significant. When testing whether the joint effect of multiple common variants explain the heritability we found that up to 81.3% of the heritability can be explained by common variants (p-value = 0.01).

Our study suggests that common variants explain a large part of the heritability of longevity. We found suggestive evidence for the involvement of rare variants in *WSCD2* and *OR52L1* in longevity, but these findings remain to be confirmed.

INTRODUCTION

In the past two centuries mean life expectancy has significantly increased [1]. In parallel, the number of individuals reaching extreme age (e.g. longevity) has increased as well. The percentage of individuals reaching 90 years of age is growing dramatically [2]. Heritability estimates of age at death range from 20 to 30% in twin registries [3, 4] and 15 to 25% in population-based samples [5, 6]. Heritability of longevity, on the other hand, was estimated to be higher, approaching 40% [7].

Despite the substantial heritability, it has been difficult to find the genes involved in longevity. As early as 1994 *APOE* was identified as one of the key longevity genes in early candidate gene studies [8] and was subsequently replicated in recent candidate gene and genome-wide association studies (GWAS) reaching genome-wide significance [9-14]. Also the second longevity gene identified, *FOXO3*, was associated in candidate gene studies across ethnicities, though it did not reach genome-wide significance in GWAS [15-19]. Other than these candidate genes, GWAS has failed to identify any consistent associations with longevity [11, 13, 14, 20-22].

As longevity is common in specific long-lived families, it has been postulated that rare, rather than common variants are involved in the longevity phenotype [23]. Such variants are not likely to be picked up in GWAS, but may be uncovered with next generation sequencing technology [24]. On the other hand, for an increasing number of traits (e.g. height) that are highly heritable and do cluster in families there is increasing evidence that the heritability is explained by a large number of common variants throughout the genome [25] which are difficult to detect in GWAS as individual effects of genes are small.

In this study we aimed to investigate the genetic architecture of longevity in the Rotterdam Study (RS-I). We determined the extent to which the heritability of longevity could be explained by rare exonic variants with large effects rather than common variants with small effects.

METHODS

Study population

From 1991 to 1995 all inhabitants of Ommoord, a district of Rotterdam, The Netherlands, who were 55 years or older, were invited to participate in the RS-I [26, 27]. Genotyping information was available for 5,974 participants. All of the participants were followed for incident diseases through linkage to the general practitioner data base and record review by trained medical investigators. General practitioners', hospital records as well as death certificates were used for identification of deaths and health events through 01.01.2011.

In the present study longevity was defined as reaching age 90 years or older. Participants who died between the ages of 55 and 80 years were used as the control group. The control group was limited to deceased participants to ensure that no one in the control group could

subsequently achieve longevity. The minimum age at death was set to match the minimum age at enrollment in the study. The maximal age at death was set at age 80 years to include the majority of deaths and at the same time to exclude those participants who survived far beyond the average life expectancy for their respective birth cohort.

In total the RS-I population included 1,004 longevity cases and 1,131 controls. For discovery of rare variants we used those samples exome sequenced, which included 202 longevity cases and 267 controls. For replication we included those samples not included in exome sequencing, but were genotyped on genome-wide array platforms (802 longevity cases and 864 controls) and subsequently imputed to the 1000 G reference panel for rare variants. Details on the measurement of general characteristics can be found elsewhere [28]. Summary statistics were calculated using SPSS 20. Unadjusted p-values were derived using the one-way anova test for continuous variables or the chi-square test for categorical variables. P-values adjusted for sex were obtained in logistic regression analysis.

Genotyping/Sequencing and imputations

Genomic DNA was extracted from whole blood samples using standard methods [29]. Exomes of 1,764 individuals from the RS-I population were sequenced, using the Nimblegen SeqCap EZ V2 capture kit on an Illumina Hiseq2000 sequencer and the TrueSeq Version3 protocol. The sequences reads were aligned to the human genome build 19 (hg19) using Burrows-Wheeler Aligner [30]. Genetic variants were called using the Unified Genotyper Tool from GATK. Samples with low concordance to genotyping array (< 95%), high heterozygote to homozygote ratio (> 2.0) and low transition/transversion ratio (< 2.3) were removed from the data. The final dataset consisted of 1,524 individuals and about 900 thousand Single Nucleotide Variants (SNVs). Annotation of the sequenced variants was performed using the SeattleSeq database (http://snp.gs.washington.edu/SeattleSeqAnnotation137/) and functional prediction of the variations were performed using Polyphen v2 [33]. For discovery we used those individuals included in the exome sequencing, including 202 longevity cases and 267 controls.

For replication we used individuals genotyped on the Illumina Infinium II HumanHap550 SNP chip array who were not included in the exome sequencing dataset, including 802 longevity cases and 864 controls. Imputations on the 1000G reference panel were performed using MACH software (http://www.sph.umich.edu/csg/abecasis/MACH/). Analysis was performed using ProbABEL [34].

Statistical analysis of rare variants

In order to assess the association of rare variants with longevity we performed single variant analysis of exome data in PLINK v1.07 software, logistic regression adjusting for sex [35]. Multiple testing correction on the number of SNVs tested was performed, resulting in a genome-wide significance threshold of $5.5*10^{-8}$ (0.05/903,316 variants).

The combined effect/burden of rare variants was assessed using SKAT v0.83, adjusted for sex [36]. Multiple testing correction on the number of genes resulted in a genome-wide significance threshold of $2.7*10^{-6}$ (0.05/18,390 genes). Association with a p-value of $< 10^{-3}$ were carried forward to replication in both the single variant and gene-based analysis. In order to assess the difference in the total burden of damaging variants between cases and controls PLINK was used to create the burden scores (option --score) and SPSS 20 to perform a logistic regression analysis with the PLINK created score adjusted for sex. The function and expression with other genes of identified genes was assessed using the genenetwork database (www.genenetwork.nl). Pathway analysis was performed using PANTHER [37]. PANTHER is a gene-based pathway analysis tool. We used the gene onthology biological processes database as reference.

Variance explained by common variants

In order to determine the extent of heritability of longevity that could be explained by common variants with small effects, we estimate the joint effect of common variation using the GCTA tool [38, 39] on the complete RS-I Illumina Infinium II HumanHap550 SNP chip array dataset (1,004 longevity cases and 1,131 controls). We excluded SNPs with a minor allele frequency (MAF) lower than 0.05, more than 0.05 missing values and Hardy-Weinberg Equilibrium (HWE) p-value lower than 0.05 and individuals with an identity by state (IBS) greater than 0.05 in order to adjust for familial relationships closer than and including second-cousins. For correction of case/control frequency in all analyses we used a frequency of longevity cases in the population of 3%, which corresponds to the Dutch population (Source: CBS).

RESULTS

General characteristics of discovery and replication set can be found in **Table 1**. Longevity cases were more likely to be female compared to controls. When adjusted for gender longevity cases smoked less, consumed less alcohol, were more likely to grade their health as better than others of their own age, had lower diastolic blood pressure, C-reactive protein (CRP), albumin, femoral neck bone mineral density and had higher uric acid levels. Longevity cases were less likely to die of cancer and more likely to die of dementia or other causes.

None of the variants passed the genome-wide significance threshold in the single variant analysis of the exome data (p-value $< 5.5*10^{-8}$). 36 loci (minor allele frequency > 0.01) with p-values lower than 10^{-3} were identified (**Table 2** and **Supplementary Figure 1**). Of these loci 5 were nominally significant (p-value < 0.05) in the replication set and of those 3 had odds ratios (ORs) in the same direction as the discovery. The strongest association for those variants consistently associated in the discovery and replication was seen for rs4979386

Table 1 | General characteristics of the discovery and replication populations.

	Discovery (RS-I ExSeq)				Replication (RS-I 1000G)			
	Controls	Cases	p-value[1]	p-value[2]	Controls	Cases	p-value[1]	p-value[2]
n	267	202	na	na	864	802	na	na
age at baseline	65.79 (5.67)	80.42 (6.36)	na	na	65.46 (5.37)	81.46 (6.66)	na	na
sex (% female)	37.1	78.7	$1.77*10^{-9}$	na	42.5	77.4	$4.75*10^{-7}$	na
BMI kg/m2)	26.16 (3.71)	26.96 (3.93)	0.026	0.630	26.08 (3.85)	26.73 (3.80)	0.001	0.225
anorexic (< 18.5), %	0.4	0.5	0.507	0.788	2.1	0.6	0.220	0.024
normal (18.5 - 25), %	39.4	33.8	reference	reference	36.9	30.2	reference	reference
overweight (> 25), %	60.2	65.7	0.414	0.454	61.0	69.2	0.276	0.129
Smoking (pack-years)	32.42 (23.18)	18.63 (20.48)	$3.40*10^{-5}$	0.001	34.39 (25.88)	23.05 (27.93)	$5.73*10^{-7}$	$6.33*10^{-5}$
never, %	16.0	59.0	reference	reference	17.7	59.2	reference	reference
former, %	45.8	35.9	$3.68*10^{-7}$	0.001	44.0	32.0	$8.16*10^{-7}$	$1.64*10^{-12}$
current, %	38.2	5.1	$1.57*10^{-12}$	$1.30*10^{-12}$	38.4	8.8	$1.55*10^{-10}$	$2.43*10^{-40}$
Alcohol (gram/day)	13.12 (19.83)	5.89 (9.63)	$9.10*10^{-5}$	0.029	17.32 (18.80)	10.38 (12.87)	$2.14*10^{-8}$	0.003
never, %	15.8	25.2	reference	reference	20.2	26.2	reference	reference
ever, %	84.2	74.8	0.100	0.588	79.8	73.8	0.315	0.931
Self-rated health								
better, %	44.2	65.1	0.006	$3.08*10^{-5}$	42.9	62.6	0.037	$8.17*10^{-10}$
same, %	40.3	25.1	reference	reference	39.5	32.6	reference	reference
worse, %	15.5	9.7	0.028	0.297	17.6	4.8	0.005	$6.60*10^{-8}$
SBP (mmHg)	141.97 (24.15)	143.22 (22.02)	0.567	0.980	139.74 (22.88)	144.35 (21.28)	$6.80*10^{-6}$	0.001
DBP (mmHg)	75.25 (12.94)	71.82 (10.80)	0.003	0.008	75.32 (11.80)	71.97 (12.07)	$1.97*10^{-9}$	$1.08*10^{-5}$
CRP	5.33 (10.59)	2.80 (3.10)	0.001	0.011	4.34 (9.52)	3.09 (4.66)	0.001	0.002
Albumin (g/l)	43.16 (2.65)	42.10 (2.78)	$3.24*10^{-4}$	0.015	43.10 (2.86)	41.65 (2.98)	$4.13*10^{-17}$	$2.77*10^{-14}$

	Discovery (RS-I ExSeq)				Replication (RS-I 1000G)			
	Controls	Cases	p-value[1]	p-value[2]	Controls	Cases	p-value[1]	p-value[2]
Uric Acid (umol/l)	325.71 (71.50)	332.27 (73.71)	0.403	0.046	336.96 (84.33)	319.28 (83.94)	0.003	0.489
ABI	1.11 (0.22)	1.07 (0.24)	0.140	0.970	1.11 (0.21)	1.05 (0.23)	$5.06*10^{-7}$	0.001
lumbar spine BMD	1.14 (0.22)	1.08 (0.23)	0.009	0.686	1.11 (0.20)	1.07 (0.21)	0.001	0.485
femoral neck BMD	0.89 (0.14)	0.81 (0.15)	$1.00*10^{-6}$	0.004	0.88 (0.15)	0.81 (0.14)	$2.16*10^{-17}$	$3.43*10^{-7}$
% died	100.0	73.8	na	na	100.0	79.4	na	na
Cause of Death*								
Cancer, %	41.2	9.4	$4.92*10^{-10}$	$2.99*10^{-9}$	41.8	7.8	$2.59*10^{-8}$	$1.22*10^{-38}$
CHD, %	36.0	31.5	0.364	0.380	31.3	29.4	0.770	0.907
Dementia, %	3.4	15.4	$4.98*10^{-5}$	0.003	3.1	15.7	0.002	$4.82*10^{-12}$
Other, %	19.4	43.7	$2.66*10^{-7}$	$2.73*10^{-7}$	23.8	47.1	$5.73*10^{-4}$	$4.67*10^{-18}$

*Percentage of those whom died
[1]Unadjusted p-value using one-way anova for continuous variables or chi-square test for categorical variables
[2]P-value from logistic regression adjusted for sex

Table 2 | Results of association analysis of single variants identified by exome sequencing with longevity in the discovery (p-value discovery < 10^{-3}).

								Discovery		Replication			Meta-analysis		
SNP	CHR	BP	Gene	Function	RA	EA	EAF	OR	P	rsq	OR	p-value	Direction	OR	p-value
rs4979386	9	117186592	DFNB31	INTRON	T	A	0.22	2.14	$4.35*10^{-5}$	1.00	1.21		++	1.34	**$2.26*10^{-4}$**
rs1366258	19	52941618	ZNF534	NS	C	A	0.18	2.35	$9.79*10^{-5}$	1.00	1.08	0.720	++	1.34	0.126
rs343376	1	42693597	FOXJ3	NS	G	A	0.24	1.98	$1.28*10^{-4}$	1.00	1.06	0.525	++	1.20	**0.020**
rs1050163	16	15811023	MYH11	S	T	C	0.46	1.75	$2.05*10^{-4}$	0.99	1.04	0.625	++	1.15	**0.037**
rs3934711	17	79532654	NPLOC4	INTRON	G	A	0.22	2.06	$2.12*10^{-4}$	1.00	0.89	0.192	+-	1.03	0.740
rs7782699	7	150915948	ABCF2	S	C	T	0.13	0.41	$2.36*10^{-4}$	1.00	0.91	0.379	--	0.78	**0.014**
rs1143676	2	182395345	ITGA4	NS	A	G	0.31	0.53	$3.10*10^{-4}$	0.98	1.08	0.325	-+	0.96	0.572
rs2853344	10	129907489	MKI67	NS	G	A	0.11	2.45	$3.13*10^{-4}$	1.00	0.98	0.871	+-	1.20	0.107
rs6827525	4	177137988	ASB5	S	C	T	0.44	1.77	$3.14*10^{-4}$	0.90	0.78	**0.002**	+-	0.93	0.298
rs74151283	10	97956830	BLNK	INTRON	T	A	0.02	9.28	$3.24*10^{-4}$	0.95	0.90	0.645	+-	1.24	0.276
rs1052637	2	118575215	DDX18	NS	C	G	0.36	0.56	$3.61*10^{-4}$	1.00	0.99	0.879	--	0.89	0.097
rs1370274	15	80452075	FAH	INTRON	C	T	0.38	1.76	$3.62*10^{-4}$	1.00	1.01	0.880	++	1.13	0.072
rs6072668	20	40877288	PTPRT	INTRON	C	T	0.06	0.24	$3.82*10^{-4}$	0.89	1.29	0.151	-+	1.00	0.982
rs12570211	10	11504752	USP6NL	S	T	C	0.06	3.28	$3.86*10^{-4}$	0.86	1.26	0.180	++	1.43	**0.021**
rs3748665	1	222713621	HHIPL2	NS	C	T	0.09	2.58	$4.61*10^{-4}$	0.03	0.77	0.738	+-	1.58	0.468
rs2072051	22	29755888	AP1B1	S	C	T	0.43	0.58	$5.29*10^{-4}$	0.99	0.97	0.678	--	0.88	0.058
rs17253753	23	106201527	MORC4	INTRON	G	C	0.09	3.07	$5.56*10^{-4}$	NA	NA	NA	NA	NA	NA
rs6761637	2	119739063	MARCO	NS	T	C	0.04	0.21	$5.73*10^{-4}$	1.00	0.92	0.666	--	0.72	0.054
rs28498091	17	48597114	MYCBPAP	S	C	T	0.15	2.11	$5.75*10^{-4}$	0.99	1.10	0.351	++	1.25	**0.017**
rs568922	11	118919206	HYOU1	S	C	T	0.17	0.48	$5.87*10^{-4}$	1.00	0.97	0.780	--	0.84	0.061
rs35770269	5	54468124	MIR449C	ncRNA_ exonic	A	T	0.32	0.56	$6.01*10^{-4}$	0.50	1.40	**0.004**	-+	1.04	0.665

								Discovery		Replication			Meta-analysis		
SNP	CHR	BP	Gene	Function	RA	EA	EAF	OR	P	rsq	OR	p-value	Direction	OR	p-value
rs61752597	11	6007294	OR52L1	S	A	G	0.03	0.07	$6.09*10^{-4}$	0.85	0.55	**0.027**	--	0.44	**0.001**
rs2873624	17	3563963	CTNS	NS	C	G	0.24	0.52	$6.12*10^{-4}$	0.79	0.99	0.895	--	0.90	0.212
rs1044193	9	138831567	UBAC1	S	T	C	0.13	2.16	$6.22*10^{-4}$	0.65	0.80	0.099	+-	0.99	0.906
rs74010036	15	34537617	SLC12A6	INTRON	G	A	0.10	2.37	$6.25*10^{-4}$	1.00	0.96	0.728	+-	1.16	0.149
rs2363844	17	65051180	CACNG1	INTRON	G	A	0.44	0.58	$6.33*10^{-4}$	0.97	1.11	0.181	-+	0.99	0.912
rs41282874	10	129870510	PTPRE	INTRON	A	G	0.12	2.19	$6.36*10^{-4}$	0.98	1.16	0.223	++	1.36	**0.005**
rs3809870	17	1636934	WDR81	NS	A	G	0.30	2.04	$6.78*10^{-4}$	0.93	0.98	0.791	+-	1.05	0.536
rs10886789	10	122624679	WDR11	S	G	A	0.32	1.73	$7.06*10^{-4}$	1.00	1.19	**0.028**	++	1.28	**0.001**
rs13009279	2	29222070	FAM179A	NS	G	A	0.28	0.52	$7.07*10^{-4}$	0.94	0.92	0.362	--	0.83	**0.013**
rs479323	1	21012575	KIF17	S	G	A	0.05	3.44	$7.66*10^{-4}$	0.56	1.01	0.968	++	1.34	0.118
rs3737135	9	86593077	HNRNPK	INTRON	A	T	0.25	0.55	$8.45*10^{-4}$	0.99	0.95	0.567	--	0.87	0.058
rs12460533	19	49422437	NUCB1	INTRON	A	T	0.11	0.43	$8.46*10^{-4}$	0.54	1.26	0.199	-+	0.94	0.704
rs12154775	7	111541692	DOCK4	INTRON	T	C	0.18	1.98	$8.63*10^{-4}$	0.99	1.19	0.071	++	1.30	**0.002**
rs11219508	11	124095525	OR8G2	NS	T	C	0.34	1.70	$9.86*10^{-4}$	1.00	1.00	0.962	+-	1.09	0.198
rs62171262	2	170148745	LRP2	INTRON	A	G	0.09	2.43	$9.89*10^{-4}$	0.56	0.89	0.588	+-	1.21	0.287

RA = Reference allele, EA = Effective allele, EAF = Effective allele frequency, rsq = imputation quality in replication; S = synonymous; NS = non-synonymous

(MAF = 0.22), which is located in an intron of the *DFNB31* gene (OR in the meta-analysis = 1.34; p-value = $2.26*10^{-4}$). The second synonymous SNV with a consistent effect was more rare (MAF = 0.03) and found in *OR52L1* (OR meta = 0.44; p-value = 0.001). The third is a synonymous variant in the *WDR11* gene (OR meta = 1.28; p-value = 0.001). As these synonymous variants do not translate into a different amino-acid, the association is unlikely to be biologically relevant. Candidate genes *FOXO3* and *APOE* did not reveal any significant associations (**Supplementary Table 1**). Of note is the absence of exonic variants in *FOXO3*, while 9 out of 10 exonic variants in *APOE* were rare (MAF < 0.01). Formal pathway analysis of the genes identified in the single variant analysis using PANTHER suggests the involvement of cell cycle control (mitosis) and protein transport in longevity (**Supplementary Table 2**).

Gene based SKAT analysis showed 19 genes with suggestive association to longevity (p-value $< 10^{-3}$; **Table 3** and **Supplementary Figure 2**). The gene closest to genome-wide significance (p-value $< 2.7*10^{-6}$) was *WSCD2* (p-value = $2.87*10^{-6}$). The six known damaging variants in *WSCD2* are rare and therefore not analyzed in single variant analysis. Because of the low frequency, the top associated variant could not be found in 1000 Genomes imputations. Three genes showed suggestive p-values in both the single variant analysis and the SKAT analysis, namely *OR52L1*, *MYH11* and *ZNF534* (**Supplementary Table 3**). For the latter two the leading SNPs did not replicate in that the effect of the same variant was in opposite direction (**Table 3**). In *OR52L1* one stop gain and 9 non-synonymous variants contributed to the p-value in the burden test. Formal pathway analysis of the genes identified in SKAT analysis resulted in no nominally significant associations (**Supplementary Table 4**). A genome-wide burden test of predicted damaging variants did not yield a significant difference between cases and controls (**Supplementary Table 5**).

Finally, we tested the hypothesis that common variants jointly explain a major part of the heritability using GWAS. In this analysis we ignore the effects of single genes but combine common variants to explain longevity. The total variance explained by common variants can be found in **Table 4**. Unidentified common variants explain 32.5% of the phenotypic variation (p-value = 0.01). As only 40% of the variance is explained by genetics, a large part of the heritability is explained by the joint effect of common variants. Based on a heritability of 40% up to 81.3% of the heritability is explained by yet unidentified common variants.

Table 3 | Results of gene-based sequence kernel association tests (SKAT) in longevity cases and controls (p-value < $1*10^{-3}$).

				Variant driving association						Replication		
Gene	**Chr**	**p-value**	**# SNPs**	**# damaging**	**SNPID**	**EAF**	**Score**	**OR**	**p-value**	**rsq**	**OR**	**p-value**
WSCD2	12	$2.87*10^{-6}$	32	6	12:108589681	0.01	NA	10.85	0.0405	NA	NA	NA
OR52L1	11	$5.58*10^{-5}$	14	0	rs61752597	0.02	NA	0.07	0.0006	0.85	0.55	0.0266
KIF5B	10	$2.01*10^{-4}$	101	10	NA	NA	NA	NA	NA	NA	NA	NA
CES2	16	$3.17*10^{-4}$	15	5	NA	NA	NA	NA	NA	NA	NA	NA
PSMB5	14	$3.51*10^{-4}$	11	0	rs11543947	0.07	NA	1.98	0.0200	1.00	1.32	0.0593
LDOC1L	22	$3.92*10^{-4}$	9	5	22:44893119	$3.34*10^{-4}$	NA	NA	NA	NA	NA	NA
ELOVL1	1	$4.13*10^{-4}$	11	0	NA	NA	NA	NA	NA	NA	NA	NA
MYH11,NDE1	16	$5.12*10^{-4}$	56	0	rs7193920	0.01	NA	10.19	0.0057	0.96	0.91	0.7896
TEX9	15	$5.98*10^{-4}$	71	20	rs795784	0.02	NA	4.68	0.0054	0.80	0.82	0.5203
YWHAQ	2	$6.54*10^{-4}$	23	7	rs41308184	0.10	NA	1.13	0.6203	0.95	0.95	0.6904
BLNK	10	$6.77*10^{-4}$	53	5	rs74151283	0.02	NA	9.28	0.0003	0.95	0.90	0.6448
MEGF11	15	$7.31*10^{-4}$	56	18	rs17819714	0.02	NA	0.09	0.0220	0.64	1.42	0.2761
ZNF562	19	$7.64*10^{-4}$	20	0	19:9763734	0.03	NA	1.80	0.2790	NA	NA	NA
VPREB1	22	$8.22*10^{-4}$	11	3	rs1320	0.06	0	0.41	0.0157	1.00	0.93	0.6499
HIST1H4B	6	$8.46*10^{-4}$	15	0	NA	NA	NA	NA	NA	NA	NA	NA
ZNF534	19	$8.65*10^{-4}$	56	0	rs1366258	0.17	NA	2.35	0.0001	1.00	1.08	0.7203
					rs11084161	0.03	NA	3.88	0.0010	0.97	1.03	0.8818
PCNP	3	$9.12*10^{-4}$	12	2	3:101293070	$3.37*10^{-4}$	2.5	NA	NA	NA	NA	NA
EFCAB7	1	$9.63*10^{-4}$	69	21	rs147375223	0.01	3	11.47	0.0322	0.83	1.71	0.1457
SLC51A	3	$9.96*10^{-4}$	18	0	NA	NA	NA	NA	NA	NA	NA	NA

SNPs = number of markers p-value is based on; # damaging = number of markers predicted to be damaging in at least 2 algorithms; EAF = Effective allele frequency; Score = number of algorithms predict this variant is damaging; NA = no single variant drives association; rsq = imputation quality; p-value for significance = $2.72*10^{-6}$

Table 4 | Total variance explained by common variants in 1004 cases and 1131 controls.

V(G)/Vp_L	variance	se	p-value	h2 (85+)	h2 explained
	0.325	0.144	0.01	0.40	0.813

V(G)/Vp_L: ratio of genetic variance to phenotypic variance

DISCUSSION

In this study investigating the association of rare variants with longevity we find suggestive evidence for *WSCD2* in a gene-based test which shows a p-value that is borderline significant. The pathways emerging from the single exonic variant analysis involve cell cycle control and protein transport. Moreover, we show that 81.3% of the heritability of longevity can be explained by common variants captured in genome-wide association arrays which include common variants in coding and non-coding regions.

Comparison of environmental factors between longevity cases and controls revealed significant differences in smoking habits and alcohol consumption. Additionally, longevity cases had lower diastolic blood pressure, CRP, albumin and femoral neck bone mineral density and they were more likely to rate their health as better than others their own age. Significant differences between all these risk factors have been described in our population previously in a less extreme longevity phenotype (80+) [28]. The present analysis shows that findings can be translated from 80+ to 90+ but also underscores the power of our study design to detect strong associations to clinical and epidemiological risk factors. In the present analysis of the Rotterdam Study, we observe significant differences in the causes of death between longevity cases and controls, with cases more likely to die of dementia and controls more likely to die of cancer. This is consistent with statistics showing that cancer is the number one cause of death in younger individuals [40], while dementia is very prevalent at old age [41].

In the single variant analysis of the exome variants we identified no rare variants for which the association with longevity reached genome-wide significance. Most of the suggestive variants were common except of the synonymous variant in *ORL521*. Rs4979386 in *DFNB31* showed the lowest p-value ($2.26*10^{-4}$). This is a common variant with a relatively small OR (1.34) that, if genotyped or imputed, should have surfaced in GWAS. *DFNB31* has been implicated in a recessive form of sensorineural hearing loss [42]. Gene Ontology suggests that this gene is involved in mineralocorticoid metabolic processes. The fact that this common variant is not found in GWAS makes it less likely to be a true positive finding. Indeed the p-value for this variant was 0.28 in a GWAS of the CHARGE consortium (personal communication). The findings on *ORL521* are of interest although the single variant analysis shows no strong evidence for association based on the p-value (p-value = 0.001). However, this gene was also suggestively associated in the gene-based analysis (p-value = $5.58*10^{-5}$),

in which 14 variants were analyzed simultaneously of which 9 were non-synonymous and 1 was a stop gain.

In the gene-based analysis using SKAT, one gene approached genome-wide significance, *WSCD2*. The leading SNVs could not be replicated in 1000 Genomes imputations as the variants are not found in the 1000 Genomes. All variants included in the gene-based analysis were rare (median = $3.55*10^{-4}$; interquartile range = $3.36*10^{-4}$ to $4.32*10^{-4}$), making it difficult to replicate this finding. Even in one of our own other studies (the Erasmus Rucphen Family Study) the variants were not found. Coverage of the region was sufficient to call variants with a median of 25.8 (interquartile range = 17.5 to 32.7). *WSCD2* has previously been associated to bipolar disorder [43]. Gene Ontology suggests a function in long-term memory for this gene linking this gene to *APOE*, which is also involved in memory. Although we reported earlier the relation of *APOE* with longevity (Deelen *et al*, *submitted*), we did not detect *APOE* in this analysis as a top finding.

There are several reasons for not finding any significant associations for rare variants in the discovery. First, our sample size is small with only 202 longevity cases and 267 controls. This implies that we have an 80% power to detect variants with a frequency of 1-3% if the effect in terms of the OR is 14 or higher. For more moderate effect sizes we are underpowered. The low power to detect moderate to small effect sizes may explain why we did not detect common variants in *APOE* or *FOXO3*. Second, our sample is population-based and not selected for long-lived families in which longevity is enriched and where rare variants are expected to have a larger role in determining longevity [10, 23].

Our study of common variants was better powered as the total dataset was used and variants are easier to detect in a genome-wide analysis investigating the joint effect of all variants. When estimating the total variance explained by common variants jointly we found that a large part of the heritability of longevity can be explained by common variants (81.3%). This study suggests that the genetics of longevity is for a large part the result of a large number of variants with small effects. Yet GWAS studies have failed to uncover them. The main reason for not finding associations with longevity in GWAS may be that the effects are small resulting in low power to detect these variants. Enlarging the sample size may help in locating common variants associated with longevity. Similar to other outcomes with a high heritability, for example height and depression [25, 44], we found for longevity that a large part of the heritability is explained by common variants which on themselves have very small effects.

In summary, we attempted to identify rare exomic variants associated with longevity. We found suggestive evidence for the association of multiple rare variants in *OR52L1* and *WSCD2* with longevity, but these findings remain to be replicated. A very compelling finding of our study is that the heritability of longevity is likely determined for a large part (81.3%) by a large number of common variants with very small effects. The challenge for the near future will be to identify these variants.

REFERENCES

1. Oeppen, J. and J.W. Vaupel, Demography. Broken limits to life expectancy. Science, 2002. 296(5570): p. 1029-31.
2. Arias, E., United States life tables, 2007. Natl Vital Stat Rep, 2011. 59(9): p. 1-60.
3. Herskind, A.M., et al., The heritability of human longevity: a population-based study of 2872 Danish twin pairs born 1870-1900. Hum Genet, 1996. 97(3): p. 319-23.
4. McGue, M., et al., Longevity is moderately heritable in a sample of Danish twins born 1870-1880. J Gerontol, 1993. 48(6): p. B237-44.
5. Kerber, R.A., et al., Familial excess longevity in Utah genealogies. J Gerontol A Biol Sci Med Sci, 2001. 56(3): p. B130-9.
6. Mitchell, B.D., et al., Heritability of life span in the Old Order Amish. Am J Med Genet, 2001. 102(4): p. 346-52.
7. Murabito, J.M., R. Yuan, and K.L. Lunetta, The search for longevity and healthy aging genes: insights from epidemiological studies and samples of long-lived individuals. J Gerontol A Biol Sci Med Sci, 2012. 67(5): p. 470-9.
8. Schachter, F., et al., Genetic associations with human longevity at the APOE and ACE loci. Nat Genet, 1994. 6(1): p. 29-32.
9. Bathum, L., et al., Apolipoprotein e genotypes: relationship to cognitive functioning, cognitive decline, and survival in nonagenarians. J Am Geriatr Soc, 2006. 54(4): p. 654-8.
10. Christensen, K., T.E. Johnson, and J.W. Vaupel, The quest for genetic determinants of human longevity: challenges and insights. Nat Rev Genet, 2006. 7(6): p. 436-48.
11. Deelen, J., et al., Genome-wide association study identifies a single major locus contributing to survival into old age; the APOE locus revisited. Aging Cell, 2011. 10(4): p. 686-98.
12. Gerdes, L.U., et al., Estimation of apolipoprotein E genotype-specific relative mortality risks from the distribution of genotypes in centenarians and middle-aged men: apolipoprotein E gene is a "frailty gene," not a "longevity gene". Genet Epidemiol, 2000. 19(3): p. 202-10.
13. Nebel, A., et al., A genome-wide association study confirms APOE as the major gene influencing survival in long-lived individuals. Mech Ageing Dev, 2011. 132(6-7): p. 324-30.
14. Sebastiani, P., et al., Genetic signatures of exceptional longevity in humans. PLoS One, 2012. 7(1): p. e29848.
15. Anselmi, C.V., et al., Association of the FOXO3A locus with extreme longevity in a southern Italian centenarian study. Rejuvenation Res, 2009. 12(2): p. 95-104.
16. Flachsbart, F., et al., Association of FOXO3A variation with human longevity confirmed in German centenarians. Proc Natl Acad Sci U S A, 2009. 106(8): p. 2700-5.
17. Njajou, O.T., et al., Association between telomere length, specific causes of death, and years of healthy life in health, aging, and body composition, a population-based cohort study. J Gerontol A Biol Sci Med Sci, 2009. 64(8): p. 860-4.
18. Soerensen, M., et al., Replication of an association of variation in the FOXO3A gene with human longevity using both case-control and longitudinal data. Aging Cell, 2010. 9(6): p. 1010-7.
19. Willcox, B.J., et al., FOXO3A genotype is strongly associated with human longevity. Proc Natl Acad Sci U S A, 2008. 105(37): p. 13987-92.
20. Malovini, A., et al., Association study on long-living individuals from Southern Italy identifies rs10491334 in the CAMKIV gene that regulates survival proteins. Rejuvenation Res, 2011. 14(3): p. 283-91.
21. Newman, A.B., et al., A meta-analysis of four genome-wide association studies of survival to age 90 years or older: the Cohorts for Heart and Aging Research in Genomic Epidemiology Consortium. J Gerontol A Biol Sci Med Sci, 2010. 65(5): p. 478-87.
22. Walter, S., et al., A genome-wide association study of aging. Neurobiol Aging, 2011. 32(11): p. 2109 e15-28.

23. Westendorp, R.G., et al., Nonagenarian siblings and their offspring display lower risk of mortality and morbidity than sporadic nonagenarians: The Leiden Longevity Study. J Am Geriatr Soc, 2009. 57(9): p. 1634-7.
24. Cirulli, E.T. and D.B. Goldstein, Uncovering the roles of rare variants in common disease through whole-genome sequencing. Nat Rev Genet, 2010. 11(6): p. 415-25.
25. Yang, J., et al., Common SNPs explain a large proportion of the heritability for human height. Nat Genet, 2010. 42(7): p. 565-9.
26. Hofman, A., et al., Determinants of disease and disability in the elderly: the Rotterdam Elderly Study. Eur J Epidemiol, 1991. 7(4): p. 403-22.
27. Hofman, A., et al., The Rotterdam Study: 2012 objectives and design update. Eur J Epidemiol, 2011. 26(8): p. 657-86.
28. Newson, R.S., et al., Predicting survival and morbidity-free survival to very old age. Age (Dordr), 2010. 32(4): p. 521-34.
29. Miller, S.A., D.D. Dykes, and H.F. Polesky, A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res, 1988. 16(3): p. 1215.
30. Li, H. and R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics, 2009. 25(14): p. 1754-60.
31. Li, H., et al., The Sequence Alignment/Map format and SAMtools. Bioinformatics, 2009. 25(16): p. 2078-9.
32. McKenna, A., et al., The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res, 2010. 20(9): p. 1297-303.
33. Adzhubei, I.A., et al., A method and server for predicting damaging missense mutations. Nat Methods, 2010. 7(4): p. 248-9.
34. Aulchenko, Y.S., M.V. Struchalin, and C.M. van Duijn, ProbABEL package for genome-wide association analysis of imputed data. BMC Bioinformatics, 2010. 11: p. 134.
35. Purcell, S., et al., PLINK: a toolset for whole-genome association and population-based linkage analysis. Am J Hum Genet, 2007. 81.
36. Ionita-Laza, I., et al., Sequence Kernel Association Tests for the Combined Effect of Rare and Common Variants. Am J Hum Genet, 2013.
37. Mi, H., A. Muruganujan, and P.D. Thomas, PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. Nucleic Acids Res, 2013. 41(Database issue): p. D377-86.
38. Lee, S.H., et al., Estimating missing heritability for disease from genome-wide association studies. Am J Hum Genet, 2011. 88(3): p. 294-305.
39. Yang, J., et al., GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet, 2011. 88(1): p. 76-82.
40. Siegel, R., D. Naishadham, and A. Jemal, Cancer statistics, 2013. CA Cancer J Clin, 2013. 63(1): p. 11-30.
41. Aronson, M.K., et al., Dementia. Age-dependent incidence, prevalence, and mortality in the old old. Arch Intern Med, 1991. 151(5): p. 989-92.
42. van Wijk, E., et al., The DFNB31 gene product whirlin connects to the Usher protein network in the cochlea and retina by direct association with USH2A and VLGR1. Hum Mol Genet, 2006. 15(5): p. 751-65.
43. Greenwood, T.A., et al., Genome-wide association study of temperament in bipolar disorder reveals significant associations with three novel Loci. Biol Psychiatry, 2012. 72(4): p. 303-10.
44. Demirkan, A., et al., Genetic risk profiles for depression and anxiety in adult and elderly cohorts. Mol Psychiatry, 2011. 16(7): p. 773-83.

Chapter 3

The role of Heat Shock Proteins in morbidiy and mortality

Chapter 3.1

Association of HSP70 and its co-chaperones with Alzheimer's Disease

Linda Broer, M. Arfan Ikram, Maaike Schuur, Anita L. DeStefano, Joshua C. Bis, Fan Liu, Fernando Rivadeneira, Andre G. Uitterlinden, Alexa S. Beiser, William T. Longstreth, Albert Hofman, Yurii Aulchenko, Sudha Seshadri, Annette L. Fitzpatrick, Ben A. Oostra, Monique M.B. Breteler, Cornelia M. van Duijn

Published: Journal of Alzheimer's Disease; 2011

ABSTRACT

The heat shock protein (HSP) 70 family has been implicated in the pathology of Alzheimer's disease (AD). In this study, we examined common genetic variations in the 80 genes encoding HSP70 and its co-chaperones.

We conducted a study in a series of 462 patients and 5238 unaffected participants derived from the Rotterdam Study, a population-based study including 7983 persons aged 55 years and older. We genotyped a total of 12,053 Single Nucleotide Polymorphisms (SNPs) using the HumanHap550K Genotyping BeadChip from Illumina. Replication was performed in two independent cohort studies, the Framingham Heart study (FHS; N = 806) and Cardiovascular Health Study (CHS; N = 2150).

When adjusting for multiple testing, we found a small but consistent, though not significant effect of rs12118313 located 32kb from *PFDN2*, with an OR of 1.19 (p-value from meta-analysis = 0.003). However this SNP was in the intron of another gene, suggesting it is unlikely this SNP reflects the effect of *PFDN2*. In a formal pathway analysis we found nominally significant evidence for an association of BAG, DNAJA and prefoldin with AD.

These findings corroborate with those of a study of 2032 AD patients and 5328 controls, in which several members of the prefoldin family showed evidence for association to AD. Our study did not reveal evidence for a genetic variant if the HSP70 family with a major effect on AD. However, our findings of the single SNP analysis and pathway analysis suggest that multiple genetic variants in prefoldin are associated with AD.

INTRODUCTION

Heat shock proteins (HSPs) have been implicated in the pathophysiology of Alzheimer's disease (AD) and its major pathological characteristics, neurofibrillary tangles (NFTs) and β-amyloid (Aβ) [1-3]. AD is an example of a protein-folding disorder, with Aβ plaques as main misfolded protein [1, 3]. HSPs are the major chaperones mediating proper (re)folding of proteins [4].

There is increasing evidence for the involvement of the heat shock protein 70 (HSP70) family in AD. It has been shown that Aβ plaques co-localize with HSP70 [1, 5, 6]. Further, HSP70 participates in the neuroprotective response to Aβ plaques [7]. HSP70 has also been found to promote tau solubility and tau binding to microtubules, thereby suppressing formation of NFTs [8].

HSP70 has numerous co-chaperones that support its function. The largest group of co-chaperones comprises the HSP40 family, also called J domain protein family. HSP40 can be subdivided into 3 subclasses (A, B and C) and each member has both common and unique functions in the cell [9]. A role of HSP40 proteins in neurodegenerative disorders, like AD, has been suggested in several studies [10]. Other important groups of co-chaperones include the BCL2-associated anthanogene (BAG) proteins, Hip, Hop, CHIP and prefoldin [11, 12]. At autopsy, the prefoldin protein complex was found to be up-regulated in brains of patients with AD [13].

We have tested 79 genes encoding all members of these HSP families and assessed their association with AD. First we have conducted a single SNP analysis. Second we performed a formal pathway analysis in which we test for evidence of multiple SNPs in various genes in a single pathway which contribute jointly to the evidence of association. To our knowledge, this is the first large-scale genomic study of the association of HSP70 and its co-chaperones with AD to be performed.

METHODS

Discovery Study

Our discovery cohort was the Rotterdam Study (RS1). RS1 is a population-based cohort study that investigates the occurrence and determinants of diseases in the elderly [14]. Baseline examinations, including a detailed questionnaire, physical examination and blood collection, were conducted between 1990 and 1993. The Medical Ethics Committee at Erasmus Medical Center approved the study protocol.

Dementia is one of the focus disorders of the Rotterdam Study. The diagnosis of dementia was made following a stringent three-step protocol [15]. Briefly, all subjects were screened at follow-up visits (1997-1999, 1999-2000 and 2000-2003) using two tests of cognition: the Mini-Mental State Examination (MMSE) [16] and Geriatric Mental State Schedule (GMS) [17]. Participants that were screen-positive (MMSE score < 26 or GMS organic level > 0)

3

underwent the Cambridge examination for mental disorders of the elderly (CAMDEX). When additional neuropsychological testing was required for diagnosis, a neuropsychologist examined subjects who were suspected of having dementia. In addition, the total cohort was continuously monitored for incident dementia through a computerized link between the study databases and digitalized medical records from general practitioners and the Regional Institute for Outpatient Mental Health Care until January 1, 2005. The diagnosis of dementia and subtypes of dementia was made in accordance with internationally accepted criteria for dementia (DSM-III-R) [18] and Alzheimer disease (NINDS-ADRDA) [19] by a panel consisting of a neurologist, a neuropsychologist and a research physician. We used only incident patients in the discovery study.

Genomic DNA was extracted from whole blood samples using standard methods [20]. Genome-wide SNP genotyping was performed using Infinium II assay on the HumanHap550 Genotyping BeadChips (Illumina Inc, San Diego, USA). Approximately 2 million SNPs were imputed using release 22 HapMap CEU population as reference. The imputations were performed using MACH software[10] (http://www.sph.umich.edu/csg/abecasis/MACH/). The quality of imputations were checked by contrasting imputed and actual genotypes at 78,844 SNPs not present on Illumina 550K for 437 individuals for whom these SNPs were directly typed using Affymetrix 500K. Using the "best guess" genotype for imputed SNPs the concordance rate was 99% for SNPs with R^2 (ratio of the variance of imputed genotypes to the binomial variance) quality measure greater than 0.9; concordance was still over 90% (94%) when R^2 was between 0.5 and 0.9.

The genome wide association study has been analyzed and is submitted elsewhere [21]. For the study of HSP proteins presented here, a total of 12,053 SNPs in 79 genes were initially selected for the association test on the basis of the following criteria: (1) position within the genes of interest with a margin of 100kb on each side of the genes according to NCBI build 36.3, (2) P-value for Hardy-Weinberg equilibrium test ≥ 0.0001, and (3) call rate ≥ 95%. For further analysis and selection of SNPs selected for evaluation of the consistency of findings a R^2 higher than 0.8 and a minor allele frequency (MAF) higher than 0.05 was used as selection criterion for a SNP.

Single SNP analyses

First we analyzed the individual SNPs using ProbABEL [22]. We used allele-based logistic regression to test the association between a single SNP and AD. Odds Ratios (ORs) for each SNP were derived adjusting for age and sex. To calculate empirical significance for SNPs, permutations were performed per region of interest (ROI) [23, 24]. Briefly, the empirical distribution of the region-wide maximum of the test statistic under the null was obtained in 10,000 replications. To estimate empirical significance each observed test statistic was compared with null statistics obtained empirically and the p-value was estimated as the proportion of replicas generating the test statistics greater than or equal to the observed

statistic. The permutation analysis keeps the original genotypes for each individual, but randomly allocates the phenotypes for each consecutive permutation. Therefore the linkage disequilibrium (LD) structure of genes is not broken up. We did not perform any additional correction for the number of ROIs after permutations as we used the permutation analysis to significantly reduce the number of SNPs to be sent for validation of the effect of the SNP. For each suggestively associated ROI ($p\text{-value}_{permuted} < 0.10$) we next selected the 'independently associated' SNPs in a backward stepwise logistic regression until only nominally significant SNPs remained in the model. These independently associated SNPs were then followed up in two independent cohorts. As effect estimates are expected to be small, a positive replication based on the p-value and odds ratio (OR) observed in individual replication studies, might be difficult to achieve due to power issues. Therefore we evaluated whether a SNP was showing an OR in the same direction in all three cohorts, i.e., whether findings were consistent over the various cohorts. Similar to genome wide association studies, we performed a joint meta-analysis of the discovery and the two independent cohorts and tested whether the joint p-value was significant using a Bonferoni correction for the number of SNPs validated in independent cohorts.

We used two studies that are part of the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium to evaluate whether findings were consistent. Data on AD were available in the Cardiovascular Health Study (CHS) and the Framingham Heart Study (FHS) [25]. Again only incident patients were studied. The CHS is a population-based cohort study of risk factors for CHD and stroke in adults ≥ 65 years conducted across four field centers. The original predominantly European ancestry cohort of 5201 persons was recruited in 1989-1990 from random samples of the Medicare eligibility lists and an additional 687 African-Americans were enrolled subsequently for a total sample of 5888 [26]. For compatibility with the discovery samples, however, these analyses were limited to the participants of self-described European ancestry. Blood was drawn for genotyping at the baseline examination. Persons were examined annually from enrollment to 1999, and the examination included a 30 minute screening cognitive battery [27]. In 1992-94 and again, in 1997-99, participants were invited to undergo brain MRI and detailed cognitive and neurological assessment as part of the CHS Cognition Study [27]. Persons with prevalent dementia were identified, and all others were followed until 1999 for the development of incident dementia and AD. Since then, CHS participants at the Maryland and Pennsylvania centers have remained under ongoing dementia surveillance [28]. The sample for this study included all incident cases between 1992 and December 2006.

FHS is a single-site study that comprises three generations of participants including the Original cohort followed since 1948 (n = 5209), [29] and their offspring and spouses of the offspring (n = 5124) followed since 1971 [30]. The original cohort has been evaluated biennially since 1948, was screened for prevalent dementia in 1974-76 and has been under surveillance for incident dementia since then [31, 32]. The offspring were examined

once every 4 years, were screened for prevalent dementia in 1979-1983 and have since been under surveillance for incident dementia [33, 34]. In order to be consistent with the sampling frame for the CHS sample, we excluded FHS subjects with a baseline age < 65 years at the time of DNA draw which was in the 1990s. To minimize survival biases persons who developed dementia prior to the date of DNA draw were treated as prevalent cases and subsequent events occurring prior to December 2006 were included in the incident analyses.

Pathway analysis

As we targeted distinct families within a larger pathway we next conducted a pathway analysis, using the SNP Ratio Test (SRT) tool, which allows for the testing of a specific, 'user defined' hypothesis [35]. Users can define their own pathways to test. We tested each HSP family, namely HSP40, its three subclasses, BAG, CHIP-like, prefoldin and HSP70 separately. We created 1000 simulation studies to test whether our observed number of 'significant' SNPs superseded what could be expected by chance. According to the principle of permutation analysis, all original genotypes remain and the phenotypes are randomly allocated with each simulation before rerunning the association analysis. This deals with any potential bias due to LD, as the same LD patterns are present in the original analysis and in each simulation analysis. All SNPs previously selected for the individual SNP analysis (SNPs within a 100kb window for the flanking region and a cut-off for the R^2 0f 0.8 and MAF of 0.05) in RS1 were entered in SRT. SRT deals with differences in gene size and number of SNPs per gene, by completely discarding the concept 'gene'. SRT simply counts the number of significant SNPs in a defined pathway and divides this by the total number of SNPs in that pathway to create a ratio. Significance of the original SNPs is determined by one of three cut-off p-values, namely 0.05, 0.01 or 0.001. In each simulation analysis SRT deals with LD and sizes of genes similar to the original analysis. The p-value is determined by dividing the number of simulations that have a more extreme ratio by the total number of simulations performed. The analysis was adjusted for age and sex. The findings of the SRT analyses were compared to a genome-wide pathway analyses published earlier by Hong *et al* [36].

RESULTS

General characteristics of the discovery population as well as the two samples used to check the ORs for consistency are shown in **Table 1**. As expected in all cohorts the cases are significantly older than controls and there are significantly more women and APOE*4 carriers among them. A summary of all SNPs tested in HSP70 and its co-chaperones in RS1 can be found in the **Supplementary Table 1**. Of note is that 3 HSP70 genes located on chromosome 6 have overlapping regions and are taken as 1 Region of Interest (ROI) with 148 SNPs. Also of note is that the HSPs are currently undergoing a change in nomenclature

[37]. In the table are the official gene names and the gene names according to the new nomenclature. Here we will use the names of the new nomenclature.

Table 1 | General Characteristics for the discovery and cohorts used to test the consistency of effects.

	RS1		CHS		FHS	
	Cases	**Controls**	**Cases**	**Controls**	**Cases**	**Controls**
N	462	5238	366	1784	76	730
Age (mean+se)	76 (7)	68 (9)	80 (6)	75 (5)	87 (6)	76 (7)
Female (%)	74	57	53	62	81	57
APOE*4 carrier (%)	45	27	38	24	38	20

RS1: Rotterdam Study; CHS: Cardiovascular Health Study; FHS: Framingham Heart Study

Figure 1 shows the observed versus the expected chi^2 of the 12,053 SNPs tested in RS1. The plot shows some excess of low p-values for the tested SNPs. **Supplementary Figure 1a-1c** shows an overview of the p-values in all genes tested in RS1. After adjusting for multiple testing by permutation analysis, 8 genes showed suggestive evidence (p-permuted < 0.10) for association to AD in RS1 (**Table 2**). These genes encode four members of the HSP40 family (*DNAJA4*, *DNAJC14*, *DNAJC17* and *DNAJC28*), one Prefoldin subunit (*PFDN2*), one BAG gene (*BAG2*) and two HSP70 genes (*HSPA5* and *HSPA12B*). The following SNPs were selected for validation of their effects in two independent studies: rs8027394 in *DNAJA4*, rs1463592 in *DNAJC14*, rs7915 in *DNAJC17*, rs7280365 in *DNAJC28*, rs12118313 in *PFDN2*, rs7760349 and rs13213618 in *BAG2*, rs13294021 in *HSPA5* and rs3899452 in *HSPA12B*. Two SNPs in *BAG2* were selected as both showed independent evidence for association to AD in RS1 when including all SNPs in the ROI in the model. In the validation phase 2 SNPs were consistent in direction of ORs across populations (**Table 2**), with a nominal p-value of the meta-analysis at or close to a Bonferoni adjusted p-value for significance (0.05/9 = 0.0056). One SNP (rs1463592) did not surpass the adjustment for multiple testing (*DNAJC14*; p-value = 0.007). **Figure 2a** shows the Forrest plot of the meta-analysis of *PFDN2* rs12118313. The summary OR for the SNP rs12118313 is 1.19 (95% CI 1.06-1.33; p-value = 0.003) for the C allele. Though this SNP is not replicated when considering the p-value corrected for testing multiple SNPs < 0.05, findings are consistent across cohorts in terms of effect size and direction. In **Figure 2b** a regional plot of the SNPs tested in *PFDN2* from the original RS1 cohort can be found. The figure shows that the p-value was most significant in RS1. Although there are multiple marginally associated SNPs in and flanking the *PFDN2* gene (see **Supplementary Table 2**), rs12118313 in located in the intron of another gene (*ARHGAP30*).

Table 2 | Consistency of the effects in the CHS and FHS cohorts of the SNPs significantly associated with AD in the Rotterdam study.

					RS1				CHS				FHS			
Gene	SNP	Chr	Position	Coded allele	Freq controls	OR	95% CI	P-value	Freq controls	OR	95% CI	P-value	Freq controls	OR	95% CI	P-value
HSPA12B	rs3899452	20	3775265	C	0.13	1.49	1.22-1.85	0.0002	0.14	0.90	0.70-1.16	0.413	0.14	0.63	0.30-1.31	0.216
HSPA5	rs13294021	9	127108209	G	0.51	1.21	1.05-1.39	0.008	0.53	0.93	0.80-1.07	0.297	0.52	0.84	0.62-1.15	0.278
PFDN2	rs12118313	1	159304603	C	0.22	1.37	1.16-1.61	0.0001	0.22	1.01	0.84-1.22	0.891	0.21	1.06	0.70-1.61	0.770
BAG2_1	rs7760349	6	57220087	A	0.45	1.28	1.11-1.47	0.0007	0.45	0.97	0.83-1.12	0.658	0.44	1.09	0.78-1.53	0.602
BAG2_2	rs13213618	6	57229922	G	0.17	1.37	1.15-1.64	0.0005	0.16	1.07	0.87-1.33	0.520	0.18	0.76	0.48-1.22	0.261
DNAJC14	rs1463592	12	54543559	A	0.22	1.28	1.09-1.52	0.003	0.18	1.06	0.87-1.29	0.556	0.22	1.27	0.88-1.84	0.208
DNAJC17	rs7915	15	38893777	C	0.37	1.24	1.08-1.43	0.003	0.35	0.83	0.71-0.96	0.013	0.37	0.81	0.56-1.15	0.230
DNAJA4	rs8027394	15	76404968	C	0.70	1.33	1.14-1.57	0.0005	0.67	0.99	0.85-1.16	0.892	0.66	1.02	0.73-1.41	0.921
DNAJC28	rs7280365	21	33874453	G	0.62	1.28	1.10-1.49	0.001	0.64	1.01	0.86-1.18	0.953	0.61	0.95	0.67-1.37	0.800

RS1: Rotterdam study; CHS: Cardiovascular Health Study; FHS: Framingham Heart Study; Chr: chromosome; Freq: frequency; OR: Odds Ratio; 95% CI: 95% Confidence Interval; Bold: The results for this gene are consistent across populations, with all ORs in the same direction.

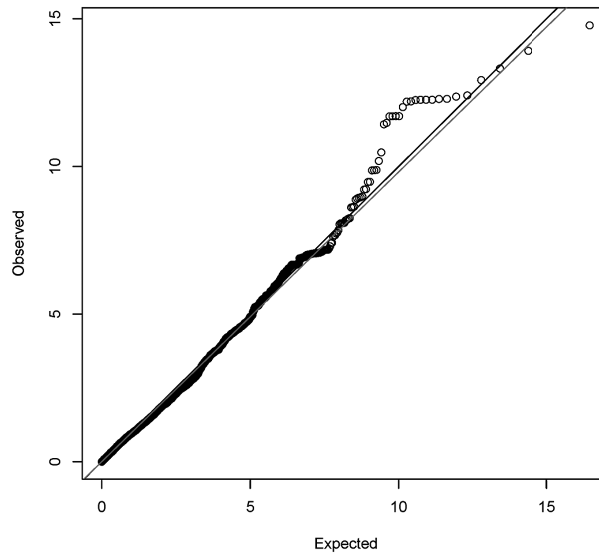
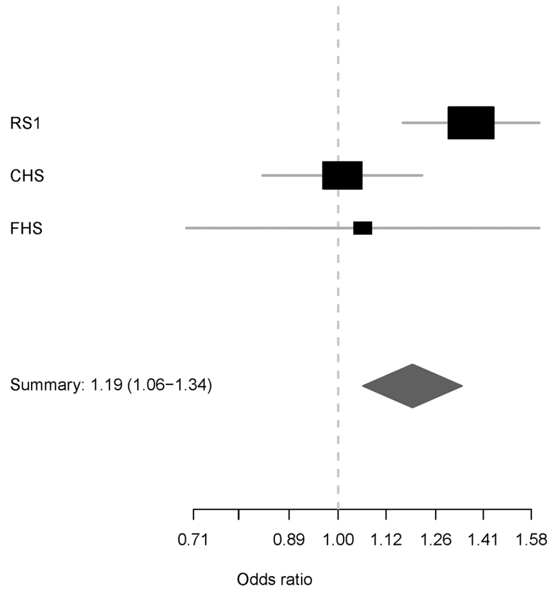
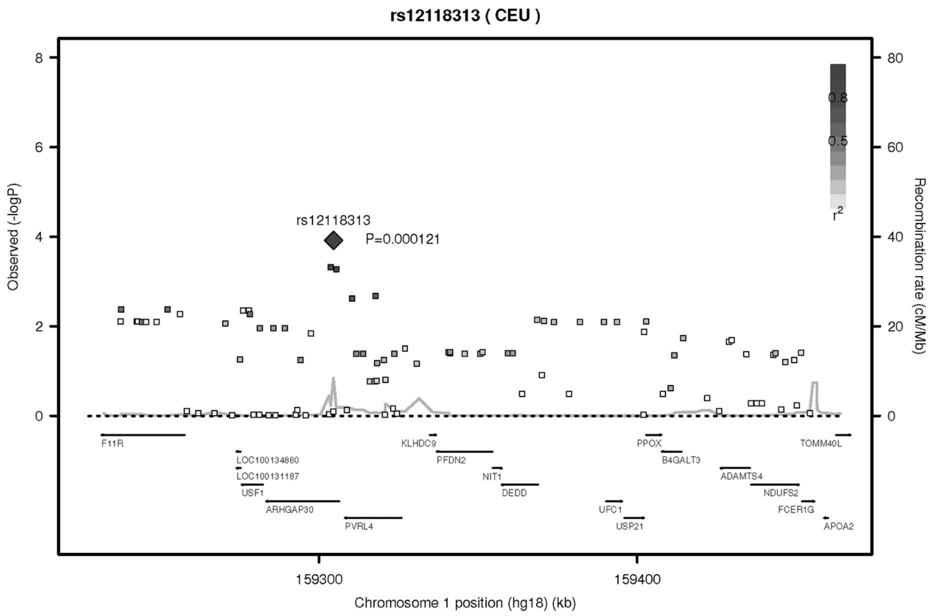


Figure 1 | Observed versus Expected p-value plot for the 12053 SNPs tested in the Rotterdam study (chi2 are given at the x-axis and y-axis).

In order to evaluate whether there is a joint effect of multiple SNPs in the *PFDN2* region and elsewhere we next conducted a formal pathway analysis. **Table 3** presents the results for the SRT pathway analysis on the different HSP families in RS1. We chose three cut-off p-values for selection of the SNPs in the discovery set (0.05, 0.01, 0.001) as was suggested in the original paper on the SRT approach. For the prefoldin HSP family we found an effect at the higher p-value cut-off (0.01), this effect is explained by 29 SNPs including 3 SNPs in *ARGHAP30*. When excluding the SNPs in *ARGHAP30* from the analysis, the p-value for the pathway was 0.03, suggesting that there are many prefoldin SNPs with smaller effects implicated. We further find nominally significant evidence for a role of the DNAJA and BAG families of HSPs when applying the smallest p-value cut-off (0.001). For the replication we used a data mining approach of a genome-wide pathway analysis of 2032 patients with AD and 5328 controls [36]. The major pathway identified in this paper was related to intracellular transmembrane protein transport. We used supplementary data to validate our SRT analyses. The PFDN gene family emerges with different genes (*PFDN1*, *PFDN2*, *PFDN6*) with p-values varying from $3.85*10^{-2}$ (*PFDN6*) to $6.80*10^{-8}$ (*PFND1*). A total of 42 SNPs in the gene family were associated with marginal p-values < 0.05. There was little support for the BAG family (best p-value 0.02 for *BAG2*) nor for the DNAJA family (*DNAJA4*: p-value = $6.11*10^{-6}$).



2a | Forrest plot



2b | Regional plot in RS-1

Figure 2 | The meta-analysis of the PFDN2 rs12118313 to AD in the three population-based studies.

Table 3 | SNP Ratio Test (SRT) pathway analysis results.

HSP family	Total # SNPs	P = 0.05	P = 0.01	P = 0.001
HSP40	6352	0.72 (279)	0.68 (50)	0.31 (7)
DNAJA	542	0.62 (13)	0.19 (9)	(7)
DNAJB	1390	0.49 (70)	0.96 (1)	1.00 (0)
DNAJC	4420	0.67 (196)	0.56 (40)	1.00 (0)
BAG	517	0.68 (16)	0.20 (10)	(9)
CHIPlike	292	1.00 (0)	1.00 (0)	1.00 (0)
Prefoldin	912	0.10 (93)	(29)	0.11 (3)
HSP70	1880	0.78 (94)	0.34 (29)	0.37 (2)

Pathway analysis results of the different HSP families at three different cut-offs for significance (P = 0.05, P = 0.01 and P = 0.001). The p-values given in the table depict the chance that the number of SNPs found to be significant at the given p-value cut-off is more than expected by chance. In bold are those p-values passing a nominal threshold of significance. In brackets are the number of SNPs passing the threshold of significance depicted. DNAJA, DNAJB and DNAJC are sub-families of HSP40.

DISCUSSION

In our study, we did not find significant evidence for a role of the individual SNPs in or directly flanking the HSP genes. However, in the pathway analysis we found nominally significant evidence for a joint effect of multiple SNPs in the DNAJA, BAG and prefoldin families. In this paper we did not adjust for multiple testing of the pathway analysis. Multiple testing adjustments on pathway analysis are difficult because many, if not all, pathways are not independent of one another. HSP70 cannot function properly without its co-chaperones and the different groups of co-chaperones compete for association to HSP70. Most other pathway tools therefore often clump all HSP families together as one huge pathway. As we wanted to test specifically which HSP family is the best candidate for association with AD we decided to divide them in separate, but not independent, pathways. It is therefore arguable whether or not to adjust for multiple testing in the pathway analysis. We have analyzed 8 pathways for association with AD. Although the HSP pathway was not the major pathway implicated in AD in an analysis of 2032 patients with AD and 5328 controls [36], there also is marginal evidence for the prefoldin family in this study. Specifically, *PFDN1*, *PFDN2* (the gene we detected), *PFDN6* surface also in this GWAs supporting our findings at the pathway level.

PFDN2 showed evidence for association with AD both in the single SNP analysis as in the pathway analysis as a part of the prefoldin HSP family. As the single SNP analysis showed an association with a SNP in another gene, the findings of the pathway analyses were most convincing, although the 3 SNPs located in *ARGHAP30* have contributed to the p-value, these do not explain the association fully. When excluding the 3 SNPs the pathway was still significant. From a biological perspective, there is no evidence that these 3 SNPs are

involved in the expression of *PFDN2*, however we cannot exclude this as this is based on eSNP analyses in lymphocytes. *PFDN2* is located in a chromosomal region known to be associated with AD in linkage studies [38] and it is a subunit of the prefoldin complex. Prefoldin is an intermediary factor between HSP70 and the TCP-1 ring complex (TRiC). The TRiC complex is involved in about 10% of the protein folding in the cytosol. Prefoldin is necessary for the transport of unfolded proteins to this complex [11]. Prefoldin has also been shown to induce *in vitro* formation of soluble Aβ oligomers similar in size to those found in AD brains [39]. It is speculated that the function of Prefoldin is to prevent aggregation, causing more of the highly toxic soluble Aβ oligomers to be present in the brain [39]. Moreover, the PFDN2 protein has been found to be upregulated in the brains of patients with AD [13]. This makes the *PFDN2* gene a plausible candidate gene for AD.

There is some evidence in our study and the study of Hong *et al* [36] for a role of DNAJA. The HSP family DNAJA is a subfamily of the larger HSP40 family. HSP40 has been shown to reduce aggregate formation in other neurodegenerative diseases [40] and recruits HSP70 to aggregates [41]. It is required for HSP70 to bind these aggregated proteins and process them [42]. Genetic variants in HSP40 genes may modify their ability to recruit and guide HSP70 to aggregates, at least partly explaining their association to AD at the protein level [5].

The last group of genes we find evidence for association for is the BAG domain protein family. A direct physical interaction with HSP70 was shown for all members of the BAG family [43]. The BAG proteins have therefore been proposed to serve as targeting factors for HSP70 [44]. However there is little evidence for association of the BAG family to AD in the paper of Hong *et al* [36].

Our study has a major advantage since the discovery cohort is embedded within a large, population based-study (RS1). Further, the studies used to evaluate the consistency of effects are also established epidemiological cohorts. We did not reach genome wide significance in the present study. A limitation in the interpretation of our data is that our findings were significant in the meta-analysis but did not reach significance in the individual cohorts except for in the RS1 cohort. The smaller number of cases in CHS and FHS in combination with the small effect in terms of OR could explain why the finding was not significant in those populations.

Until now, no large-scale studies have been performed on the role of HSP70 and its co-chaperone genes in patients with AD. In our genomic study in three population-based cohorts, we found evidence suggesting that genetic variants in the prefoldin family are associated with the risk of AD. Although larger case-series are needed to achieve genome wide significance for individual SNPs, combining these data with those of earlier functional studies makes it likely that this family of genes plays a role in AD.

REFERENCES

1. Muchowski, P.J. and J.L. Wacker, Modulation of neurodegeneration by molecular chaperones. Nat Rev Neurosci, 2005. 6(1): p. 11-22.
2. Selkoe, D.J., Translating cell biology into therapeutic advances in Alzheimer's disease. Nature, 1999. 399(6738 Suppl): p. A23-31.
3. Selkoe, D.J., Folding proteins in fatal ways. Nature, 2003. 426(6968): p. 900-4.
4. Chen, S. and I.R. Brown, Neuronal expression of constitutive heat shock proteins: implications for neurodegenerative diseases. Cell Stress Chaperones, 2007. 12(1): p. 51-8.
5. Magrane, J., et al., Heat shock protein 70 participates in the neuroprotective response to intracellularly expressed beta-amyloid in neurons. J Neurosci, 2004. 24(7): p. 1700-6.
6. Sherman, M.Y. and A.L. Goldberg, Cellular defenses against unfolded proteins: a cell biologist thinks about neurodegenerative diseases. Neuron, 2001. 29(1): p. 15-32.
7. Kakimura, J., et al., Microglial activation and amyloid-beta clearance induced by exogenous heat-shock proteins. FASEB J, 2002. 16(6): p. 601-3.
8. Dou, F., et al., Chaperones increase association of tau protein with microtubules. Proc Natl Acad Sci U S A, 2003. 100(2): p. 721-6.
9. Qiu, X.B., et al., The diversity of the DnaJ/Hsp40 family, the crucial partners for Hsp70 chaperones. Cell Mol Life Sci, 2006. 63(22): p. 2560-70.
10. Gibbs, S.J. and J.E. Braun, Emerging roles of J proteins in neurodegenerative disorders. Neurobiol Dis, 2008. 32(2): p. 196-9.
11. Hartl, F.U. and M. Hayer-Hartl, Molecular chaperones in the cytosol: from nascent chain to folded protein. Science, 2002. 295(5561): p. 1852-8.
12. Mayer, M.P. and B. Bukau, Hsp70 chaperones: cellular functions and molecular mechanism. Cell Mol Life Sci, 2005. 62(6): p. 670-84.
13. Loring, J.F., et al., A gene expression profile of Alzheimer's disease. DNA Cell Biol, 2001. 20(11): p. 683-95.
14. Hofman, A., et al., The Rotterdam Study: 2010 objectives and design update. Eur J Epidemiol, 2009. 24(9): p. 553-72.
15. Ruitenberg, A., et al., Incidence of dementia: does gender make a difference? Neurobiol Aging, 2001. 22(4): p. 575-80.
16. Folstein, M.F., S.E. Folstein, and P.R. McHugh, "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res, 1975. 12(3): p. 189-98.
17. Copeland, J.R., et al., A semi-structured clinical interview for the assessment of diagnosis and mental state in the elderly: the Geriatric Mental State Schedule. I. Development and reliability. Psychol Med, 1976. 6(3): p. 439-49.
18. Association, A.P., Diagnostic and statistical manual of mental disorders, in Diagnostic and statistical manual of mental disorders. 1987, American Psychiatric In. p. 608.
19. McKhann, G., et al., Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. Neurology, 1984. 34(7): p. 939-44.
20. Miller, S.A., D.D. Dykes, and H.F. Polesky, A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res, 1988. 16(3): p. 1215.
21. Seshadri, S., et al., Genome-wide analysis of genetic loci associated with Alzheimer disease. JAMA, 2010. 303(18): p. 1832-40.
22. Aulchenko, Y.S., M.V. Struchalin, and C.M. van Duijn, ProbABEL package for genome-wide association analysis of imputed data. BMC Bioinformatics, 2010. 11: p. 134.
23. Churchill, G.A. and R.W. Doerge, Empirical threshold values for quantitative trait mapping. Genetics, 1994. 138(3): p. 963-71.

24. Fisher, R., The Design of Experiments. 3 ed. Vol. 3. 1935, London: Oliver & Boyd Ltd. 252.
25. Psaty, B.M., et al., Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium: Design of prospective meta-analyses of genome-wide association studies from 5 cohorts. Circ Cardiovasc Genet, 2009. 2(1): p. 73-80.
26. Fried, L.P., et al., The Cardiovascular Health Study: design and rationale. Ann Epidemiol, 1991. 1(3): p. 263-76.
27. Fitzpatrick, A.L., et al., Incidence and prevalence of dementia in the Cardiovascular Health Study. J Am Geriatr Soc, 2004. 52(2): p. 195-204.
28. Lopez, O.L., et al., Evaluation of dementia in the cardiovascular health cognition study. Neuroepidemiology, 2003. 22(1): p. 1-12.
29. Dawber, T.R. and W.B. Kannel, The Framingham study. An epidemiological approach to coronary heart disease. Circulation, 1966. 34(4): p. 553-5.
30. Kannel, W.B., et al., An investigation of coronary heart disease in families. The Framingham offspring study. Am J Epidemiol, 1979. 110(3): p. 281-90.
31. Bachman, D.L., et al., Incidence of dementia and probable Alzheimer's disease in a general population: the Framingham Study. Neurology, 1993. 43(3 Pt 1): p. 515-9.
32. Beiser, A., et al., Computing estimates of incidence, including lifetime risk: Alzheimer's disease in the Framingham Study. The Practical Incidence Estimators (PIE) macro. Stat Med, 2000. 19(11-12): p. 1495-522.
33. Au, R., et al., New norms for a new generation: cognitive performance in the framingham offspring cohort. Exp Aging Res, 2004. 30(4): p. 333-58.
34. DeCarli, C., et al., Measures of brain morphology and infarction in the framingham heart study: establishing what is normal. Neurobiol Aging, 2005. 26(4): p. 491-510.
35. O'Dushlaine, C., et al., The SNP ratio test: pathway analysis of genome-wide association datasets. Bioinformatics, 2009. 25(20): p. 2762-3.
36. Hong, M.G., et al., Genome-wide pathway analysis implicates intracellular transmembrane protein transport in Alzheimer disease. J Hum Genet, 2010. 55(10): p. 707-9.
37. Kampinga, H.H., et al., Guidelines for the nomenclature of the human heat shock proteins. Cell Stress Chaperones, 2009. 14(1): p. 105-11.
38. Butler, A.W., et al., Meta-analysis of linkage studies for Alzheimer's disease--a web resource. Neurobiol Aging, 2009. 30(7): p. 1037-47.
39. Sakono, M., et al., Formation of highly toxic soluble amyloid beta oligomers by the molecular chaperone prefoldin. FEBS J, 2008. 275(23): p. 5982-93.
40. Kobayashi, Y., et al., Chaperones Hsp70 and Hsp40 suppress aggregate formation and apoptosis in cultured neuronal cells expressing truncated androgen receptor protein with expanded polyglutamine tract. J Biol Chem, 2000. 275(12): p. 8772-8.
41. Acebron, S.P., et al., DnaJ recruits DnaK to protein aggregates. J Biol Chem, 2008. 283(3): p. 1381-90.
42. Laufen, T., et al., Mechanism of regulation of hsp70 chaperones by DnaJ cochaperones. Proc Natl Acad Sci U S A, 1999. 96(10): p. 5452-7.
43. Takayama, S., Z. Xie, and J.C. Reed, An evolutionarily conserved family of Hsp70/Hsc70 molecular chaperone regulators. J Biol Chem, 1999. 274(2): p. 781-6.
44. Takayama, S. and J.C. Reed, Molecular chaperone targeting and regulation by BAG family proteins. Nat Cell Biol, 2001. 3(10): p. E237-41.

Chapter 3.2

Association of Heat Shock Proteins with Parkinson's Disease

Linda Broer, Peter J Koudstaal, Najaf Amin, Fernando Rivadeneira, Andre G. Uitterlinden, Albert Hofman, Ben A. Oostra, Monique MB Breteler, M. Arfan Ikram, Cornelia M. van Duijn

Published: European Journal of Epidemiology; 2011

Parkinson's disease (PD) is the most common neurodegenerative movement disorder, characterized by progressive degeneration of dopaminergic neurons in the *substantia nigra* and the presence of Lewy bodies in the affected neurons, comprised of aggregated α-synuclein [1-2]. PD can therefore be considered an example of a protein-folding disorder [3]. Heat Shock Proteins (HSPs) are the major chaperones mediating (re)folding of proteins, ensuring that these proteins stay in their native formations during conditions of stress to the cell [4]. We tested 27 genes encoding all members of the HSP10, HSP60, HSP70, HSP90 and the heat shock factors (HSF) families and assessed their association with PD.

Our discovery cohort was the Rotterdam Study (RS), which is a population-based cohort study that investigates the occurrence and determinants of diseases in the elderly [5]. The Medical Ethics Committee at Erasmus Medical Center approved the study protocol. Genomic DNA was extracted from whole blood samples using standard methods [6]. Genome-wide SNP genotyping was performed using Infinium II assay on the HumanHap550 Genotyping BeadChips (Illumina Inc, San Diego, USA). Approximately 2 million SNPs were imputed using release 22 HapMap CEU population as reference. The imputations were performed using MACH software (http://www.sph.umich.edu/csg/abecasis/MACH/).

A total 3828 SNPs in 27 genes were initially selected for the association test on the basis of the following criteria: (1) position within the genes of interest with a margin of 100kb on each side of the genes according to NCBI build 36.3, (2) p-value for Hardy-Weinberg equilibrium test ≥ 0.0001, and (3) call rate ≥ 95%. For further analysis and selection of SNPs for replication analysis only SNPs with an imputation quality greater than 80% and a minor allele frequency (MAF) higher than 0.05 were selected.

We analyzed the individual SNPs using ProbABEL [7]. We used allele-based logistic regression to test the association between a single SNP and PD. Odds Ratios (ORs) for each SNP were derived adjusting for age and sex. To calculate empirical significance for SNPs, permutations were performed per region of interest (ROI). Briefly, we performed 10,000 replications with randomly allocated phenotypes. To estimate empirical significance each observed test statistic was compared with null statistics obtained empirically and the p-value was estimated as the proportion of replicas generating the test statistics greater than or equal to the observed statistic.

For each suggestively associated ROI ($\text{p-value}_{\text{permuted}} < 0.10$) we next selected the 'independently associated' SNPs in a backward stepwise logistic regression until only nominally significant SNPs remained in the model. These independently associated SNPs were then followed up. Further suggestively associated SNPs inside the gene were also sent for replication. For replication we used online available genome wide association studies (GWAS) on PD and selected the study from Simón-Sánchez *et al* published in Nature Genetics in 2009 [8]. This study included a total of 5,197 PD cases and 8,803 controls. In their supplementary tables, the authors present p-values and effect estimates for all SNPs with a p-value lower than $1*10^{-4}$ in the discovery stage.

General characteristics of the discovery population are shown in **Table 1**. A total of 134 cases of Parkinson Disease were identified. The cases are significantly older than controls. A summary of all SNPs tested in the HSP genes in the discovery phase can be found in **Supplementary Table 1**. As HSP10 (*HSPE1*) and HSP60 (*HSPD1*) are located head to head on chromosome 2 these were taken as 1 Region of Interest (ROI) with 52 SNPs (Chaperonin). The same applies to 3 HSP70 genes located on chromosome 6 with 148 SNPs (HSPA1).

Table 1 | General Characteristics for cases and controls in the discovery cohort.

	Cases	Controls
N	134	5422
Age (mean ± sd)	74.2 ± 8.0	68.5 ± 8.5
Gender (% female)	58.5	53.0

Figure 1 shows an overview of the p-values in all genes tested in RS1. After adjusting for multiple testing by permutation analysis, 3 genes still showed evidence for assocation to PD in RS1. These genes encode Chaperonin, HSP70 member *HSPA13* and HSF member *HSF5*. For replication the following SNPs were selected: rs17730989 and rs788016 for Chaperonin, rs2822686 for *HSPA13* and rs9889631 for *HSF5*. Two SNPs in Chaperonin were selected as both were intronic SNPs for *HSPD1*. The results for the SNPs we attempted to replicate are in **Table 2**. None of these SNPs had a p-value lower than $1*10^{-4}$ in our replication sample. We did not find significant evidence for an association between SNPs in or directly flanking the HSP genes with PD.

Chaperonin was the best candidate for association with PD to reach significance in the discovery cohort, but none of the SNPs in *HSPD1* were replicated. Chaperonin is an interesting candidate for association with PD as it is the major folding machine of the mitochondria [9]. Its important influence on mitochondrial function is underlined by findings that yeast cells carying a null mutation in HSP60 have severe defects in folding of mitochondrial proteins and are non-viable [10]. In contrast, yeast cells with conditional mutations in HSP60 tend to accumulate misfolded proteins that are unable to perform their functions, similar to α-synculein aggregation in PD [10].

The importance of mitochondrial dysfunction in PD is widely accepted. All major genes previously identified for familial PD influence mitochondrial functioning [11]. The importance of *HSPD1* in mitochondrial functioning could be an explanation for the absence of an association with PD. Only variants with very small ORs are to be expected in such a critical gene, leaving our study underpowered to reach statistical significance.

In summary, we examined associations between 3828 SNPs in and around 27 genes encoding the Chaperonin, HSP70, HSP90 and HSF gene families in two unrelated cohorts, and found no statistically significant associations with PD. Despite evidence on the protein level for involvement of these genes, no large-scale studies have been performed on the role of HSPs in patients with PD, until now.

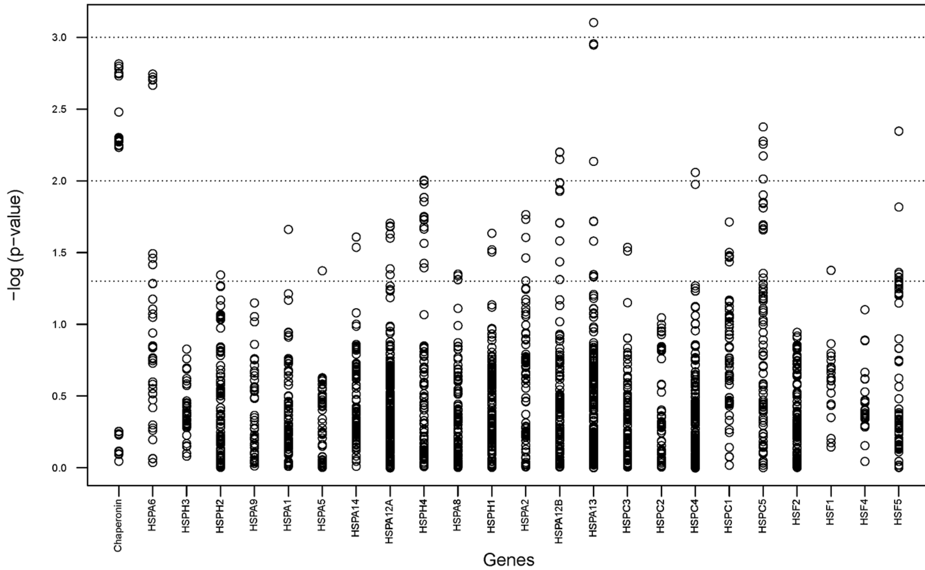


Figure 1 | The -log(p-values) of all SNPs tested are presented, where the SNPs are sorted on the genes on x-axis. The horizontal lines show the -log(p-values) of 0.05, 0.01 and 0.001 respectively.

Table 2 | Association of the SNPs significantly associated with PD in the Rotterdam Study.

					Rotterdam Study			Simón-Sánchez *et al.*		
Gene	**SNP**	**Chr**	**Position**	**RA**	**Freq**	**OR**	**95% CI**	**Freq**	**OR**	**95% CI**
HSPD1	rs17730989	2	198070769	C	0.49	0.67	0.52-0.86	NF	NF	NF
HSPD1	rs788016	2	198060538	G	0.49	0.68	0.53-0.87	NF	NF	NF
HSPA13	rs2822686	21	14773971	T	0.82	0.61	0.46-0.8	NF	NF	NF
HSF5	rs9889631	17	53985358	T	0.93	0.55	0.37-0.81	NF	NF	NF

RS1: Rotterdam study; Chr: chromosome; RA: risk allele; Freq: frequency; OR: Odds Ratio; 95% CI: 95% Confidence Interval; NF: not found with a p-value of 10^{-4}.

REFERENCES

1. Baba, M., et al., Aggregation of alpha-synuclein in Lewy bodies of sporadic Parkinson's disease and dementia with Lewy bodies. Am J Pathol, 1998. 152(4): p. 879-84.
2. Dawson, T.M. and V.L. Dawson, Molecular pathways of neurodegeneration in Parkinson's disease. Science, 2003. 302(5646): p. 819-22.
3. Selkoe, D.J., Folding proteins in fatal ways. Nature, 2003. 426(6968): p. 900-4.
4. Chen, S. and I.R. Brown, Neuronal expression of constitutive heat shock proteins: implications for neurodegenerative diseases. Cell Stress Chaperones, 2007. 12(1): p. 51-8.
5. Hofman, A., et al., The Rotterdam Study: 2012 objectives and design update. Eur J Epidemiol, 2011. 26(8): p. 657-86.
6. Miller, S.A., D.D. Dykes, and H.F. Polesky, A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res, 1988. 16(3): p. 1215.
7. Aulchenko, Y.S., M.V. Struchalin, and C.M. van Duijn, ProbABEL package for genome-wide association analysis of imputed data. BMC Bioinformatics, 2010. 11: p. 134.
8. Simon-Sanchez, J., et al., Genome-wide association study reveals genetic risk underlying Parkinson's disease. Nat Genet, 2009. 41(12): p. 1308-12.
9. Ostermann, J., et al., Protein folding in mitochondria requires complex formation with hsp60 and ATP hydrolysis. Nature, 1989. 341(6238): p. 125-30.
10. Cheng, M.Y., et al., Mitochondrial heat-shock protein hsp60 is essential for assembly of proteins imported into yeast mitochondria. Nature, 1989. 337(6208): p. 620-5.
11. Abou-Sleiman, P.M., M.M. Muqit, and N.W. Wood, Expanding insights of mitochondrial dysfunction in Parkinson's disease. Nat Rev Neurosci, 2006. 7(3): p. 207-19.

Chapter 3.3

Association of Heat Shock Proteins with all-cause mortality

L. Broer, E.W. Demerath, M.E. Garcia, G. Homuth, R.C. Kaplan, K.L. Lunetta, T. Tanaka, G.J. Tranah, S. Walter, A.M. Arnold, G. Atzmon, T.B. Harris, W. Hoffmann, D. Karasik, D.P. Kiel, T. Kocher, L.J. Launer, K.K. Lohman, J.I. Rotter, H. Tiemeier, A.G. Uitterlinden, H. Wallaschofski, S. Bandinelli, M. Dörr, L. Ferrucci, N. Franceschini, V. Gudnason, A. Hofman, Y. Liu, J.M. Murabito, A.B. Newman, B.A. Oostra, B.M. Psaty, A.V. Smith, C.M. van Duijn

Published: Age; 2013

ABSTRACT

Experimental mild heat shock is widely known as an intervention that results in extended longevity in various models along the evolutionary lineage. Heat shock proteins (HSPs), are highly up-regulated immediately after a heat shock. The elevation in HSP levels was shown to inhibit stress-mediated cell death and recent experiments indicate a highly versatile role for these proteins as inhibitors of programmed cell death.

In this study, we examined common genetic variations in 31 genes encoding all members of the HSP70, small HSP (sHSP) and Heat Shock Factor (HSF) families for their association with all-cause mortality. Our discovery cohort was the Rotterdam Study (RS1) containing 5974 participants aged 55 years and older (3174 deaths). We assessed 4430 Single Nucleotide Polymorphisms (SNPs) using the HumanHap550K Genotyping BeadChip from Illumina.

After adjusting for multiple testing by permutation analysis, 3 SNPs showed evidence for association with all-cause mortality in RS1. These findings were followed in 8 independent population-based cohorts, leading to a total of 25007 participants (8444 deaths). In the replication phase only *HSF2* (rs1416733) remained significantly associated with all-cause mortality. Rs1416733 is a known cis-eQTL for *HSF2*.

Our findings suggest a role of *HSF2* in all-cause mortality.

INTRODUCTION

Experimental mild heat shock is widely known as an intervention that results in extended longevity [1]. Brief exposure to elevated heat resulted in a 15% increase in the mean life span of *Caenorhabditis elegans*, compared to non heat-shocked controls [1-3]. Similar effects have also been seen in *Drosophila melanogaster* [4-5], in yeast [6] and in cultured human cells [7]. In the early 1960s a group of proteins, now known as heat shock proteins (HSPs), were discovered, which were highly up-regulated immediately after a heat shock [8-9]. Whether HSPs are responsible for longevity is still under debate, as their levels are only elevated for a short period of time after a heat shock [10]. However, the elevation in HSP levels during the heat shock response was shown to inhibit stress-mediated cell death, and recent experiments indicate a highly versatile role for these proteins as inhibitors of programmed cell death [11].

HSPs can be subdivided in several smaller families, including HSP90, HSP70, HSP60, HSP40, small HSP (sHSP) and HSP10 [12]. From these families, HSP70 and sHSPs show an association with longevity. In *C. elegans* extra copies of a homolog of *HSPA9* (member of HSP70), otherwise known as mortalin, extended life span up to 45% [13]. In humans decreased serum levels of HSP70 have been associated with exceptional longevity (95+) [14]. However, the same study evaluated 2 Single Nucleotide Polymorphisms (SNPs) in *HSPA1A* and *HSPA1B* which were not found to be associated to exceptional longevity [14].

The over-expression of members of the sHSP family has been shown to extend life of *C. elegans* and *D. melanogaster* by up to 32% [15-16]. Conversely, the absence of expression of a sHSP member decreases lifespan of *D. melanogaster* by 40% [17].

HSP expression is regulated by a group of transcription factors known as heat shock factors (HSFs), of which HSF1 is considered to be the master-switch of HSP expression [18]. Strong evidence exists for a highly important role for HSF1 in longevity. Reduced activity of HSF1 in *C. elegans* leads to a rapid aging phenotype with a markedly reduced lifespan of 60% [19]. Conversely, animals with an additional HSF1 gene copy lived approximately 40% longer than normal [20]. A strong relationship was found between HSF1 and DAF-16, which functions in the *C. elegans* insulin/IGF-1 signaling pathway [20]. Both genes were shown to function, at least in part, by increasing sHSP gene expression [20].

We have tested 31 genes encoding all members of the HSP70, sHSP and HSF families and assessed their association with all-cause mortality. To our knowledge, this is the first large-scale candidate gene study of these HSPs and their association to all-cause mortality to be performed.

METHODS

Discovery Study

Our discovery cohort was the Rotterdam Study (RS1). RS1 is a population-based cohort study that investigates the occurrence and determinants of diseases in the elderly [21]. Baseline examinations, including a detailed questionnaire, physical examination and blood collection, were conducted between 1990 and 1993. The Medical Ethics Committee at Erasmus Medical Center approved the study protocol. All of the participants were followed for incident diseases through linkage to the general practitioner data base and record review by trained medical investigators. General practitioners', hospital records as well as death certificates were used for identification of deaths (all-cause mortality) through 01.01.2009.

Genomic DNA was extracted from whole blood samples using standard methods [22]. Genome-wide SNP genotyping was performed using Infinium II assay on the HumanHap550 Genotyping BeadChips (Illumina Inc, San Diego, USA). Approximately 2 million SNPs were imputed using release 22 HapMap CEU population as reference. The imputations were performed using MACH software (http://www.sph.umich.edu/csg/abecasis/MACH/). The quality of imputations were checked by contrasting imputed and actual genotypes at 78,844 SNPs not present on Illumina 550K for 437 individuals for whom these SNPs were directly typed using Affymetrix 500K. Using the "best guess" genotype for imputed SNPs the concordance rate was 99% for SNPs with R^2 (ratio of the variance of imputed genotypes to the binomial variance) quality measure greater than 0.9; concordance was still over 90% (94%) when R^2 was between 0.5 and 0.9. The GWAS of all-cause mortality has been analyzed and is published elsewhere [23].

For the study of HSPs presented here, a total of 4430 SNPs in 31 genes were initially selected for the association test on the basis of the following criteria: (1) position within the genes of interest with a margin of 100kb on each side of the genes according to NCBI build 36.3, (2) P-value for Hardy-Weinberg equilibrium test ≥ 0.0001, and (3) call rate ≥ 95%. For further analysis and selection of SNPs for replication analysis only SNPs with a R^2 higher than 0.8 and a minor allele frequency (MAF) higher than 0.05 were selected.

Analysis

We performed single SNP analyses using ProbABEL [24]. We used survival analysis (semi-parametric Cox proportional hazard model), adjusted for age at DNA blood collection and for sex, to model continuous time to death in individuals that were older than 55 years at baseline. To calculate empirical significance for SNPs, permutations were performed per region of interest (ROI) [25-26]. Briefly, the empirical distribution of the region-wide maximum of the test statistic under the null was obtained in 10,000 replications. To estimate empirical significance each observed test statistic was compared with null statistics obtained empirically and the p-value was estimated as the proportion of replicas generating the test statistics greater than or equal to the observed statistic. The permutation analysis keeps

the original genotypes for each individual, but randomly allocates the phenotypes for each consecutive permutation. Therefore the linkage disequilibrium (LD) structure of genes is not broken up. We did not perform any additional correction for the number of ROIs after permutations as we used the permutation analysis to significantly reduce the number of SNPs to be selected for validation of the effect of the SNP. For each suggestively associated ROI ($p\text{-value}_{permuted} < 0.10$) we next selected the 'truly associated' SNPs in a backward stepwise survival analysis until only nominally significant SNPs remained in the model. These independently associated SNPs were then followed up in 8 independent cohorts. As effect estimates are expected to be small, a strict replication based on the p-value and hazard ratio (HR) observed in individual replication studies, might be difficult to achieve due to power issues. Therefore we evaluated whether a SNP was showing a HR in the same direction in all cohorts, i.e., whether findings were consistent over the various cohorts. Similar to genome wide association studies, we performed a joint meta-analysis of the discovery and the replication samples and tested whether the joint p-value was significant using a Bonferroni correction for the number of SNPs validated in the replication phase.

We used 4 studies that are part of the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium [27] plus an additional 4 associated cohorts to evaluate whether findings were consistent. The studies included in this replication are: Cardiovascular Health Study (CHS); Framingham Heart Study (FHS); Atherosclerosis Risk Communities Study (ARIC); Age, Gene/Environment Susceptibility -Reykjavik Study (AGES); Health, Aging and Body Composition (HABC); Baltimore Longitudinal Study of Ageing (BLSA); InCHIANTI (ICH); and Study of Health in Pomerania (SHIP). All studies are longitudinal population-based studies periodically assessing the health and vital status of their participants. All participants included in this analysis were at least 55 years of age at the time of blood draw for DNA and provided written informed consent. Detailed information on the replication studies can be found in the supplement.

RESULTS

General characteristics of the discovery cohort as well as the 8 replication cohorts are shown in **Table 1**. In the discovery cohort there were 3174 deaths with a mean age at death of 83.2 and mean follow-up of 12.5 years. In RS1 and replications combined there were 8444 deaths with a mean age at death of 81.1. The mean follow-up time ranged from 5.2 to 15.7 years. A summary of all SNPs tested in the HSP genes in RS1 can be found in **Supplementary Table 1**. Three HSP70 genes are located head to head on chromosome 6 and are taken as one ROI with 148 SNPs (HSPA1). Similarly, 2 sHSP genes on chromosome 11 with 66 SNPs were located nearby each other and taken as one ROI (ROIchr11). Also of note is that the HSPs are currently undergoing a change in nomenclature [12]. **Supplementary Table 1** provides

both the official gene names and the gene names according to the new nomenclature. Here we will use the aliases from the new nomenclature.

Table 1 | General characteristics.

Study	N	N deceased	Mean age at Baseline (±SD)	Mean age at death (±SD)	Sex, % female	Mean follow-up time in years (±SD)
RS1	5974	3174	69.4 (9.1)	83.2 (8.3)	59%	12.5 (5.2)
CHS	3267	1718	72.3 (5.4)	83.4 (6.3)	61%	12.3 (4.2)
FHS	3136	654	70.0 (10.2)	83.0 (9.3)	56%	6.0 (2.4)
ARIC	4511	1108	59.4 (2.9)	71.3 (5.4)	50%	15.7 (3.7)
AGES	3219	558	76.4 (5.5)	79.3 (5.9)	58%	5.2 (1.3)
HABC	1661	460	73.8 (2.8)	80.4 (3.7)	47%	8.2 (2.3)
BLSA	620	183	62.0 (8.8)	86.8 (8.0)	41%	15.7 (8.2)
InCHIANTI	902	183	72.5 (7.7)	85.4 (7.9)	56%	5.9 (0.9)
SHIP	1717	406	66.4 (7.2)	76.9 (7.2)	47%	9.2 (2.4)
TOTAL	25007	8444	69.0 (8.9)	81.1 (8.4)	55%	10.6 (5.4)

Figures 1 A-C illustrate the observed versus the expected chi^2 of the SNPs tested in the different HSP gene families in RS1. The plots show an excess of low p-values for tested SNPs in sHSP (**1B**) and HSF (**1C**), but not for HSP70 (**1A**). **Supplementary Figure 1** shows an overview of the p-values in all HSP70, sHSP and HSF genes tested in RS1. After adjusting for multiple testing by permutation analysis, 3 genes showed evidence for association to mortality in RS1. These genes encode HSP70 member *HSPA8*, sHSP member *HSPB1* and HSF member *HSF2*. **Table 2** shows the 3 SNPs that were selected for replication. In the replication phase only one SNP reached a Bonferroni adjusted p-value for significance (0.05/3 = 0.017). **Figure 2A** shows the Forest plot of the meta-analysis of *HSF2* rs1416733. The direction of effect was consistent across populations with only one study (HABC) showing an opposite HR. The summary HR for this SNP was 0.95 (95% CI 0.92-0.98; p-value = 0.003) for the A allele. When excluding the discovery cohort from the meta-analysis, the HR remained virtually unchanged (0.97: CI 0.93-1.00). In **Figure 2B** a regional plot of all SNPs tested in *HSF2* from the original RS1 cohort can be found. In the plot we can see that rs1416733 is located 11.7 Kb from the 5' region of the *HSF2* gene, with no other genes nearby.

Table 2 | Results of single SNP replication.

					RS1				CHARGE meta-analysis				
Family	**Gene**	**SNP**	**Non-coded Allele**	**Coded Allele**	**Frequency coded allele**	**HR**	**95% CI**	**p-value**	**Frequency coded allele**	**HR**	**95% CI**	**Study Effect Direction**	**p-value**
HSP70	HSPA8	rs12574703	G	A	0.05	1.21	1.08-1.36	0.001	0.05	1.07	0.99-1.14	+-+-++++-	0.080
sHSP	HSPB1	rs7797781	C	T	0.79	0.91	0.86-0.97	0.002	0.78	0.96	0.93-1.00	--+-+++++	0.035
HSF	**HSF2**	rs1416733	G	A	0.35	0.92	0.88-0.97	0.003	0.35	0.95	0.92-0.98	-----+---	**0.003**

Hazard Ratios (HR) and 95% Confidence Interval (95% CI) are for each additional coded allele
Study Effect Direction: study-specific information is presented in the order: RS1, CHS, FHS, ARIC, AGES, HABC, BLSA, ICH, SHIP
Direction: '+' stands for 'HR greater than 1', '-' stands for 'HR smaller than 1'
Bold shows which genes passed Bonferroni adjusted p-value for significance (0.05/3=0.017).

3

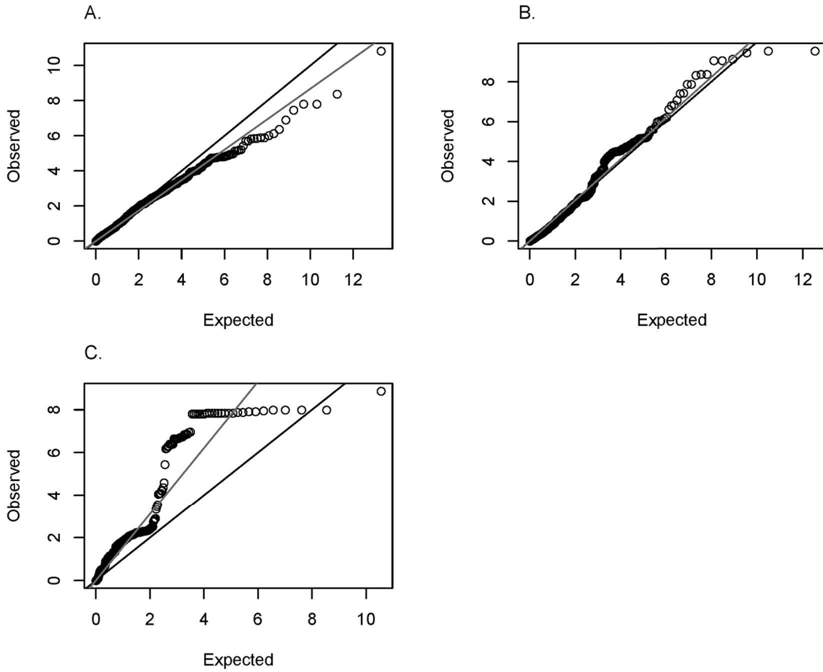
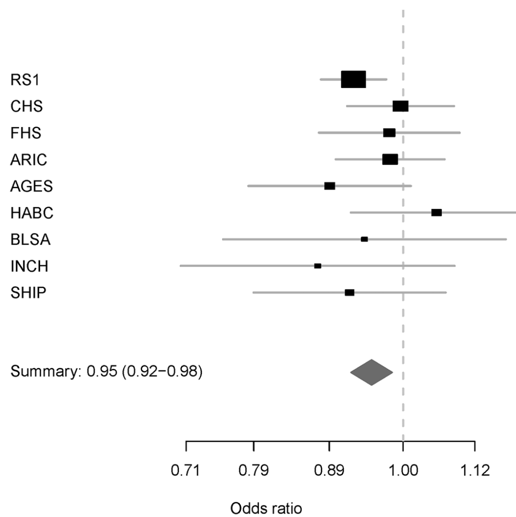
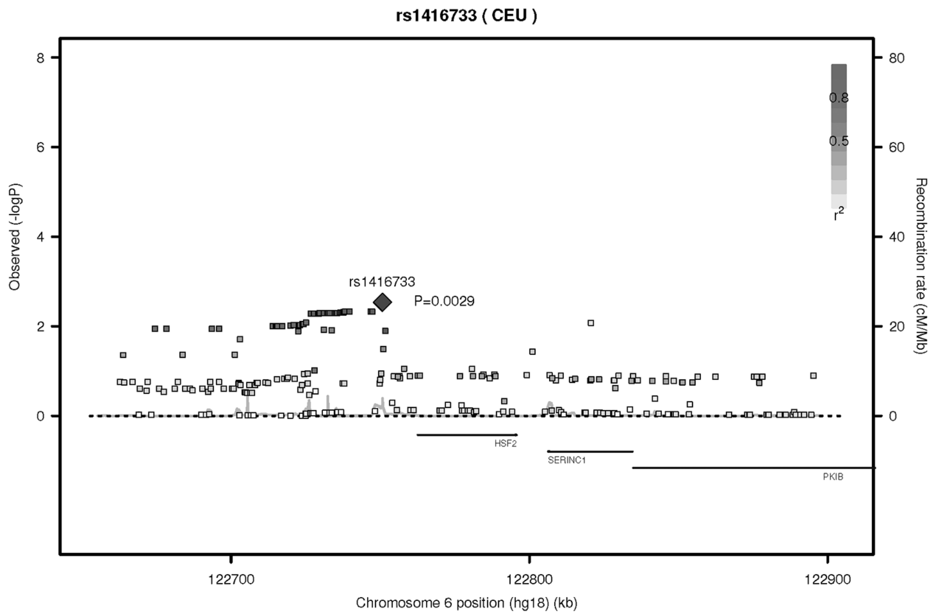


Figure 1 | Observed versus Expected p-value plot for the SNPs tested in the Rotterdam study (chi2 are given at the x-axis and y-axis) A. HSP70; B. sHSP; C. HSF

Legend: black line depicts the expected findings, red line the observed ones



2A | Forrest plot



2B | Regional plot in RS1

Figure 2 | The meta-analysis of the HSF2 rs1416733 SNP in the 9 population cohorts.

DISCUSSION

In our study, we found a significant association between *HSF2* and all-cause mortality. The top associated variant, rs1416733, is located 11.7 Kb from the 5' region of the *HSF2* gene on chromosome 6 and is a known cis-eQTL for *HSF2* [28]. Each additional copy of the A allele for SNP rs1416733 increased lifespan with a HR of 0.95 in a meta-analysis of 9 independent cohort studies. This effect was consistent in direction in 8 out of 9 cohorts.

The HSF family, like most HSP families, is a highly conserved family across species, indicating a vital role for the survival of the respective organism [29]. Unlike many other HSP families, there is virtually no redundancy in the HSF family, with HSF1 as the most crucial family member [18]. The importance of HSF1 is advocated strongly by the large effects of genetic mutations in this gene on longevity in *C. elegans* [19-20]. These large effects are not anticipated in human research, which could explain why we find no evidence for association of SNPs in *HSF1* with all-cause mortality [23]. HSF2 is less known in longevity research. It has previously been mainly described for its role as a development- and differentiation-specific factor [30-31]. The role of HSF2 in later life has remained unknown for a long time [32]. Recently evidence is emerging to suggest that HSF2 modulates HSF1 activity [18]. It has been shown that HSF2 activation leads to activation of HSF1, revealing a functional

interdependency [33]. It has been proposed that heterotrimerization of HSF1 and HSF2 integrates transcriptional activation in response to distinct stress and developmental stimuli [33]. Additionally, though HSF2 was never considered to be heat-inducible, a recent study shows that a mild heat-shock in the physiological range does activate HSF2 and has a significant impact on the proteostasis of the cell [34].

We found no evidence for association of HSP70 with all-cause mortality. HSP70 is the only HSP family for which SNPs in a couple of genes have been studied for longevity in humans previously. However, these studies only studied 2 or 3 SNPs and only in the heat-inducible members of HSP70, namely *HSPA1A*, *HSPA1B* and *HSPA1L*, all located on chromosome 6 [14, 35-36]. In our study no association was found for these 3 genes with longevity. We add a more thorough investigation of these genes and the other HSP70 genes by extensively covering common variance.

The sHSPs have been named most often in animal studies for their relationship with longevity [15-17]. However, we find no significant associations for single SNPs in these genes. Off course, what we find in animal studies doesn't always translate well to humans. For example, though genetic variations in IGF-1 signaling have been found in humans, they don't have nearly as strong effects on longevity as in animal models [37]ÿÿÿÿÿÿT. Another study has found that mutations in HSF1 lead to up-regulated sHSP expression in *C. elegans* [20]. We cannot exclude that genetic variation in *HSF2* activates a similar mechanism in humans, through its effect on HSF1.

Our study has a major advantage since the discovery cohort – the Rotterdam Study – is a large, population based-study. Further, the eight replication cohorts are also relatively large, established population-based epidemiological cohorts. A limitation in the interpretation of our data is that our findings were significant in the meta-analysis but did not reach significance in the individual cohorts except for the RS1 cohort; however a total of eight out of nine studies showed an effect in the same direction for *HSF2*. This could be explained by the markedly smaller percentage of deaths in most replication cohorts (except CHS) compared to the discovery cohort. The number of deceased is critical for the statistical power of the study. Even though some of these studies had a longer mean follow-up time (ARIC and BLSA), the population in these studies were on average younger at baseline compared to RS1, which explains the fewer number of deaths and leaves the individual replication studies underpowered to identify rs1416733.

Until now, no large-scale studies have been performed investigating the role of HSP70, sHSP or HSF genes in all-cause mortality in humans. In our candidate gene study in nine population-based cohorts, we found significant evidence suggesting that genetic variants in *HSF2* are associated with all-cause mortality. Combining these data with those of earlier functional studies, in particular in *C. elegans*, makes it likely that *HSF2* plays a role in human all-cause mortality.

REFERENCES

1. Cypser, J. and T.E. Johnson, Hormesis extends the correlation between stress resistance and life span in long-lived mutants of Caenorhabditis elegans. Hum Exp Toxicol, 2001. 20(6): p. 295-6; discussion 319-20.
2. Cypser, J.R. and T.E. Johnson, Multiple stressors in Caenorhabditis elegans induce stress hormesis and extended longevity. J Gerontol A Biol Sci Med Sci, 2002. 57(3): p. B109-14.
3. Lithgow, G.J., et al., Thermotolerance and extended life-span conferred by single-gene mutations and induced by thermal stress. Proc Natl Acad Sci U S A, 1995. 92(16): p. 7540-4.
4. Hercus, M.J., V. Loeschcke, and S.I. Rattan, Lifespan extension of Drosophila melanogaster through hormesis by repeated mild heat stress. Biogerontology, 2003. 4(3): p. 149-56.
5. Le Bourg, E., et al., Effects of mild heat shocks at young age on aging and longevity in Drosophila melanogaster. Biogerontology, 2001. 2(3): p. 155-64.
6. Shama, S., et al., Heat stress-induced life span extension in yeast. Exp Cell Res, 1998. 245(2): p. 379-88.
7. Rattan, S.I., Repeated mild heat shock delays ageing in cultured human skin fibroblasts. Biochem Mol Biol Int, 1998. 45(4): p. 753-9.
8. Ritossa, F., A new puffing pattern induced by heat shock and DNP in Drosophila. Cellular and molecular life sciences, 1962. 18: p. 571.
9. Ritossa, F., Discovery of the heat shock response. Cell Stress Chaperones, 1996. 1(2): p. 97-8.
10. Link, C.D., et al., Direct observation of stress response in Caenorhabditis elegans using a reporter transgene. Cell Stress Chaperones, 1999. 4(4): p. 235-42.
11. Garrido, C., et al., Heat shock proteins 27 and 70: anti-apoptotic proteins with tumorigenic properties. Cell Cycle, 2006. 5(22): p. 2592-601.
12. Kampinga, H.H., et al., Guidelines for the nomenclature of the human heat shock proteins. Cell Stress Chaperones, 2009. 14(1): p. 105-11.
13. Yokoyama, K., et al., Extended longevity of Caenorhabditis elegans by knocking in extra copies of hsp70F, a homolog of mot-2 (mortalin)/mthsp70/Grp75. FEBS Lett, 2002. 516(1-3): p. 53-7.
14. Terry, D.F., et al., Serum heat shock protein 70 level as a biomarker of exceptional longevity. Mech Ageing Dev, 2006. 127(11): p. 862-8.
15. Morrow, G., et al., Overexpression of the small mitochondrial Hsp22 extends Drosophila life span and increases resistance to oxidative stress. FASEB J, 2004. 18(3): p. 598-9.
16. Walker, G.A., et al., Heat shock protein accumulation is upregulated in a long-lived mutant of Caenorhabditis elegans. J Gerontol A Biol Sci Med Sci, 2001. 56(7): p. B281-7.
17. Morrow, G., et al., Decreased lifespan in the absence of expression of the mitochondrial small heat shock protein Hsp22 in Drosophila. J Biol Chem, 2004. 279(42): p. 43382-5.
18. Akerfelt, M., R.I. Morimoto, and L. Sistonen, Heat shock factors: integrators of cell stress, development and lifespan. Nat Rev Mol Cell Biol, 2010. 11(8): p. 545-55.
19. Garigan, D., et al., Genetic analysis of tissue aging in Caenorhabditis elegans: a role for heat-shock factor and bacterial proliferation. Genetics, 2002. 161(3): p. 1101-12.
20. Hsu, A.L., C.T. Murphy, and C. Kenyon, Regulation of aging and age-related disease by DAF-16 and heat-shock factor. Science, 2003. 300(5622): p. 1142-5.
21. Hofman, A., et al., The Rotterdam Study: 2012 objectives and design update. Eur J Epidemiol, 2011. 26(8): p. 657-86.
22. Miller, S.A., D.D. Dykes, and H.F. Polesky, A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res, 1988. 16(3): p. 1215.
23. Walter, S., et al., A genome-wide association study of aging. Neurobiol Aging, 2011. 32(11): p. 2109 e15-28.

24. Aulchenko, Y.S., M.V. Struchalin, and C.M. van Duijn, ProbABEL package for genome-wide association analysis of imputed data. BMC Bioinformatics, 2010. 11: p. 134.
25. Churchill, G.A. and R.W. Doerge, Empirical threshold values for quantitative trait mapping. Genetics, 1994. 138(3): p. 963-71.
26. Fisher, R., The Design of Experiments. 3 ed. Vol. 3. 1935, London: Oliver & Boyd Ltd. 252.
27. Psaty, B.M., et al., Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium: Design of prospective meta-analyses of genome-wide association studies from 5 cohorts. Circ Cardiovasc Genet, 2009. 2(1): p. 73-80.
28. Zeller, T., et al., Genetics and beyond--the transcriptome of human monocytes and disease susceptibility. PLoS One, 2010. 5(5): p. e10693.
29. Liu, X.D., et al., Conservation of a stress response: human heat shock transcription factors functionally substitute for yeast HSF. EMBO J, 1997. 16(21): p. 6466-77.
30. Kallio, M., et al., Brain abnormalities, defective meiotic chromosome synapsis and female subfertility in HSF2 null mice. EMBO J, 2002. 21(11): p. 2591-601.
31. Wang, G., et al., Targeted disruption of the heat shock transcription factor (hsf)-2 gene results in increased embryonic lethality, neuronal defects, and reduced spermatogenesis. Genesis, 2003. 36(1): p. 48-61.
32. Wu, C., Heat shock transcription factors: structure and regulation. Annu Rev Cell Dev Biol, 1995. 11: p. 441-69.
33. Sandqvist, A., et al., Heterotrimerization of heat-shock factors 1 and 2 provides a transcriptional switch in response to distinct stimuli. Mol Biol Cell, 2009. 20(5): p. 1340-7.
34. Shinkawa, T., et al., Heat shock factor 2 is required for maintaining proteostasis against febrile-range thermal stress and polyglutamine aggregation. Mol Biol Cell, 2011. 22(19): p. 3571-83.
35. Altomare, K., et al., The allele (A)(-110) in the promoter region of the HSP70-1 gene is unfavorable to longevity in women. Biogerontology, 2003. 4(4): p. 215-20.
36. Ross, O.A., et al., Increased frequency of the 2437T allele of the heat shock protein 70-Hom gene in an aged Irish population. Exp Gerontol, 2003. 38(5): p. 561-5.
37. Kuningas, M., et al., Genes encoding longevity: from model organisms to humans. Aging Cell, 2008. 7(2): p. 270-80.

Chapter 4

The role of telomere length in morbidity and mortality

Chapter 4.1

Meta-analysis of telomere length in 19,713 subjects reveals high heritability, stronger maternal inheritance and a paternal age effect

Linda Broer, Veryan Codd, Dale R Nyholt, Joris Deelen, Massimo Mangino, Gonneke Willemsen, Eva Albrecht, Najaf Amin, Marian Beekman, Eco JC de Geus, Anjali Henders, Christopher P Nelson, Claire J Steves, Margie J Wright, Anton JM de Craen, Aaron Isaacs, Mary Matthews, Alireza Moayyeri, Grant W Montgomery, Ben A Oostra, Jacqueline M Vink, Tim D Spector, P. Eline Slagboom, Nicholas G Martin, Nilesh J Samani, Cornelia M van Duijn, Dorret I Boomsma

Published: European Journal of Human Genetics; 2013

ABSTRACT

Telomere length (TL) has been associated with ageing and mortality, but individual differences also are influenced by genetic factors with previous studies reporting heritability estimates ranging from 34% to 82%.

Here we investigate the heritability, mode of inheritance and the influence of parental age at birth on TL in six large, independent cohort studies with a total of 19,713 participants.

The meta-analysis estimate of TL heritability was 0.70 (95% CI 0.64 - 0.76) and is based on a pattern of results that is highly similar for twins and other family members. We observed a stronger mother-offspring ($r = 0.42$; p-value = $3.6*10^{-61}$) than father-offspring correlation ($r = 0.33$; p-value = $7.0*10^{-5}$), and a significant positive association with paternal age at offspring birth (beta = 0.005; p-value = $7.0*10^{-5}$). Interestingly, a significant and quite substantial correlation in TL between spouses ($r = 0.25$; p-value = $2.8*10^{-30}$) was seen, which appeared stronger in older spouse pairs (mean age ≥ 55 years; $r = 0.31$; p-value = $4.3*10^{-23}$) than in younger pairs (mean age < 55 years; $r = 0.20$; p-value = $3.2*10^{-10}$).

In summary, we find a high and very consistent heritability estimate for telomere length, evidence for a maternal inheritance component and a positive association with paternal age.

INTRODUCTION

Telomeres are specialized DNA structures located at the terminal ends of chromosomes [1]. Their primary function is to maintain genomic stability. Due to the inability of DNA polymerase to fully replicate the 3' end of the DNA strand, i.e. the 'end-replication problem', telomeres naturally shorten with each cell division and, therefore, with age [2-4]. In epidemiological studies, both of cross sectional and prospective design, decreased telomere length in leukocytes was associated with increased mortality [5-8], although this finding was not consistent [9, 10]. Increased telomere length in leukocytes has been associated with longevity in a comparison of long-lived Ashkenazi Jews and their offspring with younger controls [11]. Genetic association studies have revealed associations between single nucleotide polymorphisms (SNPs) in the *TERC* and *TERT* genes and telomere length [11-17]. Other studies reported associations between SNPs in *TERC* and *POT1* with human longevity [11, 17, 18], suggesting that genes regulating telomere length may influence human longevity.

That genetic variants would be found associated with telomere length was expected from earlier twin and family studies indicating genetic influence on telomere length. An early study in twins of different ages showed a heritability of 78% for telomere length in 4-, 17- and 44-year olds [4], with no evidence for genotype x age interaction in this age range. In older twins (N = 287 pairs aged 73 to 95 years) estimates for heritability were 34% in women, while for the smaller male sample the heritability was lower [19]. In a sample of 55 female monozygotic (MZ) and 1,025 dizygotic (DZ) middle-aged twin pairs, Andrew *et al* reported a heritability estimate of 36% and an unexpectedly large shared familial effect of 49% [20]. Another study comprising 258 sib-pairs returned a heritability estimate of 82% [21]. There is a need to reconcile these divergent estimates, which might in part be caused by differences in sample composition or telomere length assessments [22]. A remaining question that also needs to be addressed is the extent to which estimates of heritability based on twin data are comparable to those based on other relatives such as parents and offspring from multigenerational studies or non-twin sibling-pairs.

Several studies have been conducted to assess the potential mechanisms for telomere inheritance. An early study suggested that telomere length was maternally inherited via an X-linked mechanism [23]. However, an increasing number of larger studies have reported stronger father-offspring than mother-offspring correlations for telomere length, suggesting the inheritance of this trait is mainly paternally determined [8, 24-26]. In addition to a paternal pattern of inheritance, a positive correlation between offspring telomere length and paternal age at birth has been observed [24, 27, 28]. Unlike somatic cells, sperm cell telomere length has been shown to increase with the age of the donor [24, 29] due to the presence of active telomerase, suggesting that offspring of older fathers would inherit longer telomeres.

In the current study, we investigated the heritability, mode of inheritance and influence of parental age at birth on telomere length in six large, independent cohort studies with a total of 19,713 participants.

MATERIALS AND METHODS

Populations

The Erasmus Rucphen Family (ERF) study is a cross-sectional cohort including more than 3,000 living descendants of 22 couples who had at least 6 children baptized in the community church between 1850 and 1900. The participants were not selected on the basis of any disease or other outcome. Details about the genealogy of the population have been previously described [30, 31]. The study protocol was approved by the medical ethics board of the Erasmus MC Rotterdam, the Netherlands.

The GRAPHIC study is comprised of individuals from 520 white Caucasian nuclear families recruited from the general population in Leicestershire, UK, between 2003 and 2005, for the purpose of investigating the genetic determinants of blood pressure and related cardiovascular traits. Inclusion criteria were that both parents (aged 40-60 years) and two offspring ≥ 18 years were willing to participate. Further details are provided elsewhere [32].

For the Leiden Longevity Study (LLS), long-lived siblings of Dutch descent were recruited together with their offspring and the partners of the offspring. Families were included if at least two long-lived siblings were alive and fulfilled the age criterion of 89 years or older for males and 91 years or older for females, representing less than 0.5% of the Dutch population in 2001 [33]. In total, 944 long-lived proband siblings were included with a mean age of 94 years (range, 89-104), 1,671 offspring (61 years, 39-81), and 744 partners (60 years, 36-79). For this study only the offspring and their partners were included.

The Netherlands Twin Register (NTR: http://www.tweelingenregister.org/) recruits twins and their family members to study the causes of individual differences in health, behavior and lifestyle. Participants are followed longitudinally; details about the cohort have been published previously [34]. A subsample of unselected twins and their family members has taken part in the NTR-Biobank [35] in which biological samples, including DNA and RNA, were collected in a standardized manner after overnight fasting. Study protocols were approved by the medical ethics board of the VUMC Amsterdam, the Netherlands.

The Queensland Institute of Medical Research (QIMR) adolescent study comprised twins and their non-twin siblings living in south-east Queensland, Australia [36]. Most (98% by self-report) are of mixed European ancestry, mainly from the British Isles. The participants are not selected on the basis of any disease or other outcome. Blood samples were collected at the end of testing sessions from participants and, if possible, from their parents. Pedigree relationships and zygosity were confirmed by genotype data. Further details are provided elsewhere [37].

The TwinsUK cohort (www.twinsuk.ac.uk) is an adult twin British registry shown to be representative of singleton populations and the United Kingdom population [38]. A total of 6,038 twins with telomere length measurement were included in the analysis. The age range of the TwinsUK cohort was 16-99. Ethical approval was obtained from the Guy's and St. Thomas' Hospital Ethics /':Committee. Written informed consent was obtained from every participant in the study.

Telomere length assessment

All samples from all studies were measured in the same laboratory under standard conditions. Mean leukocyte telomere length was measured by quantitative PCR-based technique as previously described [12, 39]. This method expresses telomere length as a ratio (T/S) of telomere repeat length (T) to copy number of a single copy gene, 36B4(S), within each sample. Samples were quantified relative to a calibrator sample used on each run (DNA from the K562 cell line) [12]. Mean inter-run coefficient of variations (CVs) were calculated for all study cohorts and these ranged from 3.30-3.72%.

Analysis

To summarize resemblance between family members, correlations of telomere length between monozygotic (MZ) and dizygotic (DZ) twins, non-twin siblings, parents and offspring and between spouses, adjusting for age, gender and potential batch effects were calculated in R, version 2.12.1 [40]. If data on more than two siblings were available, the data from the oldest two non-twin siblings were analyzed.

Maximum likelihood analyses as implemented in POLY [41] was used to evaluate different models of familial resemblance and these analyses were based on all data (i.e. allowing for more than 2 siblings per family). In these analyses telomere length was adjusted for age, gender and batch by specifying these as fixed effects on telomere length. Data from the six cohorts were first analyzed individually followed by meta-analysis, whereby each cohort took into account the unique structure in the data, such as the highly complex pedigree structure in the ERF cohort. In those studies where the pedigree structure allowed for evaluating sub models that included an additive genetic variance component, a non-additive genetic variance component (dominance) or a sibling-shared environment variance component respectively, were compared. A model in which variance was partitioned into an additive genetic variance component and a random environmental variance component gave the best description of all data and heritability was estimated based on this model.

In order to investigate differences between maternal or paternal inheritance for telomere length, family trios or duos were selected. When data on two or more children were available, data from the oldest child were selected for this analysis. We assessed the age, gender and batch adjusted correlations between fathers and their offspring and mothers and their offspring.

Paternal and maternal ages at the birth of their children were obtained based on the date of birth of the parents and offspring if available, or by subtracting the child's age from the parental age. The association of paternal and maternal age with telomere length were compared by regressing telomere length on parental age in addition to own age and gender (e.g. telomere length ~ age + gender + parental age). In order to correct significance testing for familial dependencies, these analyses were also done in POLY. Meta-analyses were performed using the R package rmeta [42].

RESULTS

General characteristics of each cohort are described in **Table 1**. The mean age ranged from 27.9 (QIMR) to 59.2 (LLS). Overall, approximately half of the participants are female in each population. Three cohorts (NTR, QIMR and TwinsUK) had twin data, with NTR and TwinsUK having more MZ twins and QIMR more DZ twins. As expected telomere length showed a significant negative association with age and was significantly longer in women in each population (**Supplementary Table 1**).

Table 1 | General characteristics of the study populations.

	ERF	GRAPHIC	LLS	NTR	QIMR	TwinsUK
Sample size	2,769	2,011	2,322	4,550	2,025	6,036
Telomere length#	1.78 (0.36)	1.60 (0.26)	1.46 (0.27)	2.75 (0.48)	3.41 (0.61)	3.72 (0.68)
Age#	49.8 (14.9)	39.3 (14.5)	59.2 (6.8)	38.8 (14.1)	27.9 (15.6)	51.1 (13.5)
Age range	17-89	18-61	30-80	15-82	7-73	16-99
Women*	1,539 (55.6%)	995 (49.5%)	1,272 (54.8%)	2,951 (64.9%)	1,075 (53.1%)	5,461 (90.5%)
MZ twins*	NA	NA	NA	834 (36.7%)	154 (15.2%)	1,546 (51.2%)
DZ twins*	NA	NA	NA	689 (30.3%)	227 (22.4%)	1,256 (41.6%)

Mean (sd); * n (%); complete twin pairs are depicted, representing the percentage of total population with telomere length measurements

MZ twins were significantly more alike than DZ twins (p-value = $1.85*10^{-24}$, **Table 2**) and siblings were significantly correlated within each cohort and this correlation was very similar to that of DZ twins (**Table 2**). Spouses had significantly correlated telomere lengths (r = 0.25; p-value = $2.82*10^{-30}$) and this effect was more pronounced in older spouses (mean age ≥ 55 years) compared to younger spouses (mean age < 55 years) (p-value = 0.010; **Table 2**) suggesting that there is a significant influence of environmental factors.

Table 2 | Sibling, twin and spousal correlations, adjusted for age, gender and batch effects.

	n	r	p-value
Siblings			
ERF	444	0.48	$5.20*10^{-27}$
GRAPHIC	481	0.60	$4.76*10^{-49}$
LLS	419	0.41	$5.30*10^{-18}$
NTR	47	0.43	0.003
QIMR	162	0.32	$3.84*10^{-5}$
TwinsUK	NA	NA	NA
Meta-analysis	1,553	0.49	$3.46*10^{-96}$
Monozygotic twins			
ERF	NA	NA	NA
GRAPHIC	NA	NA	NA
LLS	NA	NA	NA
NTR	834	0.62	$7.70*10^{-90}$
QIMR	154	0.70	$3.17*10^{-24}$
TwinsUK	1,546	0.72	$1.93*10^{-245}$
Meta-analysis	2,534	0.69	0*
Dizygotic twins			
ERF	NA	NA	NA
GRAPHIC	NA	NA	NA
LLS	NA	NA	NA
NTR	689	0.34	$2.01*10^{-20}$
QIMR	227	0.57	$4.96*10^{-21}$
TwinsUK	1,256	0.57	$5.21*10^{-110}$
Meta-analysis	2,172	0.50	$1.24*10^{-146}$
Spouses (all)			
ERF	218	0.28	$2.34*10^{-5}$
GRAPHIC	502	0.19	$1.45*10^{-5}$
LLS	702	0.34	$6.74*10^{-20}$
NTR	108	0.32	0.001
QIMR	410	0.15	0.002
TwinsUK	NA	NA	NA
Meta-analysis	1,940	0.25	$2.82*10^{-30}$
Spouses (< 55)			
ERF	78	0.21	0.065
GRAPHIC	311	0.17	0.003
LLS	183	0.39	$4.20*10^{-8}$
NTR	22	-0.05	0.809

4

	n	r	p-value
Spouses (< 55) *(cont'd)*			
QIMR	368	0.14	0.006
TwinsUK	NA	NA	NA
Meta-analysis	962	0.20	$3.24*10^{-10}$
Spouses (> 55)			
ERF	140	0.33	$8.80*10^{-5}$
GRAPHIC	191	0.24	$6.70*10^{-4}$
LLS	519	0.32	$1.94*10^{-13}$
NTR	86	0.40	$1.37*10^{-4}$
QIMR	41	0.20	0.202
TwinsUK	NA	NA	NA
Meta-analysis	977	0.31	$4.27*10^{-23}$

Complete pairs included only. In case of > 2 sibs, the oldest two non-twin pair sibs were chosen for analysis. Mean spousal age was used for cut-off of 55 years of age.
*Limits of statistical software reached

Despite the evidence for the influence of environmental factors, neither the dominant variance component nor the sibling-shared variance component influenced the estimates of familial resemblance in those studies where the analysis could be performed (**Supplementary Table 2**) suggesting that any residual shared-environment or non-additive genetic effects are small. In the heritability analyses, three cohorts (ERF, NTR and QIMR) had nearly identical heritability estimates of 0.62 (**Table 3**). The highest heritability estimate was found in LLS ($h^2 = 0.86$). In the meta-analysis of all six cohorts, the heritability was estimated to be 0.70 (95% Confidence Interval (CI): 0.64-0.76; p-value = $2.31*10^{-111}$).
The correlations between parental and offspring telomere length as a function of gender showed an effect of parental gender, but not of offspring gender. Both father-offspring (r = 0.33; p-value = $5.66*10^{-4}$) and mother-offspring (r = 0.42; p-value = $5.01*10^{-5}$) telomere lengths were significantly correlated (**Table 4**). The mother-offspring correlation was significantly larger than the father-offspring correlation (p-value = 0.007). There was no difference in correlation structure between male and female offspring and their parents.
Finally, we investigated whether parental age plays a role in telomere length (**Table 5**). In each cohort the association between paternal age and telomere length went in the positive direction. After meta-analysis, we found a significant association between paternal age and telomere length (n = 5,127; beta = 0.005; p-value = $7.01*10^{-5}$). Maternal age was found to be significantly associated with telomere length after meta-analysis (n = 5,500; beta = 0.003; p-value = 0.012). This association was no longer significant when additionally adjusting for paternal age (beta = -0.003; p-value = 0.357). The paternal age association remained significant when additionally adjusting for maternal age (beta = 0.007; p-value = 0.011).

Table 3 | Heritability of telomere length.

	h2	Confidence interval	p-value
ERF	0.65	0.59-0.72	$1.45*10^{-61}$
GRAPHIC	0.73	0.68-0.78	$2.10*10^{-94}$
LLS	0.86	0.77-0.95	$6.53*10^{-57}$
NTR	0.62	0.59-0.65	$2.51*10^{-125}$
QIMR	0.62	0.57-0.67	$4.88*10^{-63}$
TwinsUK	0.74	0.72-0.75	$2.45*10^{-92}$
Meta-analysis	0.70	0.64-0.76	$2.31*10^{-111}$

DISCUSSION

In six independent family-based cohorts from Europe and Australia we investigated three aspects of familial relationships in telomere length, namely the heritability, mode of inheritance and parental age at birth of their children. We found a high heritability, a significant stronger maternal than paternal inheritance and a positive association with paternal age.

The meta-analysis estimate for heritability of telomere length in the six populations was 0.70. Heritability estimates from the separate studies were remarkably similar across populations, with three of them showing near identical estimates. The highest estimate was observed in the oldest participants (LLS; $h^2 = 0.86$) which may reflect a selection effect [33]. Even though only the offspring with their spouses of long living participants were selected from the LLS study, the selection of the parents for old age might have influenced the total telomere length distribution in their relatives.

Spousal correlations for telomere length have not been reported before [25]. Interestingly, we observed a significant correlation in telomere length between spouses. This was more pronounced in older spouses (mean age ≥ 55; r = 0.31) compared to younger spouses (mean age < 55; r = 0.20), suggesting a possible influence on telomere length due to living together for an extended amount of time. The hypothesis that resemblance in adults is induced by living together for extended periods of time does not need to be conflicting with the finding that modeling of the twin-family data did not detect shared-environmental influences. In a twin design shared environment is reflected in the portion of the phenotypic correlation in MZ and DZ twins that exceeds above the pattern expected if the similarity of twins was only a function of their genetic resemblance. By assuming that the common environment applies to those factors that are always shared 100% between twins, or siblings, the scope of this component for quantifying shared family environment tends to be limited to influences of the early environment, and ignores factors that are shared by individuals who share a household later in life. Another explanation could be an ascertainment effect. In older

Table 4 | Inheritance patterns: father-offspring and mother-offspring correlations in telomere length.

	ERF			GRAPHIC			NTR			QIMR			Meta-analysis		
Relationship*	**n**	**r**	**p-value**	**n**	**r**	**p-value**	**n**	**r**	**p-value**	**n**	**r**	**p-value**	**n**	**r**	**p-value**
Father-offspring	320	0.29	$1.79*10^{-7}$	501	0.34	$3.23*10^{-15}$	111	0.49	$5.32*10^{-8}$	312	0.28	$5.90*10^{-7}$	1,244	0.33	$1.15*10^{-32}$
Father-son	169	0.29	$1.59*10^{-4}$	380	0.39	$6.18*10^{-15}$	44	0.43	0.003	198	0.27	$1.34*10^{-4}$	791	0.34	$2.57*10^{-23}$
Father-daughter	203	0.33	$1.86*10^{-6}$	374	0.31	$1.31*10^{-9}$	75	0.49	$9.59*10^{-6}$	230	0.31	$2.42*10^{-6}$	882	0.33	$3.99*10^{-24}$
Mother-offspring	407	0.34	$1.32*10^{-12}$	502	0.48	$9.31*10^{-30}$	110	0.49	$6.57*10^{-8}$	369	0.39	$7.80*10^{-15}$	1,388	0.42	$3.60*10^{-61}$
Mother-son	193	0.35	$5.52*10^{-7}$	380	0.50	$7.53*10^{-26}$	43	0.41	0.007	234	0.31	$1.01*10^{-6}$	850	0.42	$5.06*10^{-37}$
Mother-daughter	285	0.37	$1.22*10^{-10}$	375	0.46	$2.67*10^{-21}$	75	0.54	$7.25*10^{-7}$	270	0.38	$6.94*10^{-11}$	1,005	0.42	$2.99*10^{-45}$

*If both a son and daughter were available for analysis, the oldest was used for parent-offspring, while each was used for parent-son or parent-daughter respectively.

Table 5 | Association of paternal and maternal age at birth with telomere length.

	Paternal age					Maternal age				
	n	**mean (sd)**	**beta**	**se**	**p-value**	**n**	**mean (sd)**	**beta**	**se**	**p-value**
ERF	560	26.7 (4.1)	0.0086	0.0038	0.024	578	25.0 (4.7)	0.0035	0.0028	0.211
GRAPHIC	983	28.3 (4.2)	0.0053	0.0030	0.077	984	26.4 (4.1)	-0.0011	0.0029	0.704
LLS	1,587	35.5 (5.9)	0.0029	0.0021	0.167	1,578	32.7 (5.4)	0.0030	0.0021	0.153
NTR	1,305	30.0 (4.6)	0.0080	0.0031	0.010	1,545	28.1 (4.2)	0.0064	0.0031	0.039
QIMR	692	32.2 (5.3)	0.0066	0.0048	0.169	815	29.5 (4.6)	0.0072	0.0048	0.134
Meta-analysis	5,127	NA	0.0053	0.0013	$7.01*10^{-5}$	5,500	NA	0.0032	0.0013	0.012

spouse pairs both individuals have survived for extended periods of time, and are therefore likely to have above average telomere length for their age. The age-correction performed might not completely control for this.

Mother-offspring correlations were significantly higher than the father-offspring correlations (p-value = 0.007), though both were statistically significant on their own. This did not change when additionally adjusting the father-offspring correlation for paternal age (data not shown). In 2004, Nawrot *et al* also showed a significant correlation between mother-offspring (n = 71) and father-daughter (n = 47), but not between father-son (n = 34) [23]. Recently, a paternal inheritance model has been gaining favor [8, 24-26]. The most recent study suggesting paternal inheritance from Nordfjall *et al* contained a total of 217 parent-offspring pairs [26]. In our study we included a total of 1,244 father-offspring pairs and 1,388 mother-offspring pairs and found both significant mother-offspring and father-offspring correlations for telomere length. Our findings are in contrast with those from other studies, as we found a significantly stronger mother-offspring than father-offspring correlation, which previously was attributed to an X-linked mechanism in telomere length determination [23]. A prime candidate gene for this X-linked mechanism is *DKC1* encoding dyskerin, which is important for the function of telomerase and has been found to cause congenital dyskeratosis which is characterized by short telomeres [43]. However, if an X-linked mechanism is in effect, a stronger father-daughter correlation would be expected, compared to the father-son correlation, which we do not observe. Other potential explanations for the larger mother-offspring correlation include mitochondrial DNA or other parent-specific 'imprinting'. Non-paternities were excluded as an explanation, based on genotype data.

Finally, we found significant evidence for a positive association of paternal age with telomere length supporting earlier reports [24, 27, 28]. At first glance, any positive association of telomere length with age seems counter-intuitive. However, telomere length is known to be maintained, and even lengthened, in sperm, despite both mitotic and meiotic divisions during the proliferation and differentiation of the cells [44]. In humans, this leads to a gain of approximately 71 base-pairs of telomere length per year [29]. Recently it was shown that the paternal age effect may be cumulative over generations [45]. Our finding is opposite of the recent publication that paternal age is the major determinant of novel mutations leading to autism [46]. Thus high paternal age may have both positive and negative effects on the health of the offspring, assuming that long telomeres in children are healthy.

In summary, we find a high and very consistent heritability estimate for telomere length. In addition, we find evidence for a maternal inheritance component and a positive association with paternal age.

REFERENCES

1. Blackburn, E.H. and J.G. Gall, A tandemly repeated sequence at the termini of the extrachromosomal ribosomal RNA genes in Tetrahymena. J Mol Biol, 1978. 120(1): p. 33-53.
2. Blackburn, E.H., Switching and signaling at the telomere. Cell, 2001. 106(6): p. 661-73.
3. Lindsey, J., et al., In vivo loss of telomeric repeats with age in humans. Mutat Res, 1991. 256(1): p. 45-8.
4. Slagboom, P.E., S. Droog, and D.I. Boomsma, Genetic determination of telomere size in humans: a twin study of three age groups. Am J Hum Genet, 1994. 55(5): p. 876-82.
5. Bakaysa, S.L., et al., Telomere length predicts survival independent of genetic influences. Aging Cell, 2007. 6(6): p. 769-74.
6. Cawthon, R.M., et al., Association between telomere length in blood and mortality in people aged 60 years or older. Lancet, 2003. 361(9355): p. 393-5.
7. Kimura, M., et al., Telomere length and mortality: a study of leukocytes in elderly Danish twins. Am J Epidemiol, 2008. 167(7): p. 799-806.
8. Njajou, O.T., et al., Telomere length is paternally inherited and is associated with parental lifespan. Proc Natl Acad Sci U S A, 2007. 104(29): p. 12135-9.
9. Bischoff, C., et al., No association between telomere length and survival among the elderly and oldest old. Epidemiology, 2006. 17(2): p. 190-4.
10. Martin-Ruiz, C.M., et al., Telomere length in white blood cells is not associated with morbidity or mortality in the oldest old: a population-based study. Aging Cell, 2005. 4(6): p. 287-90.
11. Atzmon, G., et al., Evolution in health and medicine Sackler colloquium: Genetic variation in human telomerase is associated with telomere length in Ashkenazi centenarians. Proc Natl Acad Sci U S A, 2010. 107 Suppl 1: p. 1710-7.
12. Codd, V., et al., Common variants near TERC are associated with mean telomere length. Nat Genet, 2010. 42(3): p. 197-9.
13. Levy, D., et al., Genome-wide association identifies OBFC1 as a locus involved in human leukocyte telomere biology. Proc Natl Acad Sci U S A, 2010. 107(20): p. 9293-8.
14. Mirabello, L., et al., The association of telomere length and genetic variation in telomere biology genes. Hum Mutat, 2010. 31(9): p. 1050-8.
15. Rafnar, T., et al., Sequence variants at the TERT-CLPTM1L locus associate with many cancer types. Nat Genet, 2009. 41(2): p. 221-7.
16. Shen, Q., et al., Common variants near TERC are associated with leukocyte telomere length in the Chinese Han population. Eur J Hum Genet, 2011. 19(6): p. 721-3.
17. Soerensen, M., et al., Genetic variation in TERT and TERC and human leukocyte telomere length and longevity: a cross-sectional and longitudinal analysis. Aging Cell, 2011.
18. Deelen, J., et al., Gene set analysis of GWAS data for human longevity highlights the relevance of the insulin/IGF-1 signaling and telomere maintenance pathways. Age (Dordr), 2011.
19. Bischoff, C., et al., The heritability of telomere length among the elderly and oldest-old. Twin Res Hum Genet, 2005. 8(5): p. 433-9.
20. Andrew, T., et al., Mapping genetic loci that determine leukocyte telomere length in a large sample of unselected female sibling pairs. Am J Hum Genet, 2006. 78(3): p. 480-6.
21. Vasa-Nicotera, M., et al., Mapping of a major locus that determines telomere length in humans. Am J Hum Genet, 2005. 76(1): p. 147-51.
22. Horn, T., B.C. Robertson, and N.J. Gemmell, The use of telomere length in ecology and evolutionary biology. Heredity (Edinb), 2010. 105(6): p. 497-506.
23. Nawrot, T.S., et al., Telomere length and possible link to X chromosome. Lancet, 2004. 363(9408): p. 507-10.

24. Kimura, M., et al., Offspring's leukocyte telomere length, paternal age, and telomere elongation in sperm. PLoS Genet, 2008. 4(2): p. e37.
25. Nordfjall, K., et al., Telomere length and heredity: Indications of paternal inheritance. Proc Natl Acad Sci U S A, 2005. 102(45): p. 16374-8.
26. Nordfjall, K., et al., Large-scale parent-child comparison confirms a strong paternal influence on telomere length. Eur J Hum Genet, 2010. 18(3): p. 385-9.
27. De Meyer, T., et al., Paternal age at birth is an important determinant of offspring telomere length. Hum Mol Genet, 2007. 16(24): p. 3097-102.
28. Unryn, B.M., L.S. Cook, and K.T. Riabowol, Paternal age is positively linked to telomere length of children. Aging Cell, 2005. 4(2): p. 97-101.
29. Allsopp, R.C., et al., Telomere length predicts replicative capacity of human fibroblasts. Proc Natl Acad Sci U S A, 1992. 89(21): p. 10114-8.
30. Aulchenko, Y.S., et al., Linkage disequilibrium in young genetically isolated Dutch population. Eur J Hum Genet, 2004. 12(7): p. 527-34.
31. Pardo, L.M., et al., The effect of genetic drift in a young genetically isolated population. Ann Hum Genet, 2005. 69(Pt 3): p. 288-95.
32. Tobin, M.D., et al., Common variants in genes underlying monogenic hypertension and hypotension and blood pressure in the general population. Hypertension, 2008. 51(6): p. 1658-64.
33. Schoenmaker, M., et al., Evidence of genetic enrichment for exceptional survival using a family approach: the Leiden Longevity Study. Eur J Hum Genet, 2006. 14(1): p. 79-84.
34. Boomsma, D.I., et al., Netherlands Twin Register: from twins to twin families. Twin Res Hum Genet, 2006. 9(6): p. 849-57.
35. Willemsen, G., et al., The Netherlands Twin Register biobank: a resource for genetic epidemiological studies. Twin Res Hum Genet, 2010. 13(3): p. 231-45.
36. Wright, M.J. and N.G. Martin, Brisbane adolescent twin study: outline of study methods and research projects. Aust J Psychol 2004. 56(2): p. 65-78.
37. Medland, S.E., et al., Common variants in the trichohyalin gene are associated with straight hair in Europeans. Am J Hum Genet, 2009. 85(5): p. 750-5.
38. Moayyeri, A., et al., Cohort Profile: TwinsUK and Healthy Ageing Twin Study. Int J Epidemiol, 2012.
39. Cawthon, R.M., Telomere measurement by quantitative PCR. Nucleic Acids Res, 2002. 30(10): p. e47.
40. R Development Core Team, R: A language and environment for statistical computing. 2010, R foundation for Statistical Computing: Vienna, Austria.
41. Pilia, G., et al., Heritability of cardiovascular and personality traits in 6,148 Sardinians. PLoS Genet, 2006. 2(8): p. e132.
42. Lumley, T., rmeta: Meta-analysis. 2009.
43. Mitchell, J.R., E. Wood, and K. Collins, A telomerase component is defective in the human disease dyskeratosis congenita. Nature, 1999. 402(6761): p. 551-5.
44. Achi, M.V., N. Ravindranath, and M. Dym, Telomere length in male germ cells is inversely correlated with telomerase activity. Biol Reprod, 2000. 63(2): p. 591-8.
45. Eisenberg, D.T., M.G. Hayes, and C.W. Kuzawa, Delayed paternal age of reproduction in humans is associated with longer telomeres across two generations of descendants. Proc Natl Acad Sci U S A, 2012. 109(26): p. 10251-6.
46. Kong, A., et al., Rate of de novo mutations and the importance of father's age to disease risk. Nature, 2012. 488(7412): p. 471-5.

Chapter 4.2

Telomere length variation reduces with age: evidence of survivor effect

Linda Broer, Najaf Amin, Veryan Codd, Ben A. Oostra, Nilesh J. Samani, Cornelia M. van Duijn

ABSTRACT

Telomere shortening has been associated with aging, disease and mortality but findings have not been consistent. One of the reasons for inconsistencies may be a reduced variance in telomere length in the elderly: a survivor effect. A previous study attributed this survivor effect to a selection against those with short and with long telomeres.

Using a multigenerational family study in the Erasmus Rucphen Family (ERF) study we aimed to explore the variability of telomere length in different age-categories. Building upon the high heritability of telomere length, we next used the telomere length of the (grand)children of the oldest-old as a proxy for the telomere length of the oldest-old at younger ages.

We found a significant reduction in variance in telomere length from young adulthood to old age (p-value = 4.23×10^{-6}). The decline in telomere length variance was stronger in men (69.9%) compared to women (59.7%). The mean telomere length of the (grand)children of the oldest-old was significantly longer compared to others in the same age-range (p-value = 2.60×10^{-4}). In only one out of 13 families telomere length of the (grand)children was shorter than average early in life.

Our study shows a reduction in the variance of telomere length over age. Further our study suggests that those with long telomere length survive longest.

INTRODUCTION

Telomeres are specialized repetitive DNA structures located at the terminal end of chromosomes and protect against spontaneous DNA damage and thus preserve genomic integrity [1, 2]. Telomeres progressively shorten with each cell division [3], until a critically shortened length is achieved upon which the cell enters replicative senescence [4]. Telomere shortening has been associated with organismal aging, disease and mortality [5-9], though not consistently [10-15]. Further a reduction in variability in telomere length has been observed with aging [16]. This may be the result of selective mortality in those with short telomeres at birth but perhaps also those with long telomeres early in life. It has been speculated that the latter have an increased risk of cancer [16]. Most of the studies investigating the association between telomere length and mortality have been performed in the elderly where such selection could have already occurred causing a survivor effect [17].

Halaschek-Wiener *et al* [16] attributed the survivor effect to the risk associated with both long and short telomeres, suggesting there may be an optimal telomere length that is protective for healthy aging. In order to study survivor effects in a longitudinal study, a long follow-up is required assessing telomere length early and late in life. However, given the high heritability (h^2 ~ 0.70) of telomere length an alternative design is a multigenerational family study. Such a study can mimic a longitudinal study using the telomere lengths seen in the (grand)children as a proxy for the telomere length of an individual early in life [18].

In the current study we explore the variation in telomere length in various generations in a multigenerational Erasmus Rucphen Family (ERF) study using the telomere length of the (grand)children of the oldest-old as a proxy for the telomere length of the oldest-old (85+ years of age) at early life. In this way we aim to investigate the telomere length early in life that is related to healthy aging.

METHODS

The Erasmus Rucphen Family study (ERF) is a cross-sectional family-based study including 3,000 living descendants of 22 couples around 1850-1900 [19, 20]. All participants are screened for many quantitative traits related to neurodegenerative, cardiovascular and endocrine diseases. For all participants the genealogical relationships are available from church records. These relationships were validated using genetic data from genome-wide association studies. The study protocol was approved by the medical ethics board of the Erasmus MC Rotterdam, the Netherlands.

Mean leukocyte telomere length was measured using a quantitative PCR-based technique as previously described [21, 22]. This method expresses telomere length as a ratio (T/S) of telomere repeat length (T) to copy number of a single copy gene, 36B4(S), within each

sample. Samples were quantified relative to a calibrator sample used on each run (DNA from the K562 cell line) [22]. Mean inter-run CV was calculated to be 3.48%.

Heritability analysis was performed in SOLAR (version 2.05, http://solar.sfbrgenetics.org), adjusting for age and gender using an additive genetic variance component and random environment variance component. Association of age and gender with telomere length was determined in the same model. Survival analysis in SPSS version 17 was performed to determine the association of telomere length with all-cause mortality. All-cause mortality, obtained from obituaries, was complete until February 2012.

We simulated the effect of four different selection scenarios on the reduction of telomere length variance, namely A) no selection; B) selection against short telomeres; C) selection against long telomeres; D) selection against both short and long telomeres. We simulated individuals with the longest and shortest telomere length. For the rate of telomere length decline with age we took -0.1 T/S ratio for every 10 years, which is consistent across populations [23]. For the longest and shortest telomeres at young age we chose 3.3 and 1.3 T/S ratio respectively, which is consistent with the ERF data. For mortality we applied a two-fold increase in mortality in the group selected upon compared to the non-selected group.

In order to investigate the reduction in telomere length variation over age empirically, we created 10-year age-categories starting with the youngest age group included at 18 years. Using the Levene's test in SPSS version 20, we compared the variance in each age-category with the youngest age-category, as well as between consecutive age-categories. We calculated the rate of decline in telomere length variance as a percentage of the original variance. Next, we extrapolated the telomere length of the oldest-old using the telomere lengths of their (grand)children. We compared the mean telomere length of the (grand) children to all others in the age-range using an independent t-test in SPSS version 20.

RESULTS

General characteristics are found in **Table 1**. As expected telomere length is decreasing with age (p-value = $1.56*10^{-102}$) and women have longer telomere length (p-value = $2.54*10^{-7}$). In our family study, the heritability of telomere length was 65% and highly significant (p-value = $1.45*10^{-61}$).

Table 1 | General characteristics of the study population.

	mean (sd)	beta	se	p-value
telomere length	1.78 (0.36)	na	na	na
age	49.8 (14.9)	-0.0086	0.0004	$1.56*10^{-102}$
sex (% female)	55.6	0.0634	0.0123	$2.54*10^{-7}$

Regression coefficient (beta), standard error (se) and p-value from linear regression model with telomere length as outcome.

Telomere length was strongly associated with all-cause mortality (Hazard Ratio (HR) = 0.21; p-value = $1.52*10^{-13}$; **Table 2**). When adjusting for age at baseline the mortality is still 1.5 times reduced but is no longer statistically significant (HR = 0.69; p-value = 0.125) due to the strong correlation between telomere length and chronological age.

Table 2 | Association of telomere length with all-cause mortality.

				overall			adjusted for age		
	n	n_death	follow_up	HR	se	p-value	HR	se	p-value
total	2514	204	7.93 (1.44)	0.21	0.21	$1.52*10^{-13}$	0.69	0.24	0.125
men	1107	109	7.79 (1.54)	0.20	0.30	$5.4*10^{-8}$	0.62	0.34	0.158
women	1407	95	8.04 (1.35)	0.24	0.31	$3.81*10^{-6}$	0.89	0.35	0.736

HR = hazard ratio; n = total sample size; n death = number of deceased; se = standard error

Figure 1 visualizes the effect of selection on the variance in the population under four different selection scenarios, namely A) no selection; B) selection against short telomeres; C) selection against long telomeres; D) selection against both short and long telomeres. If no selection against telomere length occurs the variance in telomere length will be similar at young and old age (**Figure 1A**). However, if selection occurs a reduction in variance will be observed in all scenarios (**Figure 1B-D**). These scenarios (**B-D**) are difficult to distinguish. Independent of selection, the telomere length decreases with age according to the increasing mitotic divisions.

Figure 2 shows telomere length of the entire ERF population for both genders plotted against age. The graph shows a decline in the variance of telomere length as reflected by the funnel shape of the scatter plot from age 20 years to age 90 years. Statistical analysis using Levene's test for the differences in variance reveals a significant decline in the variance (p-value = $4.23*10^{-6}$) when categorized in 10-year age intervals (**Table 3**). The decline in telomere length variance was 69.9% in men compared to 59.7% in women. The first two age-categories (18-28 years and 28-38 years) did not show a significant difference in telomere length variance in the total cohort (p-value = 0.214) nor in the gender specific analyses.

We next tested whether there was evidence for mortality in those with long telomeres, short telomeres or both groups, assuming that this information can be deduced from the (grand) children given the high heritability. **Figure 3** shows the individuals that reached 85+ years (oldest-old) and connects their telomere length with that of the mean telomere length of their children and grandchildren. Telomere length is shown on the vertical axis. With the exception of one family (family 11; one child and one grandchild), the (grand)children of the oldest-old have longer telomeres than the oldest-old themselves. Mean telomere

length of the grandchildren was significantly longer compared to others in the same age-range (p-value = 0.001; **Table 4**), suggesting that the oldest-old were most likely born with telomeres that were longer on average than the general population.

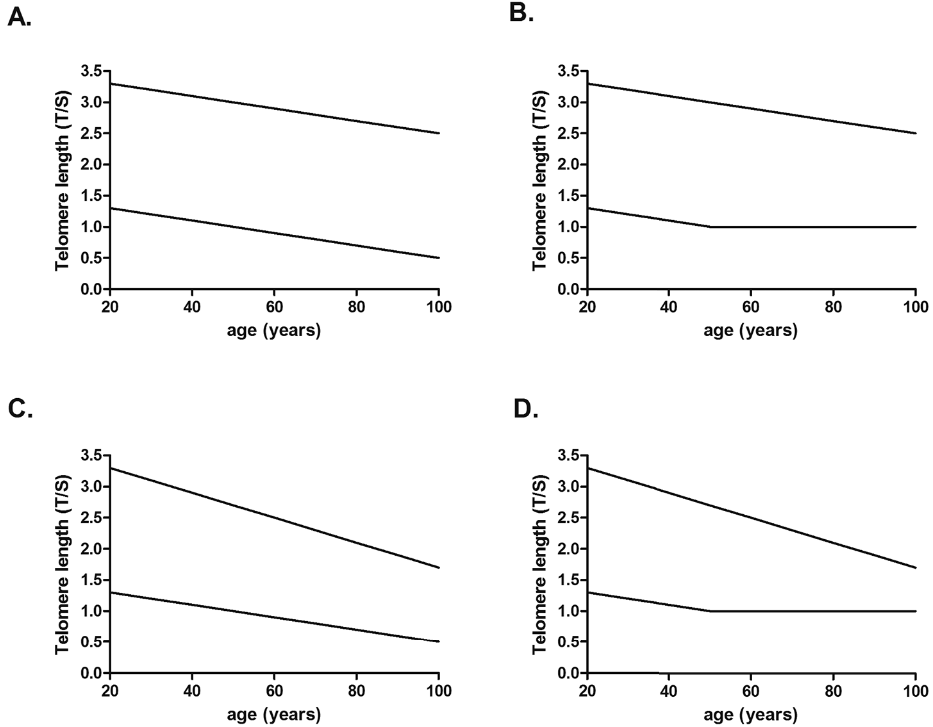


Figure 1 | Graphical representation of the relation between telomere length and age using different selection scenarios.

A. No selection
B. Selection against short telomeres only
C. Selection against long telomeres only
D. Selection against both short and long telomeres

Table 3 | Variance of telomere length by age category.

	Total				Men				Women			
age-category	**n**	**mean**	**variance**	**p-value***	**n**	**mean**	**variance**	**p-value***	**n**	**mean**	**variance**	**p-value***
18-38	646	1.938	0.150	reference	279	1.890	0.153	reference	367	1.975	0.144	reference
38-48	600	1.841	0.124	0.037	250	1.835	0.124	0.094	350	1.845	0.123	0.296
48-58	657	1.754	0.096	$3.10*10^{-9}$	306	1.702	0.086	$9.32*10^{-8}$	351	1.800	0.100	0.001
58-68	568	1.665	0.102	$1.46*10^{-5}$	261	1.612	0.095	$4.61*10^{-5}$	307	1.710	0.104	0.013
68-78	228	1.601	0.087	$7.08*10^{-6}$	109	1.591	0.091	0.002	119	1.611	0.083	0.002
78-88	70	1.522	0.053	$4.23*10^{-6}$	25	1.509	0.046	0.001	45	1.529	0.058	0.001

n = number of persons in category; *P-value from Levene's test for equal variances compared to the youngest age-category

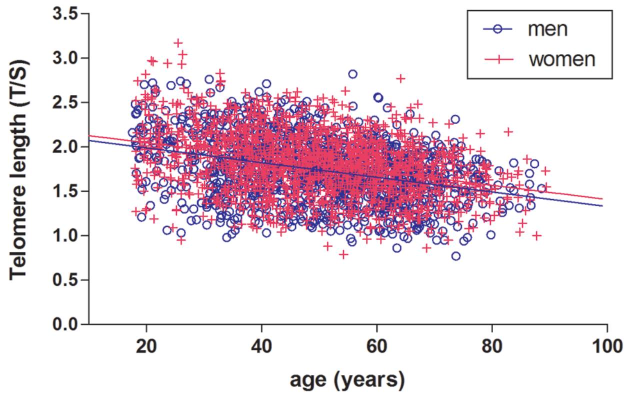


Figure 2 | Association of age and gender with telomere length.

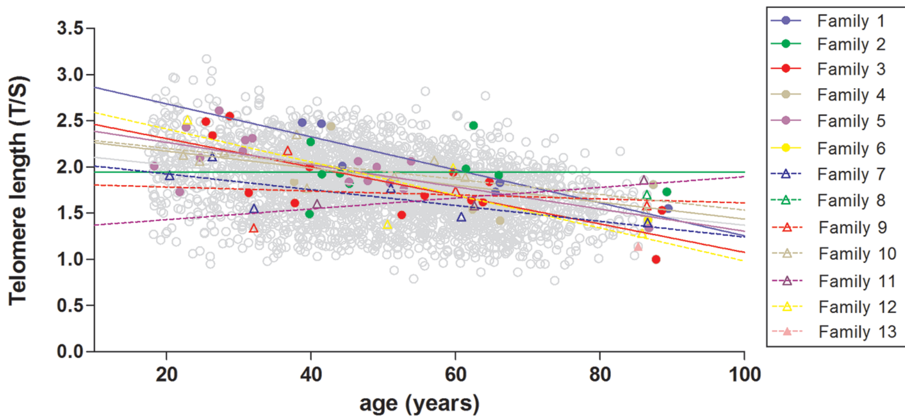


Figure 3 | The relation of the telomeres in the oldest-old to their (grand)children.

Table 4 | Telomere length of (grand)children of the oldest-old compared to the population.

generation	age range	mean (sd) (grand)children	mean (sd) population	p-value
grandchildren	18-48	2.08 (0.34)	1.89 (0.37)	0.001
children	38-68	1.81 (0.23)	1.75 (0.33)	0.220

sd = standard deviation

DISCUSSION

In this study, we observed a significant reduction in telomere length variability in the population from young adulthood until late in life. Additionally, we observed that the (grand)children of the oldest-old have long telomeres for their age, suggesting that the oldest-old themselves were also born with telomeres that were longer on average than the general population.

Our study confirmed the previously observed reduction in telomere length variability between mid-life and oldest-old as reported by Halaschek-Wiener *et al* [16]. Further, our study shows that this strong reduction in telomere length variability can be measured after the age of 38 years, implying that selection on telomere length also starts early and can be measured already around the age of 40 years, the age at which the reduction in variance was statistically significant. Another implication of our study is that telomere length at old age are not representative for those at the start of life. This could explain why almost half of previously published studies investigating the association between telomere length and mortality did not find a significant association [5-15]. In studies conducted late in life not only the telomere length is not representative anymore for early life but the variance is significantly reduced making it statistically difficult to identify associations. In our study telomere length is associated with a 1.5 fold increase in mortality, though when adjusting for age at baseline the association lost its statistical significance. Telomere length and chronological age are extremely strongly associated which explains the loss of significance. A single measurement of telomere length after age 40 seems to add little information over chronological age in predicting mortality. As has been demonstrated previously in birds, it is more likely that the differences in telomere length measured at early and middle age will be associated with mortality [24]. Indirectly our finding of the reduced variance points in this direction as this suggests that there is differential mortality based on telomere length early in life.

The most important finding of this study is that the grandchildren of the oldest-old had telomeres longer than the average for their age, which strongly suggests that the oldest-old started life with long telomeres. This finding is at odds with the hypothesis of optimal telomeres as suggested by Halaschek-Wiener *et al* [16]. Short telomeres in the (grand) children was observed in only one out of 13 families of which the grandparents reached old age (> 85 years), while all other families had long telomeres. Our data therefore suggests that predominantly short telomeres are selected against and that critically short telomeres induce replicative senescence and are associated to mortality [4, 25, 26]. A limitation of our study is the small sample size in the oldest-old, resulting in relatively few families to be investigated. Further, we did not have information on the spouses of the oldest-old and their families, which could explain the observation in the family in which the grand(children) had shorter telomeres than the oldest-old.

4

In our study, we further observed a larger rate of decline in telomere length variation in men compared to women (69.9% compared to 59.7%). Differences in telomere attrition rate between genders have been described [27]. In men, a significantly stronger attrition has been seen, in line with our findings [27].

To summarize, we found a strong association of telomere length with mortality. Our study further shows that telomere length variability in the population strongly reduces with age, which is most likely explained by a survivor effect. Our study suggests that most of the oldest-old were born with telomeres that were longer than others in the age category, indicating that those with long telomeres have a better probability of survival. These findings ask for multiple assessments of telomere length to study the relation of telomere length with mortality.

REFERENCES

1. Blackburn, E.H. and J.G. Gall, A tandemly repeated sequence at the termini of the extrachromosomal ribosomal RNA genes in Tetrahymena. J Mol Biol, 1978. 120(1): p. 33-53.
2. Blackburn, E.H., C.W. Greider, and J.W. Szostak, Telomeres and telomerase: the path from maize, Tetrahymena and yeast to human cancer and aging. Nat Med, 2006. 12(10): p. 1133-8.
3. Lindsey, J., et al., In vivo loss of telomeric repeats with age in humans. Mutat Res, 1991. 256(1): p. 45-8.
4. Campisi, J. and F. d'Adda di Fagagna, Cellular senescence: when bad things happen to good cells. Nat Rev Mol Cell Biol, 2007. 8(9): p. 729-40.
5. Cawthon, R.M., et al., Association between telomere length in blood and mortality in people aged 60 years or older. Lancet, 2003. 361(9355): p. 393-5.
6. Bakaysa, S.L., et al., Telomere length predicts survival independent of genetic influences. Aging Cell, 2007. 6(6): p. 769-74.
7. Kimura, M., et al., Telomere length and mortality: a study of leukocytes in elderly Danish twins. Am J Epidemiol, 2008. 167(7): p. 799-806.
8. Ehrlenbach, S., et al., Influences on the reduction of relative telomere length over 10 years in the population-based Bruneck Study: introduction of a well-controlled high-throughput assay. Int J Epidemiol, 2009. 38(6): p. 1725-34.
9. Epel, E.S., et al., The rate of leukocyte telomere shortening predicts mortality from cardiovascular disease in elderly men. Aging (Albany NY), 2009. 1(1): p. 81-8.
10. Martin-Ruiz, C.M., et al., Telomere length in white blood cells is not associated with morbidity or mortality in the oldest old: a population-based study. Aging Cell, 2005. 4(6): p. 287-90.
11. Bischoff, C., et al., No association between telomere length and survival among the elderly and oldest old. Epidemiology, 2006. 17(2): p. 190-4.
12. Harris, S.E., et al., The association between telomere length, physical health, cognitive ageing, and mortality in non-demented older people. Neurosci Lett, 2006. 406(3): p. 260-4.
13. Fitzpatrick, A.L., et al., Leukocyte telomere length and cardiovascular disease in the cardiovascular health study. Am J Epidemiol, 2007. 165(1): p. 14-21.
14. Njajou, O.T., et al., Association between telomere length, specific causes of death, and years of healthy life in health, aging, and body composition, a population-based cohort study. J Gerontol A Biol Sci Med Sci, 2009. 64(8): p. 860-4.
15. Zekry, D., et al., Telomere length, comorbidity, functional, nutritional and cognitive status as predictors of 5 years post hospital discharge survival in the oldest old. J Nutr Health Aging, 2012. 16(3): p. 225-30.
16. Halaschek-Wiener, J., et al., Reduced telomere length variation in healthy oldest old. Mech Ageing Dev, 2008. 129(11): p. 638-41.
17. Mather, K.A., et al., Is telomere length a biomarker of aging? A review. J Gerontol A Biol Sci Med Sci, 2011. 66(2): p. 202-13.
18. Galton, F., Regression towards mediocrity in hereditary stature. Journal of the anthropological institute, 1886. 15: p. 246-263.
19. Aulchenko, Y.S., et al., Linkage disequilibrium in young genetically isolated Dutch population. Eur J Hum Genet, 2004. 12(7): p. 527-34.
20. Pardo, L.M., et al., The effect of genetic drift in a young genetically isolated population. Ann Hum Genet, 2005. 69(Pt 3): p. 288-95.
21. Cawthon, R.M., Telomere measurement by quantitative PCR. Nucleic Acids Res, 2002. 30(10): p. e47.
22. Codd, V., et al., Common variants near TERC are associated with mean telomere length. Nat Genet, 2010. 42(3): p. 197-9.
23. Broer, L., et al., Meta-analysis of telomere length in 19 713 subjects reveals high heritability, stronger maternal inheritance and a paternal age effect. Eur J Hum Genet, 2013.

24. Heidinger, B.J., et al., Telomere length in early life predicts lifespan. Proc Natl Acad Sci U S A, 2012. 109(5): p. 1743-8.
25. Abdallah, P., et al., A two-step model for senescence triggered by a single critically short telomere. Nat Cell Biol, 2009. 11(8): p. 988-93.
26. Blackburn, E.H., Switching and signaling at the telomere. Cell, 2001. 106(6): p. 661-73.
27. Aviv, A., et al., The longevity gender gap: are telomeres the explanation? Sci Aging Knowledge Environ, 2005. 2005(23): p. pe16.

Chapter 4.3

Association of height and telomere length in the CHANCES consortium

Linda Broer, Aysel Mueezzinler, Geja J. Hageman, Najaf Amin, Eva Albrecht, Veryan Codd, Erik J. Giltay, Alessandro Martino, Paolo Boffetta, Antonia Trichopoulou, Ben A. Oostra, Aida K. Zaineddin, Nilesh J. Samani, Daan Kromhout, Hermann Brenner, Cornelia M. van Duijn

ABSTRACT

Telomere length is a trigger of cellular senescence and is considered to be a marker of biological age. A male-specific inverse association between telomere length and height has been observed across species. Here we study the association between telomere length and height in 3,625 individuals in the Consortium on Health and Aging Network of Cohorts in Europe and the United States (CHANCES).

Partial correlation analysis of telomere length and height, adjusting for age and sex, was performed in males, females and the combined sample. Additionally Mendelian randomization analysis was performed to determine causality.

We find a small but significant inverse association between telomere length and height in the total cohort (r = -0.04; p-value = 0.012). In the gender-specific analysis only men showed a statistically significant correlation (r = -0.08; p-value = $5.8*10^{-4}$), while women showed no statistical evidence for correlation (r = -0.02; p-value = 0.302). Mendelian randomization analysis are compatible with the hypothesis that differences in height cause differences in telomere length, as several genes involved in height are associated with telomere length in the inverse direction.

Our study in three populations suggests that telomere length is inversely associated to height.

INTRODUCTION

Telomeres are tandem repeats present at the ends of the chromosomes, which protect the terminal ends of the chromosomes by preventing chromosomal fusion and degradation of coding DNA [1] . Telomeres shorten with each replication until a critical length is achieved, thus achieving the point of replicative senescence [2-5]. The enzyme telomerase plays a key role in telomere restoration by adding base pairs *de novo* [6]. However this enzyme is down-regulated in humans [7]. Short telomere size and high attrition rate have previously been associated with increased mortality [8-12], though not consistently in all studies [13-18]. On average, women have longer telomeres than men when matched for age [9, 19], although mean telomere length at birth does not show any sex-related differences [20]. This suggests a slower telomeric attrition rate in women [21] and corroborates with the finding that women outlive men in every age group [22].

As a larger body size requires more cellular replications both for growth and maintenance, there is evidence that increased body size results in higher telomeric attrition rate [21], despite the fact that telomerase activation also increases with growth [23]. In line with this observation, in virtually all vertebrates studied smaller body size is associated with longevity [24-26]. This phenomenon has been attributed to the Insulin/Insulin-like Growth Factor (IGF)-1 signaling (IIS) pathway [27].

Sex is an important modifier of the association between height and life expectancy. Sex-biased mortality is more pronounced in species that show sexual size dimorphism in body size [28]. Females show greater telomerase activity and this finding is an evolutionary conserved mechanism even in species in which the body size of the female is larger than that of males [29, 30], suggesting sex-dependent telomere attrition is independent of body size [29]. In a wide variety of animal species (reptiles and birds), the association between body size and telomere length is seen predominantly in males, even in species in which males are larger than females such as dunlins [31, 32]. In humans, to our knowledge, the association between height and telomere length has never been investigated.

In this study, we aim to investigate the association between telomere length and height in 3,625 individuals in the Consortium on Health and Aging Network of Cohorts in Europe and the United States (CHANCES) overall and stratified by sex. Since both height and telomere length are highly heritable, we used the principle of Mendelian randomization analyses to evaluate the evidence for a causal relationship and the direction of the relationship between height and telomere length. That is with Mendelian randomization we aimed to answer the question whether height through growth influences telomere length or does telomere length determine height?

METHODS

Participating cohorts

The Erasmus Rucphen Family study (**ERF**) is a cross-sectional family-based study including 3,000 living descendants of 22 couples baptized in the community church around 1850-1900. The participants are not selected on any disease or other outcome. Details about the genealogy of the population have been described elsewhere [33, 34]. The study protocol was approved by the medical ethics board of the Erasmus MC Rotterdam, the Netherlands. Mean leukocyte telomere length was measured using a widely reported quantitative PCR-based technique as previously described [35, 36]. This method expresses telomere length as a ratio (T/S) of telomere repeat length (T) to copy number of a single copy gene, 36B4 (S), within each sample. Samples were quantified relative to a calibrator sample used on each run (DNA from the K562 cell line) [36]. Mean inter-run CV was calculated to be 3.48%. Height was measured with the participants standing upright without wearing shoes.

The **ESTHER** study (Epidemiological Study on the Chances of Prevention, Early Recognition, and Optimized Treatment of Chronic Diseases in the Older Population) participants were aged 50-74 years and were recruited between July 2000 and December 2002 during a general health check-up in Saarland, south-west Germany, by their general practitioners. Information on age, sex, self-reported weight and height were obtained by detailed questionnaires in a standardized manner. Serum samples were obtained from all participants from peripheral blood. Genomic DNA was extracted by high salt method and stored at -80°C. The ESTHER study is approved by the ethics committees of the medical faculty of the University of Heidelberg and of the medical board of the state of Saarland. Telomere length was measured by quantitative PCR (qPCR) [35]. In this study 36B4 was used as the single copy gene. Two quality-control samples were inserted into each PCR plate in order to assess the coefficients of inter- and intra-plate variability. All samples were measured in triplicates and the mean of the two closest measurements was used. Calculation of telomere length was based on comparison of the distinct cycle number determined by threshold values (C_t) at a constant fluorescence level. PCRs were performed on the Prism 7900 Sequence Detection System (Applied Biosystems, Foster City, USA).

The **Zutphen Elderly Study** consists of men born between 1900 and 1920. The study started in 1960 as the Dutch contribution to the Seven Countries Study [37]. The Zutphen Elderly Study consisted of 553 men in 1993, of whom 390 participated, and 345 had a blood sample. However, only 203 (59%) buffy coats could be used for DNA isolation because 142 samples were unavailable. Written informed consent was obtained from all participants. The Zutphen Elderly Study (in 1993 and in 2000) was approved by the Medical Ethics Committee of The Netherlands Organization for Applied Scientific Research (TNO). Telomere length was determined by quantitative PCR as described previously [35]. Two master mixes were prepared: one with telomere primers and one with human b-globin primers (1× IQ SYBRgreen supermix from BioRad-Laboratories BV, Veenendaal, The

Netherlands). To confirm reproducibility of the applied method, we repeated for 20 samples the telomere length measurement by using an additional reference gene, acidic ribosomal phosphoprotein PO (36B4), 300 nM forward primer (5'-CAGCAAGTGGGAAGGTGTAATCC-3'), and 500 nM reverse primer (5'-CCCATTCTATCATCAACGGGTACAA-3'). Each sample was run in triplicates. The coefficient of variation for the triplicates of the telomere reaction was 4.11% and for the reference gene 3.03%.

Statistical analysis

In order to standardize telomere length measurements across cohorts a Z-transformation was applied. Samples deviating more than 4 standard deviations from the mean for either telomere length or height were removed. We performed a partial overall correlation analysis including both men and women adjusting for age and sex, and if necessary for family relationships (ERF) and also sex-specific correlation analyses adjusted for age. Meta-analysis was performed using R-package metacor, implementing DerSimonian-Laird (DSL) random-effect meta-analytical approach with correlation coefficients as effect sizes [38, 39]. Between-study heterogeneity was assessed using rmeta [38]. We used the Bonferroni corrected p-value threshold for significance of 0.017 (= 0.05/3).

Association analysis does not establish causality. A way to defer directionality of association is to use genetic data in a so-called Mendelian randomization analysis [40]. As genotypes are randomly assigned when passed from parent to offspring during meiosis, the genotype distribution in the population is unrelated to potential confounders. Mendelian randomization assumes that the genotype can only affect the outcome indirectly via its effect on the exposure of interest [40]. If the assumption is met, Mendelian randomization can be seen as a "natural" randomized controlled trial [40]. This implies that if telomere length causally influences height, one would expect that genetic factors, for instance, Single Nucleotide Polymorphisms (SNPs) associated with telomere length will also be associated with height. We have tested three potential scenarios, namely 1) differences in telomere length cause differences in height and thus genes involved in telomere length should associate to height (**Figure 1A**); 2) differences in height cause differences in telomere length and thus genes involved in height should associate to telomere length (**Figure 1B**); 3) an untested third factor causes differences in both height and telomere length and the association between height and telomere length is explained as confounding. In order to test the first two scenarios, we use published genome-wide association studies (GWAS) [41, 42]. Assuming that the genes identified to date capture a sufficient part of the variance in height and telomere length, the absence of association supporting the first two scenarios suggests that the third scenario is likely to be the underlying biological reason for finding an association or that the association is spurious. To test for a joint-effect of all known SNPs from published GWAS on the trait to be explained (e.g. telomere length SNPs on height) we used the betas from published GWAS as weights in a meta-analysis of all SNPs in rmeta [38].

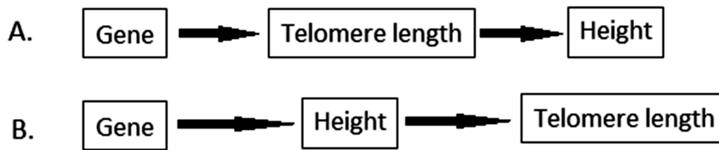


Figure 1 | Graphical representation of Mendelian Randomization.
A. Telomere length is causally associated to height
B. Height is causally associated to telomere length

As genes associated to height or telomere length are involved in various pathways, it may be speculated that only a limited number of genes from a distinct pathway are associated. The biological function of genes of the Mendelian randomization experiment were therefore evaluated using Ingenuity Pathways Analysis (Ingenuity Systems, http://www.ingenuity.com). This software analyses data in the context of known biological response, regulatory networks and pathways. Fisher's exact test was used to calculate a p-value determining the probability that each biological function or disease assigned to that dataset was lower than due to chance.

RESULTS

The general characteristics of the study populations are depicted in **Table 1** stratified for sex. The mean age of the participants ranged from 48 years of age to nearly 79 years of age. In ERF and ESTHER approximately 40% of the participants were male, while the Zutphen Elderly Study consisted of 100% male participants.

The results of the partial correlation analysis of telomere length and height are presented in **Table 2**. In the overall meta-analysis we observed a significant inverse correlation between telomere length and height ($r = -0.04$; p-value = 0.01). The sex-stratified meta-analysis revealed a stronger and more significant inverse correlation between height and telomere length in men ($r = -0.08$; p-value = $5.8*10^{-4}$), and a non-significant correlation in women ($r = -0.02$; p-value = 0.30). There was no evidence for between-study heterogeneity in the analyses (lowest p-value for heterogeneity overall or in the sex-stratified analysis p-value > 0.51).

Table 1 | General characteristics of the study populations.

	ERF						ESTHER						Zutphen					
	men			women			men			women			men			women		
	n	mean	sd	n	mean	sd	n	mean	sd	n	mean	sd	n	mean	sd	n	mean	sd
age (years)	1,083	48.43	14.24	1,364	47.63	14.41	421	61.43	6.11	571	61.09	6.33	203	78.63	4.28	NA	NA	NA
telomere length (T/S ratio)	1,083	1.73	0.36	1,364	1.82	0.36	421	0.85	0.33	571	0.88	0.29	203	5.06	0.47	NA	NA	NA
height (cm)	1,083	174.67	7.15	1,364	161.93	6.46	413	174.24	6.69	562	162.07	5.80	203	172.22	6.74	NA	NA	NA

Table 2 | Correlation between telomere length and height in the three populations.

	Total				Men				Women			
	n	r	se	p-value	n	r	se	p-value	n	r	se	p-value
ERF	2,447	-0.04	0.02	0.032	1,083	-0.10	0.03	0.001	1,364	-0.02	0.03	0.477
Esther	975	-0.05	0.03	0.139	413	-0.06	0.05	0.210	562	-0.03	0.04	0.421
Zuthpen Elderly Study	203	-0.02	0.07	0.801	203	-0.02	0.07	0.801	NA	NA	NA	NA
Meta-analysis	3,625	-0.04	0.02	0.010	1,699	-0.08	0.02	$5.8*10^{-4}$	1,929	-0.02	0.02	0.302

r = correlation coefficient; se = standard error; Total sample analysis adjusted for age and sex. Sex-stratified analysis adjusted for age

4

Table 3 shows the results of the association of known genome-wide significant telomere SNPs with height (Mendelian randomization scenario 1; **Figure 1A**). There were 7 independent SNPs known to associate with telomere length at a genome-wide significance level [41]. One SNP (rs9420907; *OBFC1*) was also significantly associated with height at a Bonferroni corrected level of significance of 0.007 (=0.05/7). The allele associated with increased telomere length was also associated with increased height. From the height to telomere relationship an inverse association was expected, which suggests that for this SNP the assumption of an effect of the genotype on height through telomere length was not met. A risk score of all variants associated with telomere length was not significantly associated with height (beta = 0.004; p-value = 0.11).

Table 3 | Association with height of SNPs genome-wide significantly associated with telomere length.

				telo_ENGAGE		height_GIANT			
SNP	**Chr**	**Gene**	**EA**	**beta**	**p-value**	**beta**	**se**	**p-value**	**N**
rs10936599	3	TERC	C	0.097	$2.54*10^{-31}$	0.002	0.005	0.640	133814
rs2736100	5	TERT	C	0.078	$4.38*10^{-19}$	-0.001	0.005	0.864	103310
rs7675998	4	NAF1	G	0.074	$4.35*10^{-16}$	0.002	0.006	0.785	133810
rs9420907	10	OBFC1	C	0.069	$6.90*10^{-11}$	0.019	0.006	0.004	133721
rs8105767	19	ZNF208	G	0.048	$1.11*10^{-9}$	0.010	0.005	0.045	133855
rs755017	20	RTEL1	G	0.062	$6.71*10^{-9}$	-0.014	0.009	0.124	75729
rs11125529	2	ACYP2	A	0.056	$4.48*10^{-8}$	0.012	0.006	0.056	133807
OVERALL						0.004	0.002	0.110	

EA = Effective allele
Telo_ENGAGE = genome-wide significant results from published telomere GWAS
Height_GIANT = look up of telomere SNPs in publically available height GWAS results

In **Table 4** the association of known height SNPs with telomere length is given. **Table 4** shows the SNPs that at least reached a p-value < 0.05 in the association with telomere length (Mendelian randomization scenario 2; **Figure 1B**). The complete list of SNPs (n = 115) can be found in **Supplementary Table 1**. Several of the height associated SNPs showed nominal association with telomere length, though none of the tested SNPs passed the Bonferroni corrected level of significance of $4.35*10^{-4}$ (= 0.05/115). There are twice as many of these SNPs associated in the expected direction (67%) than in the opposite direction. When combining the SNPs (p-value < 0.05) in a risk score, the overall effect of the height genes on telomere length is indeed inverse (beta = -0.013; p-value = 0.002) as expected. Ingenuity pathway analysis of the six genes (*POLR2B*, *ZBTB38*, *HMGA1*, *SERPINH1*, *PAPPA*

and *TGFB2*) that are associated in the expected direction showed that four of these genes (*HMGA1*, *SERPINH1*, *PAPPA* and *TGFB2*) are involved in cell proliferation (p-value = 0.04; **Supplementary Table 2**). The complete risk score of all 115 height-associated variants was not significantly associated with telomere length (beta = $-1.19*10^{-4}$; p-value = 0.925).

Table 4 | Association with telomere length of SNPs associated genome-wide significantly with height (p-value < 0.05).

				height_GIANT		telo_ENGAGE			
SNP	**Chr**	**Gene**	**EA**	**beta**	**p-value**	**beta**	**se**	**p-value**	**N**
rs17081935	4	POLR2B	T	0.030	$4.77*10^{-8}$	-0.027	0.009	0.002	37633
rs724016	3	ZBTB38	G	0.070	$4.47*10^{-52}$	-0.020	0.007	0.006	37323
rs2079795	17	TBX2	T	0.040	$1.22*10^{-16}$	0.018	0.008	0.022	37205
rs2780226	6	HMGA1	C	0.076	$1.02*10^{-18}$	-0.038	0.017	0.024	32536
rs634552	11	SERPINH1	T	0.039	$1.35*10^{-9}$	-0.026	0.012	0.026	29826
rs751543	9	PAPPA	T	0.026	$4.51*10^{-8}$	-0.019	0.009	0.034	32272
rs6569648	6	L3MBTL3	C	0.040	$8.93*10^{-12}$	0.019	0.009	0.039	34469
rs6684205	1	TGFB2	G	0.028	$1.97*10^{-11}$	-0.016	0.008	0.043	37638
rs6439167	3	C3orf47	C	0.034	$7.20*10^{-10}$	0.019	0.009	0.044	34618
OVERALL						-0.013	0.004	0.002	

EA = Effective allele
Telo_ENGAGE = look up of known height SNPs in telomere length GWAS
Height_GIANT = genome-wide significant results from published height GWAS

DISCUSSION

In this study, including 3,625 subjects derived from three population-based studies in the CHANCES consortium, we found evidence for a significant inverse correlation between telomere length and adult height. In the sex-stratified analysis the correlation was predominantly seen in men. Mendelian randomization analysis showed no evidence for a relation to height of the genes genome-wide significantly associated with telomere length. The analyses of all genetic variants genome-wide significantly associated with height failed to show an association with telomere length. However, we observed an overrepresentation of genes involved in cell proliferation in those height genes nominally associated with telomere length.

A limitation in the interpretation of our correlation analysis results is that in only the ERF population the association between height and telomere length reached significance. This may be for a large part explained by the differences in sample size. All three cohorts show effects in the same direction with similar effect sizes, strengthening our finding. Additionally,

4

no significant between-study heterogeneity was observed. Our findings may be confounded by unknown risk factors at young age which determine height and telomere length or by differential mortality later in life. However, the findings of the Mendelian randomization make this explanation less likely. Also the finding that the association is virtually limited to men makes confounding a less likely explanation as this has to occur only in men.

The inverse male-specific association between telomere length and height has been seen throughout various classes of vertebrates, pointing to an evolutionary conserved mechanism. Interestingly, this is not entirely explained by the larger body size of males, which requires more cellular replications to attain and maintain a large body size [43, 44]. In dunlins, female birds are on average larger than males, yet an association between height and telomere length is seen in males and not in females, similar to our findings in the present paper [32]. An explanation could be the difference in telomerase activity, which is lower in males compared to females [7, 30].

In the Mendelian randomization analysis, we did not find evidence supporting the hypothesis that genes determining telomere length also lead to a small body size. Although our knowledge of the genes determining telomere length is far from complete [41], the genes with a large effect on telomere length such as *TERT* do not show any evidence for association with height. Also biologically this explanation is less likely as early in life telomere length is similar in males and females [20], which predicts a similar body size in males and females if the hypothesis was true that telomere length determines height. We found that several height genes are associated with telomere length at nominal significance. Four of the six genes (67%) for which the direction of the association is in the direction anticipated are involved in cell proliferation. This is in line with the observation that high cell proliferation results in both increased height and decreased telomere length [2, 44, 45]. Our study suggests that this mechanism is partly genetically determined. Of note is however that there are over 160 genes implicated in height explaining in total 10% of the variation in height [42] while only 6 associated here with telomere length. On the one hand this may imply that not all pathways implicated in height are involved in telomere length, but rather the genes involved in a distinct pathway, i.e., cell proliferation. On the other hand, we cannot exclude a false positive association.

Of interest is whether the observed association between telomere length and height is related to the IIS pathway. Animal models including *C. elegans*, *D. melanogaster*, mice and dogs suggest that low IGF-1 levels are associated with longevity [24, 25, 46-48]. IGF-1 is also a major determinant of height in the population and has been associated with telomere length [49-51]. As measurements of IGF-1 are not available, we could not investigate this directly in the present paper. Since high IGF-1 levels relate to increased height, our findings predict an inverse association between IGF-1 and telomere length. However, previously, three studies have found a positive association between IGF-1 levels and telomere length [49-51], which is at odds with our findings. This issue remains to be resolved: the positive

association between IGF-1 levels and telomere length [49-51] is also at odds with the findings in many studies in animal models that show that lower IGF-1 levels are associated with longevity, predicting an inverse association between IGF-1 and telomeres. Up until now, IGF-1 levels have not been associated with human longevity consistently [27, 52-57]. Yet in the same study, researchers find that genes in the IIS pathway associated with lower height, also associate with survival to old age [27]. A recent report from the same group focusing on the genetic variants in a larger set of genes spanning the IIS pathway shows that several genes (*AKT1*, *AKT3*, *FOXO4*, *IGF2*, *INS*, *PIK3CA*, *SGK*, *SGK2*, and *YWHAG*) are jointly associated with longevity [58]. The GIANT GWAS hits on height contain 2 SNPs located in the vicinity of IIS pathway genes (*IGF1R* and *IGF2BP2*). However, in our study these 2 SNPs did not associate with telomere length. Our findings and that of others raise the question whether the inverse association between height and telomere length is explained by IGF-1 levels in the blood. A problem in the interpretation of human data is that IGF-1 levels decrease with age [59-61]. As a consequence, cross-sectional studies of IGF-1 levels may be biased as age related decline may be genetically determined [62].

In summary, we found consistent and significant evidence for an inverse correlation between telomere length and height, predominantly in men. We find nominal significant evidence consistent with the hypothesis that several genes involved in height through cell proliferation are also involved in telomere length.

REFERENCES

1. Blackburn, E.H., Telomere states and cell fates. Nature, 2000. 408(6808): p. 53-6.
2. Allsopp, R.C., et al., Telomere shortening is associated with cell division in vitro and in vivo. Exp Cell Res, 1995. 220(1): p. 194-200.
3. Atzmon, G., et al., Evolution in health and medicine Sackler colloquium: Genetic variation in human telomerase is associated with telomere length in Ashkenazi centenarians. Proc Natl Acad Sci U S A, 2010. 107 Suppl 1: p. 1710-7.
4. Harley, C.B., Telomere loss: mitotic clock or genetic time bomb? Mutat Res, 1991. 256(2-6): p. 271-82.
5. Shay, J.W. and W.E. Wright, Senescence and immortalization: role of telomeres and telomerase. Carcinogenesis, 2005. 26(5): p. 867-74.
6. Greider, C.W. and E.H. Blackburn, A telomeric sequence in the RNA of Tetrahymena telomerase required for telomere repeat synthesis. Nature, 1989. 337(6205): p. 331-7.
7. Gomes, N.M., et al., Comparative biology of mammalian telomeres: hypotheses on ancestral states and the roles of telomeres in longevity determination. Aging Cell, 2011. 10(5): p. 761-8.
8. Bakaysa, S.L., et al., Telomere length predicts survival independent of genetic influences. Aging Cell, 2007. 6(6): p. 769-74.
9. Cawthon, R.M., et al., Association between telomere length in blood and mortality in people aged 60 years or older. Lancet, 2003. 361(9355): p. 393-5.
10. Ehrlenbach, S., et al., Influences on the reduction of relative telomere length over 10 years in the population-based Bruneck Study: introduction of a well-controlled high-throughput assay. Int J Epidemiol, 2009. 38(6): p. 1725-34.
11. Epel, E.S., et al., The rate of leukocyte telomere shortening predicts mortality from cardiovascular disease in elderly men. Aging (Albany NY), 2009. 1(1): p. 81-8.
12. Kimura, M., et al., Telomere length and mortality: a study of leukocytes in elderly Danish twins. Am J Epidemiol, 2008. 167(7): p. 799-806.
13. Bischoff, C., et al., No association between telomere length and survival among the elderly and oldest old. Epidemiology, 2006. 17(2): p. 190-4.
14. Fitzpatrick, A.L., et al., Leukocyte telomere length and cardiovascular disease in the cardiovascular health study. Am J Epidemiol, 2007. 165(1): p. 14-21.
15. Harris, S.E., et al., The association between telomere length, physical health, cognitive ageing, and mortality in non-demented older people. Neurosci Lett, 2006. 406(3): p. 260-4.
16. Martin-Ruiz, C.M., et al., Telomere length in white blood cells is not associated with morbidity or mortality in the oldest old: a population-based study. Aging Cell, 2005. 4(6): p. 287-90.
17. Njajou, O.T., et al., Association between telomere length, specific causes of death, and years of healthy life in health, aging, and body composition, a population-based cohort study. J Gerontol A Biol Sci Med Sci, 2009. 64(8): p. 860-4.
18. Zekry, D., et al., Telomere length, comorbidity, functional, nutritional and cognitive status as predictors of 5 years post hospital discharge survival in the oldest old. J Nutr Health Aging, 2012. 16(3): p. 225-30.
19. Broer, L., et al., Meta-analysis of telomere length in 19 713 subjects reveals high heritability, stronger maternal inheritance and a paternal age effect. Eur J Hum Genet, 2013.
20. Okuda, K., et al., Telomere length in the newborn. Pediatr Res, 2002. 52(3): p. 377-81.
21. Stindl, R., Tying it all together: telomeres, sexual size dimorphism and the gender gap in life expectancy. Med Hypotheses, 2004. 62(1): p. 151-4.
22. Arias, E., United States life tables, 2007. Natl Vital Stat Rep, 2011. 59(9): p. 1-60.
23. Greider, C.W., Telomerase activity, cell proliferation, and cancer. Proc Natl Acad Sci U S A, 1998. 95(1): p. 90-2.
24. Brown-Borg, H.M., et al., Dwarf mice and the ageing process. Nature, 1996. 384(6604): p. 33.

25. Egenvall, A., et al., Age pattern of mortality in eight breeds of insured dogs in Sweden. Prev Vet Med, 2000. 46(1): p. 1-14.
26. Greer, K.A., S.C. Canterberry, and K.E. Murphy, Statistical analysis regarding the effects of height and weight on life span of the domestic dog. Res Vet Sci, 2007. 82(2): p. 208-14.
27. van Heemst, D., et al., Reduced insulin/IGF-1 signalling and human longevity. Aging Cell, 2005. 4(2): p. 79-85.
28. Clutton-Brock, T., S. Albon, and F. Guinness, Parental investment and sex differences in juvenile mortality in birds and mammals. Nature, 1985. 313: p. 131-133.
29. Barrett, E.L. and D.S. Richardson, Sex differences in telomeres and lifespan. Aging Cell, 2011. 10(6): p. 913-21.
30. Leri, A., et al., Telomerase activity in rat cardiac myocytes is age and gender dependent. J Mol Cell Cardiol, 2000. 32(3): p. 385-90.
31. Olsson, M., et al., Proximate determinants of telomere length in sand lizards (Lacerta agilis). Biol Lett, 2010. 6(5): p. 651-3.
32. Pauliny, A., et al., Age-independent telomere length predicts fitness in two bird species. Mol Ecol, 2006. 15(6): p. 1681-7.
33. Aulchenko, Y.S., et al., Linkage disequilibrium in young genetically isolated Dutch population. Eur J Hum Genet, 2004. 12(7): p. 527-34.
34. Pardo, L.M., et al., The effect of genetic drift in a young genetically isolated population. Ann Hum Genet, 2005. 69(Pt 3): p. 288-95.
35. Cawthon, R.M., Telomere measurement by quantitative PCR. Nucleic Acids Res, 2002. 30(10): p. e47.
36. Codd, V., et al., Common variants near TERC are associated with mean telomere length. Nat Genet, 2010. 42(3): p. 197-9.
37. Menotti, A., et al., Short-term all-cause mortality and its determinants in elderly male populations in Finland, The Netherlands, and Italy: the FINE Study. Finland, Italy, Netherlands Elderly Study. Prev Med, 1996. 25(3): p. 319-26.
38. R Development Core Team, R: A language and environment for statistical computing. 2010, R foundation for Statistical Computing: Vienna, Austria.
39. Schulze, R., Meta-analysis: a comparison of approaches. 2004, Gottingen, Germany: Hogrefe & Huber.
40. Smith, G.D. and S. Ebrahim, 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? Int J Epidemiol, 2003. 32(1): p. 1-22.
41. Codd, V., et al., Identification of seven loci affecting mean telomere length and their association with disease. Nat Genet, 2013. 45(4): p. 422-7.
42. Lango Allen, H., et al., Hundreds of variants clustered in genomic loci and biological pathways affect human height. Nature, 2010. 467(7317): p. 832-8.
43. Samaras, T.T., H. Elrick, and L.H. Storms, Is attainment of greater height and body size really desirable? J Natl Med Assoc, 1999. 91(6): p. 317-21.
44. Samaras, T.T., H. Elrick, and L.H. Storms, Is height related to longevity? Life Sci, 2003. 72(16): p. 1781-802.
45. Samaras, T.T. and H. Elrick, Height, body size and longevity. Acta Med Okayama, 1999. 53(4): p. 149-69.
46. Coschigano, K.T., et al., Assessment of growth parameters and life span of GHR/BP gene-disrupted mice. Endocrinology, 2000. 141(7): p. 2608-13.
47. Miller, R.A., Kleemeier award lecture: are there genes for aging? J Gerontol A Biol Sci Med Sci, 1999. 54(7): p. B297-307.
48. Shire, J.G., Growth hormone and premature ageing. Nature, 1973. 245(5422): p. 215-6.
49. Barbieri, M., et al., Higher circulating levels of IGF-1 are associated with longer leukocyte telomere length in healthy subjects. Mech Ageing Dev, 2009. 130(11-12): p. 771-6.
50. Kaplan, R.C., et al., Insulin-like growth factors and leukocyte telomere length: the cardiovascular health study. J Gerontol A Biol Sci Med Sci, 2009. 64(11): p. 1103-6.

51. Moverare-Skrtic, S., et al., Serum insulin-like growth factor-I concentration is associated with leukocyte telomere length in a population-based cohort of elderly men. J Clin Endocrinol Metab, 2009. 94(12): p. 5078-84.
52. Katic, M. and C.R. Kahn, The role of insulin and IGF-1 signaling in longevity. Cell Mol Life Sci, 2005. 62(3): p. 320-43.
53. Laron, Z., Do deficiencies in growth hormone and insulin-like growth factor-1 (IGF-1) shorten or prolong longevity? Mech Ageing Dev, 2005. 126(2): p. 305-7.
54. Rincon, M., E. Rudin, and N. Barzilai, The insulin/IGF-1 signaling in mammals and its relevance to human longevity. Exp Gerontol, 2005. 40(11): p. 873-7.
55. Rozing, M.P., et al., Human insulin/IGF-1 and familial longevity at middle age. Aging (Albany NY), 2009. 1(8): p. 714-22.
56. Rudman, D., et al., Effects of human growth hormone in men over 60 years old. N Engl J Med, 1990. 323(1): p. 1-6.
57. Steuerman, R., O. Shevah, and Z. Laron, Congenital IGF1 deficiency tends to confer protection against post-natal development of malignancies. Eur J Endocrinol, 2011. 164(4): p. 485-9.
58. Deelen, J., et al., Gene set analysis of GWAS data for human longevity highlights the relevance of the insulin/IGF-1 signaling and telomere maintenance pathways. Age (Dordr), 2013. 35(1): p. 235-49.
59. Goodman-Gruen, D. and E. Barrett-Connor, Epidemiology of insulin-like growth factor-I in elderly men and women. The Rancho Bernardo Study. Am J Epidemiol, 1997. 145(11): p. 970-6.
60. Lam, C.S., et al., Circulating insulin-like growth factor-1 and its binding protein-3: metabolic and genetic correlates in the community. Arterioscler Thromb Vasc Biol, 2010. 30(7): p. 1479-84.
61. Landin-Wilhelmsen, K., et al., Insulin-like growth factor I levels in healthy adults. Horm Res, 2004. 62 Suppl 1: p. 8-16.
62. Rietveld, I., et al., A polymorphic CA repeat in the IGF-I gene is associated with gender-specific differences in body height, but has no effect on the secular trend in body height. Clin Endocrinol (Oxf), 2004. 61(2): p. 195-203.

Chapter 4.4

Association of adiponectin and leptin with telomere length in 7 independent cohorts including 11,460 participants

Linda Broer, Julia Raschenberger, Joris Deelen, Massimo Mangino, Veryan Codd, Kirsi H Pietiläinen, Eva Albrecht, Najaf Amin, Marian Beekman, Anton JM de Craen, Christian Gieger, Margot Haun, Peter Henneman, Christian Herder, Iiris Hovatta, Ludmilla Kedenko, Wolfgang Koenig, Barbara Kollerits, Eeva Moilanen, Ben A Oostra, Bernhard Paulweber, Lydia Quaye, Aila Rissanen, Michael Roden, Ida Surakka, Ana M Valdes, Katriina Vuolteenaho, Barbara Thorand, Ko Willems van Dijk, Jaakko Kaprio, Tim D Spector, P Eline Slagboom, Nilesh J Samani, Florian Kronenberg, Cornelia M van Duijn, Karl-Heinz Ladwig

ABSTRACT

Oxidative stress and inflammation are major contributors to accelerated age-related relative telomere length shortening. Both conditions are strongly linked to leptin and adiponectin, the most prominent adipocyte-derived protein hormones. As high leptin levels and low levels of adiponectin have been implicated in inflammation, one expects adiponectin to be positively associated with telomere length while leptin should be negatively associated. Within the ENGAGE consortium, we investigated the association of telomere length with adiponectin and leptin in 7 independent cohorts with a total of 11,460 participants. We performed partial correlation analysis on Z-transformed telomere length and LN-transformed leptin/adiponectin, adjusting for age and sex. In extended models we adjusted for body mass index (BMI) and C-reactive protein (CRP).

Adiponectin showed a borderline significant association with telomere length. This appeared to be determined by a single study and when the outlier study was removed, this association disappeared. The association between telomere length and leptin was highly significant ($r = -0.05$; p-value = $1.81*10^{-7}$). Additional adjustment for BMI or CRP did not change the results. Sex-stratified analysis revealed no difference between men and women. Our study suggests that high leptin levels are associated with short telomere length.

INTRODUCTION

Both leptin and adiponectin are among the most prominent adipocyte-derived protein hormones. Their origin in adipocytes and ability to affect the expression of various markers of systemic inflammation has led to the notion of both protein hormones as adipokines [1]. High leptin levels have been implicated in contributing to inflammation, insulin resistance, glucose intolerance and atherosclerosis [2]. Low levels of adiponectin, on the other hand, have been linked to inflammation, insulin resistance, impaired endothelium-dependent vasodilatation, elevated systemic blood pressure and further deteriorating effects on cardiovascular and metabolic health [3]. Interestingly, for both adipocytokines, impaired efficacy at the level of the receptors have been considered to be a mechanism potentially akin to that seen under the condition of sustained elevated insulin levels (insulin resistance) [4].

Telomeres are tandem repeats of long hexamers (TTAGGG) at the end of chromosomes that protect against spontaneous DNA damage and thus preserve genomic integrity [5, 6]. Oxidative stress and inflammation are major causative contributors to accelerated relative telomere length shortening [7-10]. Given the associations of the adipocytokines with inflammation and oxidative stress, one would expect that adiponectin should be positively associated with telomere length while leptin should be negatively associated.

Only a limited number of studies have investigated the association of adipocytokines with telomere length and have yielded conflicting findings. Up to now, 6 studies have been published on the relation between telomere length and leptin with sample sizes varying from 317 to 2,721 subjects [11-16]. An inverse association was found in two studies [11, 15] while two did not find evidence for a significant association [12, 16]. One study found borderline significant evidence for a positive association [13]. The last study found borderline significant evidence for a positive association when adjusting for percentage body fat [14]. There have been 3 previous studies on the association between telomere length and adiponectin, including 193 to 570 subjects [12, 16, 17] with one showing a positive association [17]. In the present study we aimed to investigate the association of telomere length with adiponectin and leptin in 7 independent cohorts with a total of 11,460 participants.

METHODS

Study populations

The Erasmus Rucphen Family (**ERF**) study is a cross-sectional cohort including 3,000 living descendants of 22 couples who had at least 6 children baptized in the community church around 1850-1900. The participants are not selected on any disease or other outcome. Details about the genealogy of the population have been described elsewhere [18, 19]. The study protocol was approved by the medical ethics board of the Erasmus MC Rotterdam, the Netherlands.

The **KORA** (Cooperative Health Research in the Region of Augsburg) study is a series of independent population-based epidemiological surveys and follow-up studies of participants living in the region of Augsburg, Southern Germany [20]. All survey participants are of German nationality, identified through the registration office. Informed consent has been given by all participants. The present study includes data of the KORA F3 (2004/2005) survey which is a follow-up study of the KORA S3 survey (1994/1995), as well as data of the KORA F4 (2006-2008) study which is a follow-up study of the KORA S4 survey (1999-2001). All KORA F4 participants were fasting. In KORA F3, adjustment for fasting status was made.
For the Leiden Longevity Study (**LLS**), long-lived siblings of Dutch descent were recruited together with their offspring and the partners of thereof. Families were included if at least two long-lived siblings were alive and fulfilled the age criterion of 89 years or older for males and 91 years or older for females, representing less than 0.5% of the Dutch population in 2001 [21]. In total, 944 long-lived proband siblings from 421 families with a mean age of 94 years (range, 89-104), 1,671 offspring (61 years, 39-81), and 744 partners (60 years, 36-79) were included in the study. DNA from the LLS was extracted from samples at baseline using conventional methods [22]. For the current analysis only the offspring and their partners were used.
The **TwinsUK** cohort (www.twinsuk.ac.uk) is an adult twin British registry shown to be representative of singleton populations and the United Kingdom population [23]. A total of 6,038 twins with telomere length measurement were included in the analysis. The age range of the TwinsUK cohort was 16-99 years. Ethical approval was obtained from the Guy's and St. Thomas' Hospital Ethics Committee. Written informed consent was obtained from every participant in the study.
The **SAPHIR** Study (Salzburg Atherosclerosis Prevention program in subjects at High Individual Risk) is an observational study in a healthy working population conducted in the years 1999-2002 involving 1532 healthy unrelated subjects with telomere length measurement available: 539 females aged 42-67 years and 993 males aged 39-66. All subjects were recruited by health-screening programs in companies in and around the Austrian city Salzburg (1999-2002) [24]. At baseline, all study participants were subjected to a comprehensive screening examination with a detailed personal and family history assessed via standardized questionnaires and a physical examination. Subjects suffering from severe obesity (BMI > 40kg/m^2), established coronary artery disease, congestive heart failure, valvular heart disease, chronic alcohol consumption (more than three drinks per day), drug abuse or pregnancy were excluded.
The **Finnish twin** participants were recruited from two population-based longitudinal studies, FinnTwin16 and FinnTwin12, each consisting of five consecutive birth cohorts (1975-1959, n = 5601 subjects in FinnTwin16 and 1983-1987, n = 5184 in FinnTwin12, respectively) [25]. All pairs were of Caucasian origin. Except for one obese male subject who had recently developed type 2 diabetes and used insulin, the subjects were healthy and

did not take any medications. Zygosity was confirmed by genotyping of ten informative genetic markers [26]. For the current study only those individuals with measurements on both telomere length and adipokines were included (n = 190). They had taken part in a sub-study of metabolic factors related to obesity, and were representative of the distribution of BMI in the two twin cohorts. No further exclusion criteria were used. The protocol was designed and performed according to the principles of the Helsinki Declaration and was approved by the Ethical Committee of the Helsinki University Central Hospital.

Relative TL, adiponectin and leptin measurements

Details on the measurements of adiponectin, leptin and telomere length for the individual cohorts have been published elsewhere and are provided for this paper in the supplementary material. In all studies telomere length was measured in leukocytes by qPCR [27]. For 4 out of the 7 studies telomere length was determined in one central laboratory according to common protocol. Adiponectin was available in all cohorts, while leptin was not available in SAPHIR.

Data analysis

In order to standardize telomere length measurements across cohorts a Z-transformation was applied. Adiponectin and leptin levels were transformed using a LN-transformation. Samples deviating more than 4 standard deviations from the mean were removed. We performed partial correlation analysis, adjusting for age and sex, and if necessary, for family relationships and/or fasting status. In the extended model we additionally adjusted for body mass index (BMI) or C-reactive protein (CRP). We also performed sex-specific analyses, thus resulting in a Bonferroni corrected p-value threshold for significance of 0.006 (= 0.05/9).

RESULTS

The general characteristics of the study populations are depicted in **Table 1** stratified for men and women. Most cohorts had a mean age around 50 years, except for Finnish Twins in which the mean age was 28 years at the time of assessment. Most cohorts also had an approximately equal number of men and women, except for TwinsUK which includes only women.

Table 2 contains the partial correlation analysis of telomere length and BMI. We observed a nominally significant correlation between BMI and telomere length in the total population (r = -0.02, p-value = 0.021). We did not observe a significant correlation between CRP and telomere length (r = -0.01, p-value = 0.444, **Supplementary Table 1**).

Table 3 shows the partial correlation analysis of telomere length and adiponectin. In the overall meta-analysis, we observed a nominally significant correlation between telomere length and adiponectin (r = 0.02, p-value = 0.047). There was significant evidence for

heterogeneity (I^2 = 69.8%). The high heterogeneity was caused by one outlying study (ERF). In a sensitivity analysis we removed ERF from the meta-analysis, which completely resolved the heterogeneity and resulted in a loss of significance (r = 0.004, p-value = 0.763). Additionally adjusting for BMI (r = 0.02, p-value = 0.104) or CRP (r = 0.02, p-value = 0.083) did not change the results (**Supplementary Table 2**).

Table 1 | General characteristics of study populations.

	Men			**Women**		
	n	**mean**	**sd**	**n**	**mean**	**sd**
Telomere length (t/s ratio)						
ERF	1,230	1.74	0.35	1,539	1.81	0.36
KORA F3	1,517	1.69	0.28	1,607	1.75	0.29
KORA F4	1,456	1.79	0.32	1,568	1.90	0.32
LLS	1,043	1.43	0.25	1,262	1.48	0.26
TwinsUK	NA	NA	NA	1,428	3.50	0.63
SAPHIR	985	0.85	0.17	539	0.85	0.18
Finnish Twins	107	1.29	0.28	83	1.24	0.23
Adiponectin (mg/L)						
ERF	904	8.06	4.13	1,188	12.36	5.79
KORA F3	1,535	8.67	3.66	1,624	12.14	4.57
KORA F4	581	8.95	5.38	550	14.56	7.35
LLS	1,036	4.81	2.19	1,207	7.23	3.12
TwinsUK	NA	NA	NA	1,185	7.82	3.65
SAPHIR	985	7.05	3.30	539	11.07	5.14
Finnish Twins	107	2.30	0.77	83	3.22	1.02
Leptin (ng/mL)						
ERF	687	24.91	29.38	905	74.85	65.35
KORA F3	800	9.53	9.41	825	30.76	26.12
KORA F4	1,467	9.58	11.37	1,584	28.18	23.70
LLS	1,042	9.66	8.98	1,206	26.19	18.28
TwinsUK	NA	NA	NA	1,428	16.89	11.52
SAPHIR	NA	NA	NA	NA	NA	NA
Finnish Twins	107	5.58	4.69	83	20.96	15.89
Age (yr)						
ERF	1,230	50.17	14.65	1,539	49.44	15.05
KORA F3	1,545	57.73	13.15	1,639	57.10	12.66
KORA F4	1,486	56.64	13.41	1,594	55.51	13.14
LLS	1,063	59.87	6.85	1,288	58.60	6.76

	Men			Women		
	n	mean	sd	n	mean	sd
TwinsUK	NA	NA	NA	1,428	48.71	13.48
SAPHIR	985	48.85	5.46	539	55.63	4.18
Finnish Twins	107	28.25	1.86	83	28.08	1.60
BMI (kg/m^2)						
ERF	1,191	27.24	4.13	1,472	26.65	5.10
KORA F3	1,531	28.01	3.86	1,626	27.27	5.16
KORA F4	1,480	27.93	4.19	1,583	27.33	5.32
LLS	904	25.77	3.05	1,101	25.10	3.94
TwinsUK	NA	NA	NA	1,426	24.85	4.41
SAPHIR	984	26.91	3.70	539	26.59	4.75
Finnish Twins	107	24.92	3.72	83	24.19	5.23
CRP (mg/L)						
ERF	849	3.53	10.50	1,131	3.92	8.13
KORA F3	144	4.06	5.72	117	5.06	10.67
KORA F4	1,477	2.40	5.18	1,582	2.61	5.41
LLS	1,029	2.11	2.89	1,234	2.53	3.66
TwinsUK	NA	NA	NA	1,250	3.02	4.44
SAPHIR	983	0.26	0.73	538	0.35	0.51
Finnish Twins	107	0.75	1.03	83	2.31	3.36

NA: not available

The results of the partial correlation analysis of telomere length and leptin are depicted in **Table 4**. We observed a weak, but highly significant, correlation between telomere length and leptin in the overall meta-analysis ($r = -0.05$, p-value = $1.81*10^{-7}$). No heterogeneity was observed ($I^2 = 0\%$). When removing the cohort with the strongest association (KORA F4) from the analysis, the correlation did not change ($r = -0.04$, p-value = $1.17*10^{-3}$). Additionally adjusting for BMI ($r = -0.04$, p-value = $1.93*10^{-5}$) or CRP ($r = -0.06$, p-value = $2.10*10^{-8}$) did not change the results (**Supplementary Table 3**). The multivariate analysis shows that leptin explains 0.16% of the telomere length variance. The sex-stratified analysis also showed very similar correlations for men and women ($r = -0.06$ and $r = -0.05$ respectively), which indicates that the association between telomere length and leptin is not sex-specific.

4

Table 2 | Correlation between telomere length and BMI.

	Total				Men				Women			
Study	**n**	**r**	**se**	**p-value**	**n**	**r**	**se**	**p-value**	**n**	**r**	**se**	**p-value**
ERF	2,663	0.003	0.02	0.891	1,191	-0.01	0.03	0.664	1,472	-0.01	0.03	0.749
KORA F3	3,095	-0.01	0.02	0.674	1,501	0.001	0.03	0.984	1,594	-0.02	0.03	0.415
KORA F4	2,998	-0.07	0.02	$1.20*10^{-4}$	1,442	-0.09	0.03	$3.90*10^{-4}$	1,556	-0.06	0.03	0.027
LLS	1,949	-0.003	0.02	0.882	883	0.01	0.03	0.776	1,066	-0.01	0.03	0.718
TwinsUK	1,426	-0.02	0.03	0.450	0	NA	NA	NA	1,426	-0.02	0.03	0.450
SAPHIR	1,520	-0.01	0.03	0.607	982	-0.04	0.03	0.181	538	0.02	0.04	0.591
FTC	190	0.06	0.07	0.423	107	0.01	0.10	0.904	83	0.10	0.11	0.373
Meta-analysis	13,841	-0.02	0.01	0.021	6,106	-0.03	0.01	0.020	7,735	-0.02	0.01	0.084

NA: not available.

Table 3 | Correlation between telomere length and adiponectin.

	Total				Men				Women			
Study	**n**	**r**	**se**	**p-value**	**n**	**r**	**se**	**p-value**	**n**	**r**	**se**	**p-value**
ERF	2,092	0.09	0.02	$6.38*10^{-5}$	904	0.07	0.03	0.043	1,188	0.11	0.03	$1.03*10^{-4}$
KORA F3	3,120	0.02	0.02	0.171	1,515	0.001	0.03	0.963	1,605	0.05	0.02	0.056
KORA F4	1,116	-0.04	0.03	0.178	574	-0.07	0.04	0.082	542	-0.005	0.04	0.914
LLS	2,233	0.001	0.02	0.944	1,018	0.02	0.03	0.034	1,215	-0.02	0.03	0.109
TwinsUK	1,185	0.002	0.03	0.942	NA	NA	NA	NA	1,185	0.00	0.03	0.942
SAPHIR	1,524	0.01	0.03	0.845	985	-0.002	0.03	0.950	539	0.01	0.04	0.799
FTC	190	-0.08	0.07	0.247	107	0.02	0.10	0.815	83	-0.25	0.11	0.024
Meta-analysis	11,460	0.02	0.01	0.047	5,103	0.01	0.01	0.526	6,357	0.03	0.01	0.026

NA: not available.

Table 4 | Correlation between telomere length and leptin.

	Total				Men				Women			
Study	**n**	**r**	**se**	**p-value**	**n**	**r**	**se**	**p-value**	**n**	**r**	**se**	**p-value**
ERF	1,592	-0.03	0.03	0.201	687	-0.03	0.04	0.368	905	-0.05	0.03	0.156
KORA F3	1,613	-0.03	0.02	0.247	797	-0.01	0.04	0.816	816	-0.05	0.04	0.170
KORA F4	3,010	-0.08	0.02	$2.84*10^{-5}$	1,443	-0.11	0.03	$3.83*10^{-5}$	1,567	-0.04	0.03	0.087
LLS	2,260	-0.05	0.02	0.013	1,031	-0.05	0.03	0.090	1,229	-0.05	0.03	0.070
TwinsUK	1,428	-0.05	0.03	0.054	NA	NA	NA	NA	1,428	-0.05	0.03	0.054
SAPHIR	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
FTC	190	-0.03	0.07	0.680	107	-0.11	0.10	0.262	83	0.09	0.11	0.422
Meta-analysis	10,093	-0.05	0.01	$1.81*10^{-7}$	4,065	-0.06	0.02	$7.38*10^{-5}$	6,028	-0.05	0.01	$3.26*10^{-4}$

NA: not available.

4

DISCUSSION

In this study including data of 11,460 subjects derived from population-based studies, we found evidence for a significant negative association between telomere length and serum leptin levels. The relationship was independent of age, sex, BMI and CRP. However, no relationship was found between telomere length and adiponectin. A nominally significant association of telomere length with BMI was also observed.

The correlation between leptin and telomere length, adjusted for age and sex was small but highly significant ($r = -0.05$, p-value = $1.81*10^{-7}$). Based on the assumption of an association between high leptin levels and inflammation [2], we observed an inverse association between leptin and telomere length. The current study is the largest study on the association between telomere length and leptin to date with 10,093 subjects studied compared to the previous studies with sample size ranging from 317 to 2,721 subjects [11-16]. Two out of the six previously performed studies on the association between leptin and telomere length also found an inverse association only in women [11, 15]. The present study therefore provides a strong confidence in the observed association and elucidates the association in both men and women.

High leptin levels have been implicated in contributing to inflammation, to insulin resistance, glucose intolerance and atherosclerosis [2]. Furthermore, leptin is an important mediator in stress-induced cardiovascular activity mainly by raising circulating catecholamine concentrations and by stimulating heart rate and arterial blood pressure [1, 28]. Though the percentage of variance explained is small (~0.16%), the observed association between leptin and telomere length is interesting, due to the biological link with both inflammation and oxidative stress of both variables.

The findings on adiponectin were not consistent over studies with evidence for substantial heterogeneity. Removing the outlier study (ERF) led to a non-significant association. There may be various explanations for this finding. Although differences between populations may exist in terms of BMI and metabolic syndrome (to which adiponectin is strongly related) [29], it is less likely that these differences fully explain the discrepancies as the leptin association is highly consistent and is also strongly associated to BMI and the metabolic syndrome [30]. Additionally we attempted to elucidate why ERF showed a very different association from the other populations by additionally adjusting for metabolic syndrome, but this did not change the results. Therefore, it is most likely that the association in ERF was a chance finding.

It is currently unclear whether high adiponectin levels are part of a protective effect due to an ongoing inflammation or whether individuals with low adiponectin levels are prone to inflammation. Another possibility would be that changes in adiponectin levels might be an epiphenomenon. Several prospective studies reported an association of high rather than low adiponectin levels with outcomes such as (cardiovascular) mortality [31-35], progression of chronic kidney disease [36-38] and dementia [39]. This positive association

was mainly observed in patients suffering already at baseline from a chronic disease. Other studies reported low adiponectin levels to be associated with several diseases such as insulin resistance [40-42], type 2 diabetes [43-45], atherosclerosis or cardiovascular disease [46]. It has been proposed that reverse epidemiology [47] or adiponectin resistance [48, 49] may be an explanation.

A limitation of our study is that we do not have detailed measurements of adiposity, except BMI. Though adjusting for BMI did not change our results for either adiponectin or leptin and most adiposity measures are highly correlated to each other, we cannot completely exclude that adjusting for some other adiposity measurement might have changed the results. Observed differences in telomere length measurements across cohorts are mainly related to age differences between the various populations.

A major strength of the present study is its large sample size. This large number of samples requires telomere measurement to be done by a qPCR method instead of Southern Blot techniques. This approach for assessing telomere length has some important advantages, including the high-throughput capacity and the requirement of only low amounts of DNA. However, we are also aware of the potential limitations of this method which are e.g. the type of the reference sample [50]. However, if performed in a highly standardized way, telomere qPCR is a valuable method for large epidemiological studies without consuming large amounts of DNA. The alternatively used method of telomere length measurement is the telomere restriction fragment (TRF) analysis (Southern Blot) where the absolute telomere length (kb) is estimated directly from a telomeric smear signal from blot pictures. However, this approach consumes a considerable amount of DNA, the required reagents are costly, and the methodology itself is cumbersome and labor-intensive. The results of the Southern Blot method can be influenced by several factors such as restriction enzymes, hybridization targets, hybridization probes and hybridization conditions, gel calibration, background subtraction, as well as the calculation formula and the analysis window [50]. Despite these limitations, to compare results of different studies, accurately measured absolute telomere length might be a more adequate and concrete method as a relative approach by qPCR. However, since the measurements in our meta-analyses were performed to a large extent in one laboratory and difference between methods were considered by using a Z-transformation of the data, we consider the qPCR method with the large number of samples as a major advantage.

Another advantage of our study is the population-based and multicenter design which allows the evaluation of consistent findings over populations. Though the meta-analysis also includes studies with relatedness (ERF, LLS, TwinsUK, Finnish twins), the relatedness of subjects within families was taken into account statistically in order to derive correct standard errors and p-values. The impact of modern living environment is undisputable. Nutrition, activity as well as the exposure to a wide range of man-made chemicals plays an important role and may have an effect on insulin action, metabolic rate and other

physiological processes [51, 52]. Nevertheless, we observed an association of telomere length and leptin in population-based studies of unrelated individuals as well as studies with related individuals sampled from the population without ascertainment for any disease.
In summary, our study is by far the largest addressing the relation between telomere length and adipocytokines. We found consistent and significant evidence for an inverse correlation between telomere length and leptin, which was not caused by increased BMI or CRP. However, we could not find a consistent association between telomere length and adiponectin. This study has elucidated the association between adipocytokines and telomere length in both sexes.

REFERENCES

1. Tilg, H. and A.R. Moschen, Adipocytokines: mediators linking adipose tissue, inflammation and immunity. Nat Rev Immunol, 2006. 6(10): p. 772-83.
2. Conde, J., et al., At the crossroad between immunity and metabolism: focus on leptin. Expert Rev Clin Immunol, 2010. 6(5): p. 801-8.
3. Hui, X., et al., Adiponectin and cardiovascular health: an update. Br J Pharmacol, 2012. 165(3): p. 574-90.
4. Enriori, P.J., et al., Leptin resistance and obesity. Obesity (Silver Spring), 2006. 14 Suppl 5: p. 254S-258S.
5. Blackburn, E.H. and J.G. Gall, A tandemly repeated sequence at the termini of the extrachromosomal ribosomal RNA genes in Tetrahymena. J Mol Biol, 1978. 120(1): p. 33-53.
6. Blackburn, E.H., C.W. Greider, and J.W. Szostak, Telomeres and telomerase: the path from maize, Tetrahymena and yeast to human cancer and aging. Nat Med, 2006. 12(10): p. 1133-8.
7. Halvorsen, T.L., et al., Accelerated telomere shortening and senescence in human pancreatic islet cells stimulated to divide in vitro. J Endocrinol, 2000. 166(1): p. 103-9.
8. Kurz, D.J., et al., Chronic oxidative stress compromises telomere integrity and accelerates the onset of senescence in human endothelial cells. J Cell Sci, 2004. 117(Pt 11): p. 2417-26.
9. von Zglinicki, T., Oxidative stress shortens telomeres. Trends Biochem Sci, 2002. 27(7): p. 339-44.
10. von Zglinicki, T., et al., Mild hyperoxia shortens telomeres and inhibits proliferation of fibroblasts: a model for senescence? Exp Cell Res, 1995. 220(1): p. 186-93.
11. Aviv, A., et al., Menopause modifies the association of leukocyte telomere length with insulin resistance and inflammation. J Clin Endocrinol Metab, 2006. 91(2): p. 635-40.
12. Diaz, V.A., et al., Telomere length and adiposity in a racially diverse sample. Int J Obes (Lond), 2010. 34(2): p. 261-5.
13. Njajou, O.T., et al., Shorter telomeres are associated with obesity and weight gain in the elderly. Int J Obes (Lond), 2011.
14. Njajou, O.T., et al., Shorter telomeres are associated with obesity and weight gain in the elderly. Int J Obes (Lond), 2012. 36(9): p. 1176-9.
15. Valdes, A.M., et al., Obesity, cigarette smoking, and telomere length in women. Lancet, 2005. 366(9486): p. 662-4.
16. Zhu, H., et al., Leukocyte telomere length in healthy Caucasian and African-American adolescents: relationships with race, sex, adiposity, adipokines, and physical activity. J Pediatr, 2011. 158(2): p. 215-20.
17. Al-Attas, O.S., et al., Adiposity and insulin resistance correlate with telomere length in middle-aged Arabs: the influence of circulating adiponectin. Eur J Endocrinol, 2010. 163(4): p. 601-7.
18. Aulchenko, Y.S., et al., Linkage disequilibrium in young genetically isolated Dutch population. Eur J Hum Genet, 2004. 12(7): p. 527-34.
19. Pardo, L.M., et al., The effect of genetic drift in a young genetically isolated population. Ann Hum Genet, 2005. 69(Pt 3): p. 288-95.
20. Wichmann, H.E., et al., KORA-gen--resource for population genetics, controls and a broad spectrum of disease phenotypes. Gesundheitswesen, 2005. 67 Suppl 1: p. S26-30.
21. Schoenmaker, M., et al., Evidence of genetic enrichment for exceptional survival using a family approach: the Leiden Longevity Study. Eur J Hum Genet, 2006. 14(1): p. 79-84.
22. Beekman, M., et al., Chromosome 4q25, microsomal transfer protein gene, and human longevity: novel data and a meta-analysis of association studies. J Gerontol A Biol Sci Med Sci, 2006. 61(4): p. 355-62.
23. Moayyeri, A., et al., Cohort Profile: TwinsUK and Healthy Ageing Twin Study. Int J Epidemiol, 2012.
24. Heid, I.M., et al., Genetic architecture of the APM1 gene and its influence on adiponectin plasma levels and parameters of the metabolic syndrome in 1,727 healthy Caucasians. Diabetes, 2006. 55(2): p. 375-84.

25. Kaprio, J., Twin studies in Finland 2006. Twin Res Hum Genet, 2006. 9(6): p. 772-7.
26. Pietilainen, K.H., et al., Growth patterns in young adult monozygotic twin pairs discordant and concordant for obesity. Twin Res, 2004. 7(5): p. 421-9.
27. Cawthon, R.M., Telomere measurement by quantitative PCR. Nucleic Acids Res, 2002. 30(10): p. e47.
28. Knudson, J.D., et al., Leptin and mechanisms of endothelial dysfunction and cardiovascular disease. Curr Hypertens Rep, 2008. 10(6): p. 434-9.
29. Okamoto, Y., et al., Adiponectin: a key adipocytokine in metabolic syndrome. Clin Sci (Lond), 2006. 110(3): p. 267-78.
30. Considine, R.V., et al., Serum immunoreactive-leptin concentrations in normal-weight and obese humans. N Engl J Med, 1996. 334(5): p. 292-5.
31. Kistorp, C., et al., Plasma adiponectin, body mass index, and mortality in patients with chronic heart failure. Circulation, 2005. 112(12): p. 1756-62.
32. Kizer, J.R., et al., Adiponectin and risk of coronary heart disease in older men and women. J Clin Endocrinol Metab, 2008. 93(9): p. 3357-64.
33. Maiolino, G., et al., Plasma adiponectin for prediction of cardiovascular events and mortality in high-risk patients. J Clin Endocrinol Metab, 2008. 93(9): p. 3333-40.
34. Menon, V., et al., Adiponectin and mortality in patients with chronic kidney disease. J Am Soc Nephrol, 2006. 17(9): p. 2599-606.
35. Pilz, S., et al., Adiponectin and mortality in patients undergoing coronary angiography. J Clin Endocrinol Metab, 2006. 91(11): p. 4277-86.
36. Jorsal, A., et al., Serum adiponectin predicts all-cause mortality and end stage renal disease in patients with type I diabetes and diabetic nephropathy. Kidney Int, 2008. 74(5): p. 649-54.
37. Kollerits, B., et al., Gender-specific association of adiponectin as a predictor of progression of chronic kidney disease: the Mild to Moderate Kidney Disease Study. Kidney Int, 2007. 71(12): p. 1279-86.
38. Saraheimo, M., et al., Serum adiponectin and progression of diabetic nephropathy in patients with type 1 diabetes. Diabetes Care, 2008. 31(6): p. 1165-9.
39. van Himbergen, T.M., et al., Biomarkers for Insulin Resistance and Inflammation and the Risk for All-Cause Dementia and Alzheimer Disease: Results From the Framingham Heart Study. Arch Neurol, 2012.
40. Costacou, T., et al., The prospective association between adiponectin and coronary artery disease among individuals with type 1 diabetes. The Pittsburgh Epidemiology of Diabetes Complications Study. Diabetologia, 2005. 48(1): p. 41-8.
41. Pischon, T., et al., Plasma adiponectin levels and risk of myocardial infarction in men. JAMA, 2004. 291(14): p. 1730-7.
42. Schulze, M.B., et al., Adiponectin and future coronary heart disease events among men with type 2 diabetes. Diabetes, 2005. 54(2): p. 534-9.
43. Arita, Y., et al., Paradoxical decrease of an adipose-specific protein, adiponectin, in obesity. Biochem Biophys Res Commun, 1999. 257(1): p. 79-83.
44. Hu, E., P. Liang, and B.M. Spiegelman, AdipoQ is a novel adipose-specific gene dysregulated in obesity. J Biol Chem, 1996. 271(18): p. 10697-703.
45. Maeda, N., et al., Diet-induced insulin resistance in mice lacking adiponectin/ACRP30. Nat Med, 2002. 8(7): p. 731-7.
46. Han, S.H., et al., Antiatherosclerotic and anti-insulin resistance effects of adiponectin: basic and clinical studies. Prog Cardiovasc Dis, 2009. 52(2): p. 126-40.
47. Kalantar-Zadeh, K., et al., Epidemiology of dialysis patients and heart failure patients. Semin Nephrol, 2006. 26(2): p. 118-33.
48. Furuhashi, M., et al., Possible impairment of transcardiac utilization of adiponectin in patients with type 2 diabetes. Diabetes Care, 2004. 27(9): p. 2217-21.
49. Kadowaki, T. and T. Yamauchi, Adiponectin and adiponectin receptors. Endocr Rev, 2005. 26(3): p. 439-51.

50. Horn, T., B.C. Robertson, and N.J. Gemmell, The use of telomere length in ecology and evolutionary biology. Heredity (Edinb), 2010. 105(6): p. 497-506.

51. Koch, H.M. and A.M. Calafat, Human body burdens of chemicals used in plastic manufacture. Philos Trans R Soc Lond B Biol Sci, 2009. 364(1526): p. 2063-78.

52. Newbold, R.R., Impact of environmental endocrine disrupting chemicals on the development of obesity. Hormones (Athens), 2010. 9(3): p. 206-17.

Chapter 4.5

Associations between telomere length and metabolite profiles provide new insight into the biology of aging

Linda Broer, Harmen HM Draisma, René Pool, Eva Albrecht, Marian Beekman, Massimo Mangino, Mait Raag, Dale R Nyholt, Harish K Dharuri, Veryan Codd, Najaf Amin, Eco JC de Geus, Joris Deelen, Ayse Demirkan, Idil Erte, Krista Fischer, Toomas Haller, Anjali K Henders, Aaron Isaacs, Sarah E Medland, Grant W Montgomery, Simon P Mooijaart, Ben A Oostra, Konstantin Strauch, H Eka D Suchiman, Anika AM Vaarhorst, Diana van Heemst, Rui Wang-Sattler, John B Whitfield, Gonneke Willemsen, Margaret J Wright, Nicholas G Martin, Nilesh J Samani, Andres Metspalu, P Eline Slagboom, Tim D Spector, Dorret I Boomsma, Cornelia M van Duijn, Christian Gieger

ABSTRACT

Telomere shortening has been associated with multiple age-related diseases and with mortality. However, biological mechanisms responsible for these associations are partially unknown. Recent metabolomics techniques enables capturing a wide range of intermediate metabolites in serum. In an effort to better understand the biology behind the association of telomere length with age-related disease and mortality, we investigate the association between telomere length and serum metabolites in 8,192 individuals from seven independent cohorts from European ancestry.

We measured 163 serum metabolites with mass spectrometry and telomere length was determined using quantitative polymerase chain reaction. Using partial correlation analysis, we identified six metabolites to be significantly associated with telomere length, including three phosphatidylcholines (PCs), two amino acids (methionine and tyrosine) and one acyl-carnitine. Additionally we performed pathway analysis using Taverna to delve deeper into the biological mechanisms and identified homocysteine metabolism as the biological mechanism underlying the in common associations with telomere length between PCs and methionine.

In conclusion we found evidence for the involvement of lipid, homocysteine and thyroid metabolism in telomere length attrition, shedding light on the association of telomere length with age-related diseases and mortality.

INTRODUCTION

Telomeres are located at the end of chromosomes and protect against spontaneous DNA damage, thus preserving genomic integrity [1, 2]. The progressive shortening of telomere length with each subsequent cell division underlies the so-called mitotic clock, i.e. the limited replicative capacity of a cell [3]. Replicative senescence and subsequent cell death occurs when the mean telomere length reaches a critical value and telomere length is therefore seen as a marker for biological age [4-6].

Short telomere length has been associated with multiple age-related diseases including coronary heart disease, hypertension and dementia [7-10]. In prospective studies, short telomere length has been associated to mortality [11-14], although not all studies confirm this [15-17]. Also, the postulated causal factors of aging, oxidative stress and inflammation, are major contributors to accelerated telomere length shortening [18-21]. Although there is some evidence that telomere length is a marker of disease risk and early mortality, biological mechanisms underlying the association of telomere length with age-related diseases and mortality are unknown.

Longevity in humans has been associated with various metabolic traits in the elderly, including blood pressure and lipids suggesting that changes at metabolic level are a key feature in longevity. Metabolic profiles, as generated by metabolomics [22], have been associated with age and various aging-related diseases, such as type 2 diabetes, atherosclerosis, cancer and Alzheimer's Disease [23-27]. Metabolic profiles capture the effect of genes and environment, providing new and exciting opportunities to gain new insight in the underlying mechanism of the association between telomere length and mortality.

In this study we investigated the association between telomere length and serum metabolites in 8,192 individuals from 7 independent population-based cohorts from Europe and Australia. Additionally we aimed to diverge which metabolic pathways are represented by the metabolites significantly associated with telomere length in order to better understand the biology behind the associations.

METHODS

Cohort descriptions and measurements of telomere length and metabolites

Details on the separate cohorts can be found in the supplementary material. Details on the measurement of telomere length and the metabolites as measured by the AbsoluteIDQ p150 kit (Biocrates AG, Innsbruck, Austria) are also provided in the supplementary material. Both telomere length and metabolite concentrations were measured centrally according to a common protocol. Telomere length and metabolite measurements were performed in samples taken at the same time point. To ensure data quality each metabolite had to meet three criteria for inclusion in the study, namely 1) coefficient of variation not exceeding 25%; 2) less than 5% missing values; 3) median of metabolite concentrations

above the limit of detection. This was done on a per cohort basis. The metabolites measured include hexoses (H1), amino acids (AA), acyl-carnitines (AC), sphingomyelins (SMs), diacyl phosphatidylcholines (PC aa), acyl-alkyl PCs (PC ae) and lyso PCs.

Analysis and multiple testing correction

In order to standardize telomere length measurement across populations we used a Z-transformation. Metabolite concentrations were natural log transformed to attain a better approximation of the normal distribution. We performed partial correlation analysis, adjusting for age and sex, and, if necessary, for family relationships. In the extended model we additionally adjusted for body mass index (BMI). Meta-analysis was performed using R-package metacor, implementing DerSimonian-Laird (DSL) random-effect meta-analytical approach, per metabolite over all cohorts for which this metabolite passed pre-defined QC parameters [28, 29]. **Supplementary Table 1** contains the reasons for exclusion from the analysis of each metabolite.

A multiple testing-corrected statistical significance threshold for association of metabolite concentrations with telomere length was defined at the meta-analysis level. We accounted for multiple testing by Bonferroni-Sidak correction based on the effective number of independent variables (VeffLi) in the metabolite concentration data [30] (http://gump.qimr.edu.au/general/daleN/matSpDlite/). The number of VeffLi was determined on the correlation matrix of the quality controlled and log-transformed metabolomics data in the ERF and NTR cohorts.

Biological interpretation of significantly associated metabolites

For interpretation of the observed associations we followed two bioinformatics approaches in parallel. First we employed a bioinformatics pipeline based on a workflow management software tool called Taverna (http://www.taverna.org.uk) [31]to determine if the metabolites significantly associated with telomere length shared a network space within two reaction steps. We took as input any two metabolites; all the reactions within a radius of two steps in the reaction space of the first metabolite are obtained from the KEGG database [32]. The second metabolite is searched against the substrates and the products of the reactions obtained in the previous step. The presence of the second metabolite in any of the reaction steps is an indication that the two metabolites participate in reactions within two steps of each other; in the final step the path between the two metabolites is returned to the user. In order to prevent non-specific connections, an intermediate step filters out hub metabolites such as ATP, ADP, and NADP etc.

Second we derived a correlation matrix for the significantly associated metabolites and visualized this in a correlogram using R package corrgram [28]. In the correlogram a blue color represents a positive correlation while a red color represents a negative correlation. A darker color indicates a stronger correlation between two variables.

RESULTS

General characteristics of the study populations are depicted in **Table 1**. Mean age of the cohorts ranged from 18.4 to 62.9 years. BMI was not available at the time of metabolite assessment in QIMR and was on average between 25 and 27 kg/m^2 in the other cohorts.

Separate analyses to determine the effective number of independent variables of the metabolite concentration data for two cohorts that participated in this study (NTR and ERF) resulted in 46 VeffLi in both ERF and NTR, illustrating the consistency in the number of independent variables present in the metabolomics data from both cohorts. This resulted in a p-value of 0.001 (= 0.05/46) as significance level.

Table 1 | General characteristics of study populations.

				telomere length (T/S ratio)		Age (years)		BMI (kg/m^2)	
	n	**n_BMI***	**% female**	**mean**	**sd**	**mean**	**sd**	**mean**	**sd**
KORA	3003	2988	51.8	1.85	0.33	56.08	13.25	27.61	4.80
NTR	1314	1307	33.3	2.54	0.47	50.60	14.13	25.97	3.80
EGCUT	1084	1081	50.3	1.90	0.30	37.78	15.70	25.16	4.56
TwinsUK	1149	1149	100.0	3.58	0.64	53.72	10.76	26.44	5.35
ERF	806	806	53.7	1.79	0.37	47.76	13.97	27.17	4.81
LLS	643	643	50.1	1.44	0.27	62.91	6.61	26.65	4.01
QIMR	193	0	48.2	3.43	0.56	18.44	12.65	NA	NA

*Number of samples with BMI measurements at the time of sampling for metabolite and telomere length analysis.

Of the 163 measured metabolites 131 passed quality control in at least one study. 27 metabolites showed at least nominally significant correlation (p-value <= 0.05) with telomere length (**Table 2**). The associations of six of these metabolites, lysoPC a C17:0, Met, Tyr, PC aa C32:1, C3-OH, PCaeC38:4, with telomere length passed the multiple-testing corrected statistical significance threshold. When additionally adjusting for BMI the association with telomere length for five metabolites remained significant (lysoPC a C17:0, Met, Tyr, PC aa C32:1, C3-OH). The top hit is for lysoPC a C17:0 with a correlation coefficient of 0.05 in the age and sex adjusted model (p-value = $7.10*10^{-6}$) and a correlation coefficient of 0.04 in the model additionally adjusted for BMI (p-value = $4.72*10^{-4}$). A list of all metabolite correlations for both models with telomere length can be found in **Supplementary Table 2**. **Supplementary Table 3** contains study-specific results for the age and sex-adjusted model.

Table 2 | Meta-analysis of partial correlation analysis of telomere length and metabolites (p-value in model 1 < 0.05).

	Model 1: age + sex				Model 2: age + sex + BMI				
metabolite	**n**	**direction**	**r**	**p-value**	**n**	**direction**	**r**	**p-value**	**metabolite full name**
lysoPC a C17:0	8192	++-++++	0.05	**$7.10*10^{-6}$**	7977	++-+++	0.04	**$4.72*10^{-4}$**	lysoPhosphatidylcholine acyl C17:0
Met	8192	------+	-0.04	**$9.20*10^{-5}$**	7977	------	-0.05	**$7.51*10^{-5}$**	Methionine
Tyr	7386	----?-+	-0.04	**$2.14*10^{-4}$**	7171	----?-	-0.04	**0.001**	Tyrosine
PC aa C32:1	8192	----+-+	-0.04	$2.44*10^{-4}$	7977	----++	-0.04	**$3.38*10^{-4}$**	Phosphatidylcholine diacyl C32:1
C3-OH	1449	????--?	-0.10	**$2.64*10^{-4}$**	1449	????--	-0.10	**$2.75*10^{-4}$**	Hydroxypropionylcarnitine
PC ae C38:4	8192	+-+-+++	0.04	**0.001**	7977	+-+-++	0.03	0.005	Phosphatidylcholine acyl-alkyl C38:4
PC ae C40:3	8192	+++++++	0.04	0.002	7977	++++++	0.03	0.009	Phosphatidylcholine acyl-alkyl C40:3
PC ae C40:5	8192	+-+++++	0.04	0.002	7977	+-++++	0.03	0.019	Phosphatidylcholine acyl-alkyl C40:5
SM C20:2	8192	+-+-+++	0.03	0.002	7977	+-+-++	0.03	0.003	Sphingomyeline C20:2
C9	5601	+??-++?	0.04	0.003	5586	+??-++	0.04	0.011	Nonaylcarnitine
PC ae C40:4	8192	+0+++++	0.03	0.004	7977	+-++++	0.03	0.028	Phosphatidylcholine acyl-alkyl C40:4
PC aa C38:3	8192	----+++	-0.03	0.007	7977	----++	-0.02	0.035	Phosphatidylcholine diacyl C38:3
PC ae C36:1	8192	+-+-+++	0.03	0.007	7977	+-+-++	0.02	0.058	Phosphatidylcholine acyl-alkyl C36:1
PC aa C36:1	8192	----+++	-0.03	0.009	7977	----++	-0.03	0.004	Phosphatidylcholine diacyl C36:1
PC ae C40:6	8192	+-+-+++	0.03	0.009	7977	+-+-++	0.02	0.085	Phosphatidylcholine acyl-alkyl C40:6
SM (OH) C16:1	7386	+-+-?++	0.03	0.011	7171	+-+-?+	0.02	0.053	Hydroxysphingomyeline C16:1
C2	8192	--+---+	-0.03	0.014	7977	--+---	-0.03	0.015	Acetylcarnitine
PC ae C36:2	8192	++--+++	0.03	0.014	7977	+---++	0.02	0.132	Phosphatidylcholine acyl-alkyl C36:2
PC ae C38:3	8192	+++-+++	0.03	0.017	7977	+++-++	0.02	0.036	Phosphatidylcholine acyl-alkyl C38:3
PC aa C42:0	8192	+++++++	0.03	0.020	7977	+++-++	0.02	0.089	Phosphatidylcholine diacyl C42:0
PC aa C32:0	8192	----+++	-0.03	0.021	7977	----++	-0.03	0.006	Phosphatidylcholine diacyl C32:0
PC aa C40:5	8192	----+++	-0.03	0.022	7977	----++	-0.03	0.017	Phosphatidylcholine diacyl C40:5

metabolite	Model 1: age + sex				Model 2: age + sex + BMI				metabolite full name
	n	direction	r	p-value	n	direction	r	p-value	
PC aa C38:1	836	?????++	0.08	0.026	643	?????+	0.05	0.173	Phosphatidylcholine diacyl C38:1
PC aa C36:2	8192	-----++	-0.02	0.030	7977	-----+	-0.03	0.010	Phosphatidylcholine diacyl C36:2
PC aa C34:1	8192	----+++	-0.02	0.031	7977	----++	-0.03	0.013	Phosphatidylcholine diacyl C34:1
PC ae C42:4	8192	+-+++++	0.02	0.035	7977	+-++++	0.01	0.209	Phosphatidylcholine acyl-alkyl C42:4
SM C26:0	5794	+??-+++	0.03	0.043	5586	+??-++	0.02	0.229	Sphingomyeline C26:0

*Direction: KORA, NTR, EGCUT, TwinsUK, ERF, LLS, QIMR; Direction of effect represented by – (negative correlation) + (positive correlation) or ? (not included) for each study.

4

The pathway analysis using the Taverna Workflow shows that PCs and Met participate within two reaction steps of each other in homocysteine metabolism. Briefly, PC is converted to betaine which is used to convert homocysteine to methionine. A correlogram of the six metabolites significantly associated with telomere length in the first model is presented in **Figure 1**. The amino-acids Metionine and Tyrorsine are highly correlated with one another. Metionine also shows strong correlation with the PC and lysoPC species.

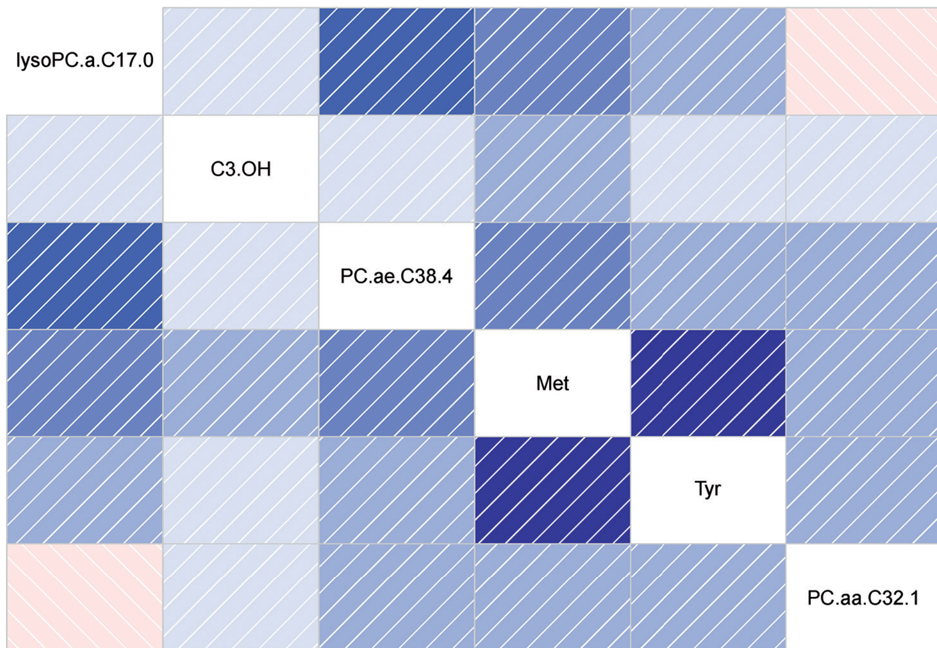


Figure 1 | Correlogram of 6 metabolites associated with telomere length after correction for multiple testing.

DISCUSSION

In this study we find significant association between telomere length and 6 metabolites, including three phosphocolines, two aminoacids and one acylcarnitine. Longer leukocyte telomeres are associated with higher levels of lysoPCaC17:0 and PCaeC38:4, and with lower levels of Methionine, Tyrosine, PCaaC32:1 and C3-OH. Additional pathway analysis points to an involvement of homocysteine metabolism in telomere length.

PCs constitute more than two-thirds of the total phospholipid content in HDL-cholesterol [33], which is known to have anti-oxidant and anti-inflammatory properties [34, 35].

Additionally, a common feature of cellular senescence, which is triggered by short telomeres, is an increased cell surface and a higher uptake of PCs from HDL-Cholesterol with long-chain fatty-acid residues thereby removing them from the serum [36]. Therefore, a possible biological mechanism can be constructed tying lipid metabolism to telomere length and aging. Short telomeres trigger cellular senescence which leads to an increase in the uptake of PCs, thereby lowering their serum levels.

Lower concentrations of methionine are associated with longer leukocyte telomere length. Methionine is an essential amino acid and is important for methylation reactions and as a precursor for cysteine. However, excessive intake of methionine is known to be highly toxic and leading to increased oxidative stress [37]. Though some of the harmful effects of methionine have been attributed to its product homocysteine, there is evidence for a direct toxic effect as well [38, 39]. Additionally, in animal studies it has been shown that restricting methionine consumption increases lifespan [40].

Furthermore, lower tyrosine levels associate with longer leukocyte telomere length. Tyrosine is a non-essential amino acid that is a precursor for several neurotransmitters, including dopamine, as well as thyroid hormones (T_3 and T_4) [41, 42]. Low thyroid hormone levels have been associated with increased lifespan in multiple animal models [43-45]. This association has also been observed in humans [46-48]. The observed inverse association between telomere length and tyrosine therefore makes biological sense and links telomere length to thyroid metabolism in longevity determination.

Pathway analysis using the Taverna Workflow revealed that four of the metabolites that associated with telomere length (i.e., the three PCs and Met) participate in homocysteine metabolism. Elevated homocysteine levels have previously been associated with shortened telomere length [49]. Additionally, exposure of endothelial cells to homocysteine has been shown to significantly increase senescence and to increase the amount of telomere length lost per population doubling [50]. Homocysteine has also been associated to age-related disease and mortality previously [51-54]. Taking all evidence together, a biological link between homocysteine metabolism and telomere length in the determination of longevity seems likely. The correlogram shows strong positive correlation between the two amino-acids as well as between methionine and the PC and LPC species. Given that methionine and PCs are in the same pathway (homocysteine metabolism) the correlogram supports the findings of the Taverna Workflow.

A major strength of this study is that both telomere length and the metabolites were measured centrally and using standard protocols. A limitation of the Taverna pathway analysis is that it does not accept individual PC species as separate terms in the analysis nor provides p-values.

In conclusion we find evidence for the association of homocysteine, thyroid and lipid metabolism with telomere length in 8192 individuals from 7 independent cohorts. Thereby shedding light on the biology underlying the association between telomere length and mortality and age-related related diseases.

REFERENCES

1. Blackburn, E.H. and J.G. Gall, A tandemly repeated sequence at the termini of the extrachromosomal ribosomal RNA genes in Tetrahymena. J Mol Biol, 1978. 120(1): p. 33-53.
2. Blackburn, E.H., C.W. Greider, and J.W. Szostak, Telomeres and telomerase: the path from maize, Tetrahymena and yeast to human cancer and aging. Nat Med, 2006. 12(10): p. 1133-8.
3. Vaziri, H., et al., Evidence for a mitotic clock in human hematopoietic stem cells: loss of telomeric DNA with age. Proc Natl Acad Sci U S A, 1994. 91(21): p. 9857-60.
4. Abdallah, P., et al., A two-step model for senescence triggered by a single critically short telomere. Nat Cell Biol, 2009. 11(8): p. 988-93.
5. Blackburn, E.H., Switching and signaling at the telomere. Cell, 2001. 106(6): p. 661-73.
6. Campisi, J. and F. d'Adda di Fagagna, Cellular senescence: when bad things happen to good cells. Nat Rev Mol Cell Biol, 2007. 8(9): p. 729-40.
7. Brouilette, S., et al., White cell telomere length and risk of premature myocardial infarction. Arterioscler Thromb Vasc Biol, 2003. 23(5): p. 842-6.
8. Brouilette, S.W., et al., Telomere length, risk of coronary heart disease, and statin treatment in the West of Scotland Primary Prevention Study: a nested case-control study. Lancet, 2007. 369(9556): p. 107-14.
9. Demissie, S., et al., Insulin resistance, oxidative stress, hypertension, and leukocyte telomere length in men from the Framingham Heart Study. Aging Cell, 2006. 5(4): p. 325-30.
10. Martin-Ruiz, C., et al., Telomere length predicts poststroke mortality, dementia, and cognitive decline. Ann Neurol, 2006. 60(2): p. 174-80.
11. Bakaysa, S.L., et al., Telomere length predicts survival independent of genetic influences. Aging Cell, 2007. 6(6): p. 769-74.
12. Cawthon, R.M., et al., Association between telomere length in blood and mortality in people aged 60 years or older. Lancet, 2003. 361(9355): p. 393-5.
13. Ehrlenbach, S., et al., Influences on the reduction of relative telomere length over 10 years in the population-based Bruneck Study: introduction of a well-controlled high-throughput assay. Int J Epidemiol, 2009. 38(6): p. 1725-34.
14. Kimura, M., et al., Telomere length and mortality: a study of leukocytes in elderly Danish twins. Am J Epidemiol, 2008. 167(7): p. 799-806.
15. Bischoff, C., et al., No association between telomere length and survival among the elderly and oldest old. Epidemiology, 2006. 17(2): p. 190-4.
16. Harris, S.E., et al., The association between telomere length, physical health, cognitive ageing, and mortality in non-demented older people. Neurosci Lett, 2006. 406(3): p. 260-4.
17. Martin-Ruiz, C.M., et al., Telomere length in white blood cells is not associated with morbidity or mortality in the oldest old: a population-based study. Aging Cell, 2005. 4(6): p. 287-90.
18. Halvorsen, T.L., et al., Accelerated telomere shortening and senescence in human pancreatic islet cells stimulated to divide in vitro. J Endocrinol, 2000. 166(1): p. 103-9.
19. Kurz, D.J., et al., Chronic oxidative stress compromises telomere integrity and accelerates the onset of senescence in human endothelial cells. J Cell Sci, 2004. 117(Pt 11): p. 2417-26.
20. von Zglinicki, T., Oxidative stress shortens telomeres. Trends Biochem Sci, 2002. 27(7): p. 339-44.
21. von Zglinicki, T., et al., Mild hyperoxia shortens telomeres and inhibits proliferation of fibroblasts: a model for senescence? Exp Cell Res, 1995. 220(1): p. 186-93.
22. Psychogios, N., et al., The human serum metabolome. PLoS One, 2011. 6(2): p. e16957.
23. Cao, M., et al., NMR-based metabolomic analysis of human bladder cancer. Anal Sci, 2012. 28(5): p. 451-6.
24. Matsumoto, T., T. Kobayashi, and K. Kamata, Role of lysophosphatidylcholine (LPC) in atherosclerosis. Curr Med Chem, 2007. 14(30): p. 3209-20.

25. Oresic, M., et al., Metabolome in progression to Alzheimer's disease. Transl Psychiatry, 2011. 1: p. e57.
26. Suhre, K., et al., Metabolic footprint of diabetes: a multiplatform metabolomics study in an epidemiological setting. PLoS One, 2010. 5(11): p. e13953.
27. Yu, Z., et al., Human serum metabolic profiles are age dependent. Aging Cell, 2012. 11(6): p. 960-7.
28. R Development Core Team, R: A language and environment for statistical computing. 2010, R foundation for Statistical Computing: Vienna, Austria.
29. Schulze, R., Meta-analysis: a comparison of approaches. 2004, Gottingen, Germany: Hogrefe & Huber.
30. Li, J. and L. Ji, Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. Heredity (Edinb), 2005. 95(3): p. 221-7.
31. Hull, D., et al., Taverna: a tool for building and running workflows of services. Nucleic Acids Res, 2006. 34(Web Server issue): p. W729-32.
32. Kanehisa, M. and S. Goto, KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res, 2000. 28(1): p. 27-30.
33. Engelmann, B., et al., Transfer of phosphatidylcholine, phosphatidylethanolamine and sphingomyelin from low- and high-density lipoprotein to human platelets. Biochem J, 1996. 315 (Pt 3): p. 781-9.
34. Negre-Salvayre, A., et al., Antioxidant and cytoprotective properties of high-density lipoproteins in vascular cells. Free Radic Biol Med, 2006. 41(7): p. 1031-40.
35. Norata, G.D. and A.L. Catapano, Molecular mechanisms responsible for the antiinflammatory and protective effect of HDL on the endothelium. Vasc Health Risk Manag, 2005. 1(2): p. 119-29.
36. Naru, E., et al., Increased levels of a particular phosphatidylcholine species in senescent human dermal fibroblasts in vitro. Hum Cell, 2008. 21(3): p. 70-8.
37. Mori, N. and K. Hirayama, Long-term consumption of a methionine-supplemented diet increases iron and lipid peroxide levels in rat liver. J Nutr, 2000. 130(9): p. 2349-55.
38. Harper, A.E., N.J. Benevenga, and R.M. Wohlhueter, Effects of ingestion of disproportionate amounts of amino acids. Physiol Rev, 1970. 50(3): p. 428-558.
39. Troen, A.M., et al., Lifespan modification by glucose and methionine in Drosophila melanogaster fed a chemically defined diet. Age (Dordr), 2007. 29(1): p. 29-39.
40. Miller, R.A., et al., Methionine-deficient diet extends mouse lifespan, slows immune and lens aging, alters glucose, T4, IGF-I and insulin levels, and increases hepatocyte MIF levels and stress resistance. Aging Cell, 2005. 4(3): p. 119-25.
41. During, M.J., I.N. Acworth, and R.J. Wurtman, Effects of systemic L-tyrosine on dopamine release from rat corpus striatum and nucleus accumbens. Brain Res, 1988. 452(1-2): p. 378-80.
42. Evans, R.M., The steroid and thyroid hormone receptor superfamily. Science, 1988. 240(4854): p. 889-95.
43. Brown-Borg, H.M., et al., Dwarf mice and the ageing process. Nature, 1996. 384(6604): p. 33.
44. Ooka, H., S. Fujita, and E. Yoshimoto, Pituitary-thyroid activity and longevity in neonatally thyroxine-treated rats. Mech Ageing Dev, 1983. 22(2): p. 113-20.
45. Tatar, M., A. Bartke, and A. Antebi, The endocrine regulation of aging by insulin-like signals. Science, 2003. 299(5611): p. 1346-51.
46. Atzmon, G., et al., Extreme longevity is associated with increased serum thyrotropin. J Clin Endocrinol Metab, 2009. 94(4): p. 1251-4.
47. Rozing, M.P., et al., Familial longevity is associated with decreased thyroid function. J Clin Endocrinol Metab, 2010. 95(11): p. 4979-84.
48. Rozing, M.P., et al., Low serum free triiodothyronine levels mark familial longevity: the Leiden Longevity Study. J Gerontol A Biol Sci Med Sci, 2010. 65(4): p. 365-8.
49. Richards, J.B., et al., Homocysteine levels and leukocyte telomere length. Atherosclerosis, 2008. 200(2): p. 271-7.

4

50. Xu, D., R. Neville, and T. Finkel, Homocysteine accelerates endothelial cell senescence. FEBS Lett, 2000. 470(1): p. 20-4.
51. Bostom, A.G., et al., Nonfasting plasma total homocysteine levels and stroke incidence in elderly persons: the Framingham Study. Ann Intern Med, 1999. 131(5): p. 352-5.
52. Dangour, A.D., et al., Plasma homocysteine, but not folate or vitamin B-12, predicts mortality in older people in the United Kingdom. J Nutr, 2008. 138(6): p. 1121-8.
53. Nygard, O., et al., Plasma homocysteine levels and mortality in patients with coronary artery disease. N Engl J Med, 1997. 337(4): p. 230-6.
54. Vollset, S.E., et al., Plasma total homocysteine and cardiovascular and noncardiovascular mortality: the Hordaland Homocysteine Study. Am J Clin Nutr, 2001. 74(1): p. 130-6.

Chapter 5

General Discussion

Chapter 5.1

Distinguishing true from false positives in genomic studies: p-values

Linda Broer, Christina M. Lill, Maaike Schuur, Najaf Amin, Johannes T. Roehr, Lars Bertram, John P.A. Ioannidis, Cornelia M. van Duijn

Published: European Journal of Epidemiology; 2012

ABSTRACT

Distinguishing true from false positive findings is a major challenge in human genetic epidemiology. Several strategies have been devised to facilitate this, including the positive predictive value (PPV) and a set of epidemiological criteria, known as the "Venice" criteria. The PPV measures the probability of a true association, given a statistically significant finding, while the Venice criteria grade the credibility based on the amount of evidence, consistency of replication and protection from bias. A vast majority of journals use significance thresholds to identify the true positive findings. We studied the effect of p-value thresholds on the PPV and used the PPV and Venice criteria to define usable thresholds of statistical significance.

Theoretical and empirical analyses of data published on AlzGene show that at a nominal p-value threshold of 0.05 most "positive" findings will turn out to be false if the prior probability of association is below 0.10 even if the statistical power of the study is higher than 0.80. However, in underpowered studies (0.25) with a low prior probability of $1*10^{-3}$, a p-value of $1*10^{-5}$ yields a high PPV (> 96%).

Here we have shown that the p-value threshold of $1*10^{-5}$ gives a very strong evidence of association in almost all studies. However, in the case of a very high prior probability of association (0.50) a p-value threshold of 0.05 may be sufficient, while for studies with very low prior probability of association ($1*10^{-4}$; genome-wide association studies (GWAS) for instance) $1*10^{-7}$ may serve as a useful threshold to declare significance.

INTRODUCTION

Distinguishing true from false positive findings is one of the most important challenges in genetic epidemiology [1, 2]. Even though reaching formal statistical significance (p-value < 0.05), many published research findings were not replicated in further studies [3]. Several methodologies have been developed to address this issue and identify the true positive findings. One of the approaches that were proposed to distinguish true from false positives is the False Positive Report Probability (FPRP) [4]. FPRP measures the probability of no true association between a genetic variant and disease given a statistically significant finding. The FPRP is a function of type 1 error (α), power (1-β) and prior probability (π) ($FPRP = \frac{\alpha(1-\pi)}{\alpha(1-\pi)+(1-\beta)\pi}$). FPRP has been shown to never reach 'acceptable' values if the prior probability is low [5]. This makes its application rather impractical in case of genome-wide association studies (GWAS) in which millions of genetic variants are tested and only hundreds to thousands are expected to be associated [6, 5]. Although, a general problem of any Bayesian method is the necessity of a prior probability, however, in the case of the FPRP it is not possible to update it directly with a new prior probability [6-9, 5]. This is crucial in moving findings from gene discovery (GWAS) to a replication study in which the prior probability has increased substantially.

As an alternate Weitkunat and Ioannidis *et al* suggested true report probability (TRP) also known as the positive predictive value (PPV) [3, 5], which measures the probability of a true association, given a statistically significant finding. The PPV is also a function of type 1 error, power and prior probability ($PPV = \frac{(1-\beta)\pi}{\pi(1-\beta)+(1-\pi)\alpha}$). As one would expect the FPRP and PPV are complementary probabilities to each other [3, 5]. In contrast to the FPRP, the PPV can be viewed as the posterior probability of an association, which reflects the update of the prior probability through the results of a study [5]. The PPV can be used as prior for subsequent studies and was used to show that replication rather than power in the discovery phase increases the credibility of a genetic association [5]. This conclusion has also been supported by others [3, 10].

A more epidemiological approach is the evidence based criteria that were developed at a workgroup meeting of the Human Genome Epidemiology Network (HuGENet) in Venice, Italy, (subsequently we refer it to as the "Venice criteria") to grade the credibility of cumulative genetic association evidence [11]. The Venice criteria assess genetic associations on the basis of the 1. "amount of evidence based on sample size"; 2. "consistency of replication" across studies, and 3. "protection from bias". Using these criteria, statistically significant meta-analyses in large populations with little heterogeneity between the results of the individual studies and without evidence of bias are considered to display "strong" credibility in favor of the genetic association, whereas one or more violations of the criteria result in only "moderate" or "weak" credibility. As the criteria are only applied to statistically significant associations, an implicit fourth criterion of the Venice criteria is that statistical significance can be determined [11]. A problem not addressed in the criteria is how to

define significance. As recognized for long in classical and clinical epidemiology, the major problem in gene discovery studies is to define the threshold.

While most classical and clinical epidemiological studies use the threshold of 0.05 and ignore the multiple testing, most GWAS adjust for multiple testing and use the p-value threshold of $5*10^{-8}$ [12]. The rationale to adjust for multiple testing is based on the large number of tests performed (millions of genetic markers in a single study) in combination with the low prior probability for association of a single variant. A less conservative p-value for GWAS is based on the false discovery rate [13]. For 2.5 million SNPs tested, a p-value of $1*10^{-7}$ implies that the expected number of false positive results is equal to or less than 1, suggesting that the generally accepted p-value threshold of $5*10^{-8}$ might be slightly conservative [14]. Conversely a p-value of 0.05 for candidate gene studies is often used [11], based on the argument that these targeted studies should be penalized less severely for multiple testing as fewer tests are conducted. However, this threshold is recognized to be lenient and often results in false-positive associations [11, 15].

Although we recognize the arbitrariness of defining a threshold for significance, in practice not only the Venice criteria but also the vast majority of (clinical) journals uses thresholds (< 0.05, < 0.01, < 0.001) [16]. These thresholds have been reviewed recently, showing that the threshold determines whether a finding is published as well as the impact of the journal in which it is published [16]. In this paper we studied the effect of p-value thresholds on the PPV extending the work of Weitkunat and Ioannidis *et al*, we then used the PPV and the Venice criteria to define usable thresholds of statistical significance for various types of (genetic) association studies.

MATERIALS AND METHODS

First, we performed simulations to determine the PPV for a combination of different p-value thresholds, statistical power and prior probability of association. Second, we verified the theoretically suggested thresholds by assessing the robustness of formally significant genetic associations (p-value < 0.05) in the AlzGene database, which uses the Venice criteria to grade nominally significant meta-analysis results after systematically combining genetic association data in Alzheimer's disease (AD), between 2005 and 2010 [17].

Venice criteria

The Venice criteria grade the strength of the epidemiological credibility of genetic associations by taking into account the amount of evidence, consistency of replication and protection from bias [11]. In most applications to-date, the amount of evidence is graded as "strong" (grade 'A') when the combined number of cases and controls in the minor genetic group exceeds 1000. The amount of evidence is scored as "moderate" (grade 'B') when the number is between 100 and 1000, and "weak" (grade 'C') when it is below 100. For consistency

of replication, the degree of heterogeneity between study results (I^2) is considered [18]. Point estimates of I^2 below 25% receive an 'A', I^2 between 25% and 50% receive a 'B' and I^2 exceeding 50% receive a 'C'. This criterion does not apply to meta-analyses with a p-value < $1*10^{-7}$ after exclusion of the initial dataset(s), as described in Khoury *et al*, 2009 [15]. For protection from bias, the guidelines propose to consider potential sources of bias at the level of individual studies including errors in phenotypes, genotypes and confounding, and at the level of meta-analysis including publication, first study and other selective reporting biases [11]. Associations receive an 'A' when bias is not likely to affect the presence of the association, a 'B' when there is no demonstrable bias, but important information is missing for its appraisal and a 'C' when there is potential or evident bias that may have invalidated the association. "Strong" epidemiological credibility for significant association is assigned to the meta-analyses that receive three A's, "moderate" credibility to those that receive any 'B', but no 'C', and "weak" credibility to those that receive a 'C' in any of the three criteria.

AlzGene database

The AlzGene database is a publicly available database of published genetic association studies in AD (available at www.alzgene.org) and provides regularly updated meta-analyses of eligible polymorphisms after inclusion of all available, published data [17]. Genetic association studies are identified through systematic, regularly updated literature searches. Studies are included in AlzGene if they assess the association of polymorphisms and AD risk and if they are published in English in peer-reviewed journals. Meta-analyses using the DerSimonian-Laird random-effects model [19] are conducted for polymorphisms for which at least four independent datasets are available. Furthermore, meta-analyses are performed after stratification for ancestry if three or more datasets are available, which typically is only the case for studies using Caucasian and Asian descent populations. Nominally significant meta-analysis results are highlighted in a specific section of AlzGene ("Top Results"); the credibility of these associations is assessed on the basis of the Venice criteria, and results of this assessment are displayed online alongside the meta-analysis results (http://www.alzgene.org/TopResults.asp). The criteria are implemented as described above with the amount of evidence quantified as the number of minor alleles. The degree of heterogeneity between study results (I^2) is also considered. For protection from bias, the following potential reasons for bias in the meta-analysis results are systematically assessed: summary OR < 1.15, loss of significance after exclusion of first study, loss of significance after exclusion of studies violating HWE in control populations, evidence for publication/small-study bias using a modified regression test [20], and evidence for excess of statistically significant results [21]. These protection from bias tests are not applied to meta-analyses with a p-value < $1*10^{-7}$ after exclusion of the initial studie(s).

AlzGene currently includes almost 1400 studies covering over 650 genes and 2900 polymorphisms (as accessed on 5th January 2012) [17]. Meta-analysis results are available for 320 polymorphisms.

Analysis

To investigate the p-value thresholds we used the PPV. The PPV is based on power (1-β; proportional to the product of effect and sample size), type 1 error (α) and prior probability of a true association (π). The PPV is calculated as: $PPV = \frac{(1-\beta)\pi}{\pi(1-\beta)+(1-\pi)\alpha}$ [3, 5].

We investigated the PPV for five different prior probabilities for association: $1*10^{-4}$, $1*10^{-3}$, 0.01, 0.10 and 0.50. The lower-bound prior probability reflects a GWAS situation in Caucasians where 500,000 linkage disequilibrium blocks exist and are tested. Assuming that 50 of these blocks are truly associated with the trait of interest yields a prior probability of association of $1*10^{-4}$. The upper-bound prior probability of 0.50 reflects the replication of a genome-wide significant finding or a finding that approaches genome-wide significance. The prior probability of $1*10^{-3}$ to 0.10 were chosen to reflect the prior probability in candidate gene studies which may be low ($1*10^{-3}$) when there is only limited prior evidence to suggest an association and higher (0.01-0.10) for SNPs that already show fairly compelling evidence for association. In order to investigate the influence of the p-value on the PPV we varied the p-value from 0.05 to $5*10^{-8}$ at a power of 0.25, 0.50, 0.80 and 0.95.

In order to evaluate the p-value threshold at which grading of formally significant genetic association results no longer empirically changed from strong to moderate or weak credibility, we monitored the AlzGene top results list between 2005 and 2010 and followed changes in the credibility assessment by the Venice criteria (see above). Meta-analyses of the AlzGene data and Venice grading were performed using the R programming language, packages HardyWeinberg version 1.4, rmeta version 2.16, and a customized R code.

To determine the overlap between the top results at each time point (end of the year), we counted the number of associations that had changed in credibility since the previous time point and overall. Specifically, we focused on meta-analysis results with strong credibility that became weak or non-significant as well as results with moderate or weak credibility that transitioned to strong credibility.

RESULTS

Investigating the influence of p-value on the PPV

Figure 1 shows the effect of varying the p-value threshold on the PPV. **Figure 1A** most closely resembles the situation in a GWAS, where the prior probability of association to the outcome is very low ($1*10^{-4}$) and power to detect an association may vary from very low (0.25) to high (0.95). As expected based on our experience in GWAS, "significance" defined using p-value thresholds up until 10^{-5} implies a low PPV. However, the PPV rapidly increases when using lower p-value thresholds, even if the power to detect an association is low

(0.25). Significance defined at a p-value threshold of $1*10^{-7}$ or lower results in an excellent PPV (> 99%) even if the study is underpowered (0.25).

For a prior probability of association of 0.10 (**Figure 1D**) we find that a p-value of 0.05 to determine significance reaches a PPV of 67.9% when testing with a power of 0.95. However, when the prior probability of association is lower (e.g. 0.01; **Figure 1C**) the PPV is 16.1%. When testing at a p-value of $1*10^{-5}$ the PPV is over 96% if the prior probability is low ($1*10^{-3}$; **Figure 1B**) and to 100% if the prior probability is 0.01 to 0.50 (**Figure 1C to 1E**).

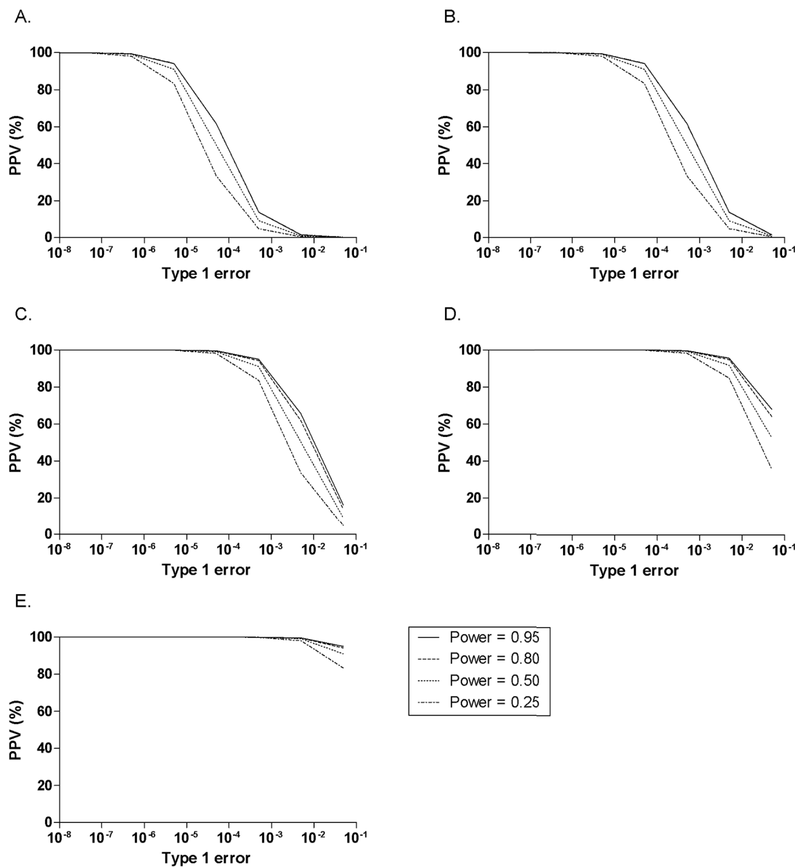


Figure 1 | Positive Predictive Value (PPV) of varying p-value thresholds.

Solid: power is 0.95; short dash: power is 0.80; dotted: power is 0.50; dot-dashed: power is 0.25

1A. PPV of p-value with a fixed prior probability of association of $1*10^{-4}$

1B. PPV of p-value with a fixed prior probability of association of $1*10^{-3}$

1C. PPV of p-value with a fixed prior probability of association of 0.01

1D. PPV of p-value with a fixed prior probability of association of 0.10

1E. PPV of p-value with a fixed prior probability of association of 0.50 obustness of grading over time in AlzGene's top list

Supplementary table 1 displays the top results lists of the AlzGene database from 2005 to 2010. All genes that contained at least one polymorphisms showing at least nominally significant (p-value < 0.05) association in the AlzGene meta-analyses and that were assigned "strong" credibility of association at any point in time between 2005 and 2010 are presented in **Table 1**. In 2005, in the pre-GWAS era, five genes were graded to have strong credibility of association, while in 2010 only two of these five genes (40%) still showed strong credibility of association. At that time point the meta-analysis result for one of the five genes was no longer nominally significant, while two genes were down-graded to show only weak credibility of association. Of the two associations that retained their strong credibility, *APOE* had a p-value of $9.27*10^{-120}$ and *CST3* a p-value of 0.022. It should be noted that the meta-analysis of the ε4-allele of *APOE* is not being updated in AlzGene due to its well-established unequivocal role in AD [17]. Accordingly, the Venice grading of *APOE* has not changed over time. Regarding *CST3*, only two datasets have been added to the meta-analysis since 2005 making the stability of the findings on this gene difficult to interpret. Conversely, no gene increased from weak/moderate to strong credibility grading between 2005 and 2010.
For a total of 12 out of 21 genes that had been graded to show strong credibility of association at any point in time between 2005 and 2010 the credibility had changed from strong to weak or non-significance by 2010 (57%). The lowest p-value for a gene initially classified with strong credibility that subsequently lost its high grading was $1*10^{-4}$ (*SORL1*).

DISCUSSION

In this study, we found a strong association of the p-value to the PPV, implying that the lower the p-value, the higher the probability that a finding is true. Although the p-value has for long been ignored in epidemiology, the findings underscore its relevance in interpretation of the findings of a study. In line with the finding of Weitkunat *et al* we find that when the prior probability of association is below 0.10, the PPV of a finding with a p-value of 0.05 is low even if the statistical power of the study is higher than 0.80. On the other hand, even in an underpowered study (25%), a p-value threshold of $1*10^{-5}$ yields a high PPV (> 96%) if the prior probability of association is at least $1*10^{-3}$. Empirical analyses of the AlzGene database, which uses the epidemiological approach of the Venice criteria (evidence, consistency and bias), show that when using a p-value threshold of 0.05 as a threshold for "significance" upon meta-analysis of candidate genes in at least four independent datasets, still more than half of the findings (12/21) will turn out to be false.

Table 1 | Genes with strong grading over time.

Gene	Polymorphism	Grade 2005	Grade 2006	Grade 2007	Grade 2008	Grade 2009	Grade 2010
APOE	APOE_e2/3/4	A ($9.27*10^{-120}$)	A ($9.27*10^{-120}$)	A ($9.27*10^{-120}$)	A ($9.27*10^{-120}$)	A ($9.27*10^{-120}$)	A ($9.27*10^{-120}$)
TF	rs1049296	A (0.002)	A (0.003)	C (0.012)	C (0.006)	C (0.002)	C (0.002)
SOAT1	rs1044925	A (0.002)	A (0.002)	NR	NR	NR	NR
ACE	rs1800764	A (0.002)	A (0.002)	A (0.002)	A (0.002)	A (0.002)	C ($3.12*10^{-4}$)
CST3	rs1064039	A (0.022)	A (0.022)	A (0.007)	A (0.043)	A (0.043)	A (0.043)
DAPK1	rs4878104	NR	A (0.002)	A (0.002)	C (0.002)	C (0.002)	C (0.002)
IL1B	rs1143634	NR	A (0.013)	A (0.013)	A (0.013)	C (0.021)	C (0.021)
PON1	rs662	NR	A (0.018)	NR	NR	NR	NR
MAPT/STH	rs1467967	NR	A (0.050)	C (0.042)	C (0.037)	NR	NR
SORL1	rs2070045	NR	NR	A ($1.01*10^{-4}$)	B (0.007)	C ($1.74*10^{-4}$)	C ($1.74*10^{-4}$)
GALP	rs3745833	NR	NR	A ($1.20*10^{-4}$)	NR	NR	NR
GWA_14q32.13	rs11622883	NR	NR	A ($3.38*10^{-4}$)	A ($3.38*10^{-4}$)	A ($3.38*10^{-4}$)	A ($3.38*10^{-4}$)
TNK1	rs1554948	NR	NR	A (0.001)	A (0.001)	A (0.001)	A (0.001)
PCK1	rs8192708	NR	NR	A (0.001)	NR	NR	NR
TFAM	rs2306604	NR	NR	A (0.002)	A (0.003)	C (0.036)	C (0.036)
LOC651924	rs6907175	NR	NR	A (0.009)	C (0.005)	C (0.005)	C (0.005)
PVLR2	rs6859	NR	NR	NR	NR	A ($2.44*10^{-28}$)	A ($2.44*10^{-28}$)
CLU	rs11136000	NR	NR	NR	NR	A ($3.06*10^{-16}$)	A ($4.72*10^{-23}$)
PICALM	rs3851179	NR	NR	NR	NR	A ($3.49*10^{-9}$)	A ($5.20*10^{-18}$)
IL8	rs4073	NR	NR	NR	NR	A (0.004)	A (0.004)
LDLR	rs5930	NR	NR	NR	NR	A (0.039)	A (0.039)

Genes with strong grading followed over time. Grading with the Venice Criteria determined at the end of each year. 'A' is strong credibility, 'B' is moderate credibility, 'C' is weak credibility and 'NR' is not-rated due to non-significant meta-analysis. Between brackets is the p-value of association. Note that owing to the established role of the ε4- allele of *APOE* in Alzheimer's disease, the *APOE* meta-analysis is not being updated, but instead as a proof of concept, based on the 43 samples included in the previous meta-analysis by Farrer *et al.* [30].Therefore the Venice grade and significance cannot change over time.

5

The finding that a p-value threshold of 0.05 yields a low PPV when the prior probability of association is low corroborates findings of others [22, 23], and provides one explanation as to why many candidate gene studies using this p-value threshold were not replicated. This finding also extends to non-genetic studies of epidemiological risk factors. The use of p-values in epidemiological studies has been a matter of debate for long [24]. However, p-values are important determinants of the PPV. If the prior probability of association is high (e.g. 0.50) the PPV of a p-value threshold of 0.05 is nearly 90% when the power of the study is 0.80. A high prior probability can be anticipated for phase III clinical trials of new medical compounds, which are usually conducted after extensive research of the compounds starting from cell cultures to animal models and early clinical studies in humans. One may argue that the industry will not embark on phase III clinical trials unless the prior probability of success is high (> 0.50). In observational studies, high prior probabilities could, for example, be assumed for genetic association studies that aim to replicate the association of SNPs that were originally identified by GWAS with genome-wide significance (p-value < $1*10^{-7}$).

In the context of a very low prior probability of association, e.g. a GWAS that aims to discover novel genes with prior probabilities as low as $1*10^{-4}$ when a limited number of variants (n = 50) determines the trait, it is clear that only very low p-values (< $1*10^{-7}$) yield a high PPV. This corroborates the threshold of $1*10^{-7}$ or $5*10^{-8}$ that is commonly used in GWAS based on multiple comparison adjustment using Encode sequence data [12]. The recent finding that most GWAS-derived associations with p-values between $1*10^{-7}$ and $5*10^{-8}$ are very often replicated in subsequent studies also corroborates with this finding [25].

The results of our study may also be relevant for other "-omics" studies in which the general aim is discovery without a specific hypothesis. Thousands of molecules and their ratios are studied for which almost no prior proof of association to any trait exist. The prior probability for findings in these studies to be true will be very low. In this context it is interesting that findings from such studies have also been difficult to replicate in the past [26, 27]. It should be noted, that with the PPV we do not aim to determine whether a finding of association is true or false but rather aim to give a probability that a finding is true or false. In hypothesis-free discovery studies we are never able to discriminate between true positive and false positive results. However, some findings are more likely to be true positives, as measured by the PPV [3].

Though not explicitly tested, using the formula for PPV, for genome wide interaction studies, we can extrapolate that the significance threshold needed is probably substantially lower as the prior probability of association is substantially lower than in classical GWAS. For example: if we study the main effects of 500,000 SNPs in a GWAS, in an interaction study (combination of 2 SNPs only) we test (500,000*500,000)/2 – 500,000 comparisons. This will reduce the prior probability considerably. With a prior probability of $1*10^{-6}$ a PPV of 95% can

still be reached with a p-value threshold of $5*10^{-8}$, even when the power is low. However, if the prior probability is $1*10^{-7}$ a PPV of 95% is reached only with a p-value threshold of $5*10^{-9}$. A key issue in the estimation of PPV (or FPRP) is the need for the specification of the prior probability of association. There will be many instances where researchers will have little knowledge of the prior probability of association and in these situations, choosing an appropriate p-value threshold may prove to be difficult or even impossible. This is particularly true for candidate gene studies where genes are typically selected based on their potential biochemical/functional implication. However, even though the biochemical involvement may be true, genetic variations in the genes may not be relevant to pathogenesis if the levels of the protein are predominantly determined by environmental factors. In addition, studies that fine-map GWAS regions may fall in this category as even though the region has been implicated it usually includes hundreds to thousands of variants, most of which are not associated causally [28]. The prior probability that such a variant is causally related to the disease is therefore still small, but considerably higher compared to the discovery GWAS. Here a Bonferroni correction is often conservative, as many variants will be strongly correlated to each other. A similar problem occurs in epidemiological research when findings of animal studies are translated to humans [29]. The present study shows that over a wide range of prior probabilities ($1*10^{-3}$ – 0.50) and power (0.25 – 0.95), a p-value of $1*10^{-5}$ yields a high PPV (> 96%). These conclusions are supported by meta-analysis based on genetic association data as posted on AlzGene. Longitudinal evaluation of the AlzGene results by the Venice criteria also supports a more conservative p-value threshold of $1*10^{-5}$ to define “significance” for candidate associations. The lowest p-value of a marker that lost its significance was 10^{-4} for *SORL1*.

The advantage of using a p-value threshold rather than the PPV or FPRP is that it is independent of a prior probability. Although one should consider the p-value as continuous, many journals will want a threshold to determine which associations are true. The general threshold of $1*10^{-5}$ could be used for this. However, it may be too conservative in some instances (e.g. clinical trials or replication of GWAS) or too lenient in the case of gene discovery (e.g. GWAS). We therefore suggest two additional thresholds for high prior probability (> 0.50) and extremely low prior probability (< $1*10^{-4}$). These exploratory suggestions for p-value thresholds are summarized in **Table 2**.

The low p-value thresholds needed to define “significance” imply the necessity of large datasets in genetics and epidemiology, in particular when the effect estimates are expected to be small. An alternative approach in order to be able to employ less stringent significance criteria to the initial dataset is to validate the results in a replication phase as is common practice in GWAS designs [10].

We have shown that the p-value is an important determinant of the PPV, i.e. of the probability that a finding is true. Whenever there is a high prior probability of association (> 0.50) a p-value threshold of 0.05 may be sufficient, while for studies with a low prior probability

(< $1*10^{-4}$) a threshold of $1*10^{-7}$ is more appropriate. Perhaps the most important finding of our study is that over a wide range of prior probabilities ($1*10^{-3}$ – 0.50) and power (0.25 – 0.95), a p-value of $1*10^{-5}$ yields a high PPV (> 96%). Similar to the use of the 0.05 threshold at present, the threshold of $1*10^{-5}$ may be used for evaluating the significance of findings in candidate gene studies.

Table 2 | Suggestions for context specific p-value thresholds for significance.

Context	**High prior probability of association (> 0.50)**	**Unknown prior probability of association**	**Very low prior probability of association (< $1*10^{-4}$)**
Example	1. Clinical trial 2. Replication of significant GWAS finding ($1*10^{-7}$)	1. Candidate gene study 2. Translating findings from model organisms to humans in epidemiology	1. GWAS 2. Other "-omics", like metabolomics and lipidomics
P-value threshold	0.05	$1*10^{-5}$	$1*10^{-7}$

GWAS = genome wide association study

REFERENCES

1. Hirschhorn, J.N., et al., A comprehensive review of genetic association studies. Genet Med, 2002. 4(2): p. 45-61.
2. Abou-Sleiman, P.M., M.G. Hanna, and N.W. Wood, Genetic association studies of complex neurological diseases. J Neurol Neurosurg Psychiatry, 2006. 77(12): p. 1302-4.
3. Ioannidis, J.P., Why most published research findings are false. PLoS Med, 2005. 2(8): p. e124.
4. Wacholder, S., et al., Assessing the probability that a positive report is false: an approach for molecular epidemiology studies. J Natl Cancer Inst, 2004. 96(6): p. 434-42.
5. Weitkunat, R., et al., Effectiveness of strategies to increase the validity of findings from association studies: size vs. replication. BMC Med Res Methodol, 2010. 10: p. 47.
6. Lucke, J.F., A critique of the false-positive report probability. Genet Epidemiol, 2009. 33(2): p. 145-50.
7. Matullo, G., M. Berwick, and P. Vineis, Gene-environment interactions: how many false positives? J Natl Cancer Inst, 2005. 97(8): p. 550-1.
8. Rebbeck, T.R., et al., SNPs, haplotypes, and cancer: applications in molecular epidemiology. Cancer Epidemiol Biomarkers Prev, 2004. 13(5): p. 681-7.
9. Thomas, D.C. and D.G. Clayton, Betting odds and genetic associations. J Natl Cancer Inst, 2004. 96(6): p. 421-3.
10. Moonesinghe, R., M.J. Khoury, and A.C. Janssens, Most published research findings are false-but a little replication goes a long way. PLoS Med, 2007. 4(2): p. e28.
11. Ioannidis, J.P., et al., Assessment of cumulative evidence on genetic associations: interim guidelines. Int J Epidemiol, 2008. 37(1): p. 120-32.
12. International HapMap, C., The International HapMap Project. Nature, 2003. 426(6968): p. 789-96.
13. Benjamini, Y. and Y. Hochberg, Controlling the False Discovery Rate: A practical. J.R. Statist. Soc. B, 1995. 57(1): p. 289-300.
14. Gordon, A., et al., Control of the mean number of false discoveries, Bonferroni and stability of multiple testing. Annals of Applied Statistics, 2007. 1: p. 179-190.
15. Khoury, M.J., et al., Genome-wide association studies, field synopses, and the development of the knowledge base on genetic variation and human diseases. Am J Epidemiol, 2009. 170(3): p. 269-79.
16. Ridley, J., et al., An unexpected influence of widely used significance thresholds on the distribution of reported P-values. J Evol Biol, 2007. 20(3): p. 1082-9.
17. Bertram, L., et al., Systematic meta-analyses of Alzheimer disease genetic association studies: the AlzGene database. Nat Genet, 2007. 39(1): p. 17-23.
18. Higgins, J.P., et al., Measuring inconsistency in meta-analyses. BMJ, 2003. 327(7414): p. 557-60.
19. DerSimonian, R. and N. Laird, Meta-analysis in clinical trials. Control Clin Trials, 1986. 7(3): p. 177-88.
20. Harbord, R.M., M. Egger, and J.A. Sterne, A modified test for small-study effects in meta-analyses of controlled trials with binary endpoints. Stat Med, 2006. 25(20): p. 3443-57.
21. Kavvoura, F.K., et al., Evaluation of the potential excess of statistically significant findings in published genetic association studies: application to Alzheimer's disease. Am J Epidemiol, 2008. 168(8): p. 855-65.
22. Ioannidis, J.P., Calibration of credibility of agnostic genome-wide associations. Am J Med Genet B Neuropsychiatr Genet, 2008. 147B(6): p. 964-72.
23. Ioannidis, J.P., R. Tarone, and J.K. McLaughlin, The false-positive to false-negative ratio in epidemiologic studies. Epidemiology, 2011. 22(4): p. 450-6.
24. Rothman, K.J., Epidemiology: An Introduction. 1 ed. 2002: Oxford University Press. 221.
25. Panagiotou, O.A., J.P. Ioannidis, and P. for the Genome-Wide Significance, What should the genome-wide significance threshold be? Empirical replication of borderline genetic associations. Int J Epidemiol, 2011.

26. Kaiser, J., Biotechnology. Researcher, two universities sued over validity of prostate cancer test. Science, 2009. 325(5947): p. 1484.
27. van Duijn, C.M., STROBE-ME too! Eur J Epidemiol, 2011. 26(10): p. 761-762.
28. Sanna, S., et al., Fine mapping of five Loci associated with low-density lipoprotein cholesterol detects variants that double the explained heritability. PLoS Genet, 2011. 7(7): p. e1002198.
29. Boseley, S., Six men in intensive care after drug trial goes wrong, in The Guardian. 2006.
30. Farrer, L.A., et al., Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease. A meta-analysis. APOE and Alzheimer Disease Meta Analysis Consortium. JAMA, 1997. 278(16): p. 1349-56.

Chapter 5.2

Findings of this thesis and future research

In this thesis we aimed to identify genetic markers associated with longevity in genome-wide association studies (GWAS) and candidate gene studies on the one hand. On the other hand we aimed to gain a better understanding of telomere length and its relation to longevity and mortality.

FINDINGS OF GENETIC RESEARCH IN LONGEVITY

We performed a GWAS on the longevity phenotype, yet failed to identify any new consistent associations (**Chapter 2.1**). In **Table 1** we summarize the GWAS previously performed on longevity. The only consistent finding is *APOE*, originally a candidate gene [1], which was found to be associated with longevity at genome-wide significant level in half of the GWAS performed to date [2-7].

Table 1 | Summary of previous genome-wide association study findings.

Author	Year	Phenotype definition	Sample size	Main finding(s)†
Newman, AB	2010	Longevity: 90+ Comparison: died between 55-80	1,836 cases 1,955 controls	*MINPP1* ($6.8*10^{-7}$)
Deelen, J	2011	Longevity: 85+ Comparison: middle age	4,149 cases 7,582 controls	*TOMM40*** ($3.4*10^{-17}$)
Nebel, A	2011	Longevity: 90+ Comparison: middle age	763 cases 1,085 controls	*APOC1*** ($1.8*10^{-10}$)
Malovini, A	2011	Longevity 90+ Comparison: 18-45	582 cases 784 controls	*CAMKIV* ($1.7*10^{-6}$)
Walter, S	2011	Survival: all-cause mortality Follow-up: 10.6 (5.4) years	25,007 total (8,444 deaths)	*OTOL1* ($1.6*10^{-6}$)
Sebastiani, P	2012	Longevity 100+ Comparison: middle age	801 cases 914 controls	*TOMM40*** ($2.4*10^{-10}$)

†Gene-name and p-value reported in the publication; **In high LD with *APOE*

In a recent and currently unpublished GWAS performed by Deelen *et al* in the European Longevity Consortium the *APOE* region was confirmed and a new region (5q33) was found to be associated with longevity (personal communication). The study of Deelen *et al* consisted of 5,406 longevity cases (90+) and 16,121 younger controls (< 65 years of age). We could not replicate the newly identified variant (5q33) in our own GWAS results. Compared to our study in **Chapter 2.1**, Deelen *et al* had a similar number of longevity cases (6,036 in our study) but more controls (3,757). In **Chapter 2.1** we chose to only include controls deceased before 80 years of age to increase the contrast between cases and controls. However, frequency distributions of suggestively associated variants in both studies across

different age-categories reveal that the allele frequencies are stable until 85 years of age. For example, the identified variant in *FOXO3* has a minor allele frequency of 0.28 in all age-categories below 85 years of age, while in those older than 85 years of age the allele frequency is 0.34. This suggests that the differences in findings between **Chapter 2.1** and Deelen *et al* are not explained by the differences in controls. At this point the novel finding of Deelen *et al* (5q33) may be a false positive. A future joint-analysis enlarging the case sample will be performed to elucidate the association. The stable allele frequency until age 85 suggests that restricting our control group to individuals who have died is not necessary. Therefore, in order to increase our control population in the future joint-analysis our control definition will include all individuals younger than 65 years of age at baseline, as is the case in the current GWAS of Deelen *et al*.

Despite the lack of success of GWAS in identifying genetic variants associated with longevity, our investigation of the total variance explained by common variants tagged by GWAS shows that over 80% of the heritability can be explained by these variants (**Chapter 2.2**). This strongly suggests that very many variants all with very small effects are involved in longevity following the infinitesimal model of genetic contribution to common traits. The infinitesimal model suggests that eventually nearly every gene, and also variants in regulatory elements, contribute to longevity, mostly with effect sizes that are so small that it would take very large study sample sizes to detect them [8, 9]. For example, detecting an odds ratio of 1.1 with an allele frequency ranging from 0.05 to 0.50 would require 34,906 to 175,931 longevity cases. This number will only increase as the effect size decreases. Increasing the sample size beyond that which was used in this thesis (6,036 cases and 3,757 controls; **Chapters 2.1**) may help identifying some of these variants with larger effects. However, another avenue may be to consider other definitions of the longevity phenotype. It has been argued that longevity defined as 90+ is not a sufficiently extreme age for gene discovery. Centenarians (100+) should be used instead [6]. However, our finding that over 80% of the heritability of longevity (90+) can be explained by common variants contradicts this. After all, in order to become a centenarian you must first reach the age of 90. Variants associated to centenarian status should also be associated to reaching the age of 90+. Though studies focusing on centenarians exist, only three have genome-wide association data allowing hypothesis-free testing of the whole genome [6, 10]. Moreover, due to the phenotype (100+) being rare, these studies are small (less than 1,000 cases) which hinders genetic discoveries as well [6, 10]. In population-based settings very few individuals included in the study are expected to reach centenarian status, which is why the longevity definition of 90+ is more commonly used (**Table 1**) [11]. For example, in the Rotterdam Study 1,004 individuals have reached 90 years of age since the start of the study over 20 years ago. Of these less than 5% (46) have reached centenarian status. Additionally, it is known that the longevity phenotype is very heterogeneous even in centenarians, namely consisting of the survivors (onset of aging disease < 80 years), delayers (onset of aging

disease between 80 and 100 years) and escapers (onset of aging disease > 100 years) [12]. This heterogeneity is expected to be worse in younger age cut-offs for longevity, like 90+, but is present in centenarians as well. Such heterogeneity makes identifying new genetic associations challenging, though not impossible as variants are being found in diseases like type 2 diabetes which is also known to be heterogenic [13, 14].

A previously suggested approach to solve the heterogeneity issue in longevity could be the development of novel phenotypes, such as the healthy aging index (HAI) [15]. The HAI uses markers of 5 major organ systems that are known to predict mortality and disability. The systems included are vascular (carotid wall thickness), brain (white matter grade on MRI), kidney (cystatin-c), lung (forced vital capacity) and metabolic (fasting glucose levels) [15]. The HAI is able to distinguish a wide risk gradient, but is most remarkable for its advantage in identifying low risk individuals. As a single factor, the HAI prediction of mortality is similar in magnitude to age itself. When investigated together, age remained partly independent, but the HAI explained 40% of the effect of age [15].

Another phenotype previously explored is all-cause mortality [7]. As prospective follow-up studies with a continuous outcome are more powerful than case-control analyses, this might provide new insight in the phenotype. Though this design has been successfully applied in animal models [16] and in human genetic research of blood pressure [17], no new genome-wide significant associations were identified in the GWAS by Walter *et al* [7].

Another avenue to consider is further investigation of rare variants. Though most of the heritability of longevity can be explained by common variants, the high recurrence rate of longevity in some families suggests the involvement of rare genetic variants in these families [18]. Next Generation Sequencing (NGS) may help us uncover such variants [19]. However, we could not identify such variants in the Rotterdam Study by sequencing the exomes of 202 longevity cases and 267 controls at an average depth of 45x. Suggestive evidence for the involvement of rare variants in *WSCD2* in longevity was observed (**Chapter 2.2**). Possible reasons for our failure to identify rare variants are characteristics of our sample, which was small and not pre-selected for long-lived families in which rare variants are expected to play a larger role, and limiting sequencing to only the exome, that is about 1% of the entire human genome. Even though rare variants are unlikely to explain the variability of the phenotype in the general population, such variants can give us important clues to genes and biological pathways involved in longevity.

HEAT SHOCK PROTEINS IN RELATION TO AGE-RELATED NEURODEGENERATIVE DISEASES AND LONGEVITY

In animal studies the insulin pathway appears to be the key pathway determining mortality and longevity (see **Chapter 1**). Studies in *Caenorhabditis elegans* and *Drosophila melanogaster* suggest that heat shock proteins (HSPs) may play a role in lifespan [20, 21]. HSPs play in

particular a role in protein folding and have been implicated in neurodegenerative disorders that are characterized by protein misfolding [22-24].

In this thesis we performed a comprehensive search for the involvement of common genetic variation in HSPs in common age-related neurodegenerative diseases and longevity. Previous candidate gene studies investigating HSPs only tested a limited number of SNPs in a limited number of genes [25-27]. First, we aimed to elucidate the association of these genes with Alzheimer's Disease (AD) and Parkinson's Disease (PD). These are two diseases in which protein folding is known to be involved and are therefore of interest in relation to aging and longevity (**Chapters 3.1** and **3.2**).

Pathological research has implicated several HSP families in AD including HSP70 and prefoldin (see **Chapter 3.1**). In this thesis, we studied a total of 79 genes from the HSP70 family and its major co-chaperones, including HSP40 and prefoldin, in 462 patients and 5,238 controls from the Rotterdam study. We found prefoldin 2 (*PFDN2*) to be associated with AD. The minor allele of rs12118313 was associated with increased AD risk, with an odds ratio (OR) of 1.19. Further, pathway analysis pointed to an overall involvement of the prefoldin family of genes (p-value = 0.03; see **Chapter 3.1**). The findings were not replicated in the ongoing GWAS of AD [28, 29]. This raises the question whether the prefoldin association with AD as reported in **Chapter 3.1** are a false or true positive. Our simulation studies in the **Chapter 5.1** may shed light on the question. The p-value for the association is as high as 0.003. As discussed in **Chapter 5.1**, findings with such a high p-value have a low positive predictive value. For the *PFDN2* finding the positive predictive value ranges between 40 and 80% depending on the prior probability of association, which is difficult to estimate. The same argument holds for the findings of the pathway analyses which were not replicated in the recent genome wide association studies of AD [28, 29]. Although the HSP10, HSP60, HSP70 and HSP90 families play a key role in the pathogenesis of PD, our genetic studies of the genes encoding for these protein are negative. In total we studied 3,828 variants in 27 genes in 134 patients and 5,422 controls. We could not find any statistical evidence for HSP gene involvement (**Chapter 3.2**). Also in GWAS of PD HSP genes have not emerged [30].

Next we studied the relation of HSP70, small HSP (sHSP) and Heat Shock Factor (HSF) with all-cause mortality as a major determinant of longevity (**Chapter 3.3**). We chose for all-cause mortality as definition for longevity in this study as it has been suggested that using the entire age-range of deaths is a more powerful approach in gene discovery compared to a dichotomization of the outcome as has been used previously [7]. As discussed above, this approach allows us to maximize the statistical power of research using the information of all participants (N= 5,974). In total we studied 4,430 markers in 31 genes. Of these only *HSF2* was associated with mortality at a p-value of 0.003. In the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) we could replicate the finding in 8 cohorts. However in the joint analyses including 25,007 subjects we still obtained a p-value as high

as 0.003 yielding a positive predictive value of 40 to 80% depending on the prior probability of association.

Recently evidence has come to light that the mitochondrial unfolded protein response, which is mitigated by HSP60 and HSP70, is involved in longevity in animal models, such as *C. elegans* and *Mus Musculus* [31]. Whether this can be translated to humans is currently unknown. HSP complexity is increased in humans as compared to lower organisms [32]. However, in the mitochondria only HSP70 member *HSPA9* or mortalin [33], and the chaperonin complex (consisting of *HSP60* and *HSP10*) [34] are expressed in both lower organisms as well as humans. Interestingly, mortalin has been associated with longevity in *C. elegans* previously [35]. Again we did not find any evidence for the involvement of this gene in longevity in humans in the GWAS (**Chapter 2.1**) or the HSP targeted analyses (**Chapter 3.3**). Rare variants rather than common variants in these genes may be interesting as these variants are expected to have stronger effects on the phenotypes. In our exome study (**Chapter 2.2**) we identified 40 exome variants in *HSPA9*, 38 in *HSP60* and 6 in *HSP10*. Burden test analysis did not reveal an association of these genes with longevity (p-value > 0.20).

From a methodological perspective it should be noted that difficulties of defining the longevity phenotype as discussed earlier also hamper the findings of theses analyses and have been recognized in the literature [11]. HSPs are emerging in GWAS of other traits like chronic widespread pain [36] and lean body mass (confidential). Both pain and loss of lean mass (muscle) are associated with mortality [11]. Thus, systematic investigation of other traits associated with longevity may elucidate the role of HSP genes in morbidity and mortality.

TELOMERE LENGTH DISCOVERIES

Although the sequence of telomere length and the discovery of telomerase, the enzyme that can rebuild telomeres, have led to the awarding of the Nobel prize (2009; Elizabeth Blackburn, Carol Greider and Jack Szostak), little is known about the biological processes and environmental factors that influence telomere length. The only well-replicated associations with telomere length in the literature are age, sex and genetic factors [37, 38]. Many other associations, including smoking, diet, body fat, leptin, adiponectin and blood pressure, have been proposed but never consistently associated [39-45]. Problems with the preferred telomere length measurement technique (gel electrophoresis Telomere Restriction Fragment) contributed to the non-replication. This technique has a high cost with regard to both DNA and money, resulting in small sample sizes [46] and consequently results in false positive and false negative findings.

With the new qPCR method to estimate (relative) telomere length based on the copy number of a single copy gene (usually 36B4) it has become feasible to measure telomere length in large cohorts and at multiple time points [46]. In this thesis, we made a start

at untangling previously suggested associations with telomere length using the qPCR method for telomere length measurement. In our first study we determined the heritability (**Chapter 4.1**). As previously reported, the heritability is high (70%; see **Chapter 4.1**). We also observed a mother-offspring (0.42) and father-offspring (0.33) correlation. That this did not differ between male and female children, suggests that this maternal effect on telomere length determination is not X-chromosomal in nature. In **Chapter 4.1** we found a significant positive paternal age effect on telomere length determination, i.e., children of older fathers have longer telomeres. This may be related to high telomerase levels in sperm causing telomere length to be maintained, and even lengthened, despite both mitotic and meiotic divisions during the proliferation and differentiation of the sperm cells [47]. This is of interest as increasing paternal age has an increased risk of new mutations which previously have been associated to autism [48].

Halaschek-Wiener *et al* reported a reduced variance in telomere length in the elderly compared to midlife controls [49]. They attributed this to the risk associated with both long and short telomeres, suggesting there may be an optimal telomere length that is protective for healthy aging. We investigated whether there is a reduced variance in telomere length in the elderly due to a survivor effect in telomere length, as this could explain the inconsistent results for the association between telomere length and mortality (**Chapter 4.2**). We found a significant reduction in telomere length variance in the population from young adulthood to old age, which was stronger in men compared to women. Additionally we found that the (grand)children of the oldest-old had significantly longer telomeres compared to others in the same age-range, which suggests that the oldest-old started life with long telomeres. This finding is relevant when interpreting the inconsistent findings with longevity and mortality [40, 50-59] as telomere length measured at old age is not representative for those at the start of life. In studies conducted late in life not only the telomere length is not representative anymore for early life but the variance is significantly reduced making it statistically difficult to identify associations.

We also find an association of telomere length with height. A male-specific association has been observed in two separate animal species previously (**Chapter 4.3**). We also find a significant inverse association between telomere length and height in males ($r = -0.08$; p-value = $5.8*10^{-4}$), but not in females ($r = -0.02$; p-value = 0.302). Results from a Mendelian randomization analysis are compatible with the hypothesis that differences in height cause differences in telomere length, as several genes involved in height are associated with telomere length in the inverse direction.

Finally, we investigated associations with metabolic traits and metabolome-wide data. We first studied adiponectin [39, 45, 60] and leptin [39, 42, 44, 45, 61, 62] for which inconsistent associations with telomere length have been reported. We investigated the association of both adiponectin and leptin with telomere length in 7 studies including 11,460 participants of the European Network for Genetic and Genomic Epidemiology (ENGAGE) consortium

(**Chapter 4.4**). We identified a highly significant association between leptin and telomere length ($r = -0.05$; p-value = $1.81*10^{-7}$), but not between adiponectin and telomere length ($r = 0.004$, p-value = 0.763). A full examination of the association of 163 metabolomic traits with telomere length is presented in **Chapter 4.5**. Previously a set of 22 age-associated metabolites has been shown to explain 5% of the total telomere length variation in the population by Menni *et al* [63]. We identified six metabolites that implicated the involvement of homocysteine, lipid and thyroid metabolism in telomere length biology (**Chapter 4.5**). Menni *et al* used a different approach to measure metabolites, which makes a one on one comparison between the two studies impossible. However, several lipid metabolites were identified in the study by Menni *et al* [63], supporting our findings.

Recent GWAS has provided seven genetic loci to be associated with telomere length, but the very high heritability (h^2 = 70%) remains unexplained [38]. Also the intriguing maternal aspect to telomere length inheritance we observed remains to be identified [37]. The investigation of mitochondrial DNA variations as well as anything expressed *in utero* will be interesting to investigate in this sense. Additionally, NGS shows promise to further unravel the high heritability of telomere length. NGS technology captures most variations in the genome including the variants which are neither well represented on the regular genome-wide genotyping arrays nor can be imputed with confidence [19]. Genetic research on telomere length has been focused on common variation, following the common trait-common variant hypothesis [64]. However, as shown in other highly heritable traits such as height, which can be explained by the infinitesimal model of genetic contribution to common traits, also rare variants are expected to play a role in telomere length [9, 65]. Sequencing may unravel the rare/ individual specific/population specific variants, to identify which also large samples will be required. Given the ongoing large scale sequencing efforts in the world and the developments of tools for the analysis and pooling of such data, analyzing rare variants seems to be the most logical next step.

So far, most research has focused on single measurement of telomere length. Our study of telomere length variation presented in **Chapter 4.2** suggests that this may not be representative. Recent studies show that telomere length changes over time might be more informative and has an independent effect of telomere length at baseline in the determination of mortality and other age-related diseases [54]. Here again studies are struggling with small sample sizes and lack of samples for replication. To truly unravel telomere length in an epidemiological setting, more studies with repeated telomere length measurements are needed. Existing large follow-up studies should consider investing in telomere length, both as an outcome and as a determinant of major age-related diseases and longevity.

Currently, most research on telomere length is focused on older individuals with only a few studies on adolescents and children [66, 67]. Our study on the reducing telomere length variation with age, suggests that telomere length is already selected on in young adulthood

(**Chapter 4.2**). A question to be answered is if such a trend starts even earlier in life. Though we do not see a change in telomere length variation in the population between 20 and 40 years of age, we cannot exclude that a change in telomere length variation before the age of 20 has occurred. Understanding telomere length dynamics at young age is vitally important if we want to understand the influence of these dynamics on age-related traits and longevity. Early life events may predispose to telomere shortening. Understanding these dynamics will be important for future preventive studies and translation.
However there are hindrances to overcome. Maintaining telomere length is important for maintaining cells, however early experiments increasing telomerase levels in order to maintain telomere length in mice resulted in high cancer rates [68]. Also in humans telomerase is activated in 90% of cancerous cells, maintaining telomere length despite the large number of population doublings [69]. Finding out how this delicate balance works will be an important step in increasing longevity.

OMICS AND MULTIPLE TESTING

Although not a main point of this thesis, we briefly touched on metabolomics in relationship to telomere length in **Chapter 4.5.** As more studies obtain metabolomics data, large scale research into this field becomes possible, yet very challenging. Running a GWAS on 130+ metabolic traits is computationally challenging, although software to handle this data is in development (http://www.genabel.org/packages/OmicABEL).
Most of the metabolites measured in metabolomics are still poorly understood. Although genetic associations are uncovered, if we ever want to make full use of this technology it will be vital to better understand these metabolites and what effects they have in disease and in health [70]. Also here challenges, like the influence of medication and lifestyle, will need to be overcome.
A final challenge with metabolomics and indeed all "-omics" is multiple testing correction. In **Chapter 5.1** we show, based on simulations and the Alzgene database, that independent of power (related to sample size, frequency and effect size) and prior probability of association, a p-value of $1*10^{-7}$ gives an excellent posterior probability of association. In GWAS this threshold is already commonly used. For metabolome-wide association studies (MWAS) the prior probability of association will also be low, leading to a strong likelihood of false positives. However, as these metabolites are highly correlated, we might not have to be as stringent as in GWAS.
Currently, to adjust for multiple testing 3 strategies are used in literature, namely 1) Bonferroni correction; 2) PCA adjustment; 3) adjustment for number of independent variables. The first strategy, where one adjusts for the exact number of metabolites tested, will be very conservative due to the high correlation of the metabolites. The second strategy is based on principal component analysis and adjusts for the number of significant principal

components explaining a given percentage of the variance. This strategy is scientifically not correct as you do not *a priori* know whether the metabolites one will deem to be significant, are explained well by the principal components you adjust for. It might be that the significant finding is not properly captured by the multiple testing correction. The third strategy, which we used in this thesis, estimates the number of independent variables to be adjusted for based on the eigenvalues of a correlation matrix. This approach was first proposed for genetic associations where linkage disequilibrium creates a similar problem [71]. More research will be needed to come to a consensus within the metabolomics research community to determine a strategy to adjust for multiple testing.

CONCLUDING REMARKS

In this thesis we have performed a comprehensive investigation of candidate genes (HSPs) as well as a GWAS to identify genetic variants associated with longevity, yet we did not reveal any new associations. Despite this lack of success over 80% of the heritability of longevity can be explained by very many common variants with very small effects. Given this lack of success it may be more fruitful to focus on telomere length as a phenotype in the near future. Telomere length has yielded genome-wide significant results [38] and the telomere length maintenance pathway has been shown to be associated with longevity previously [72]. In my opinion, future research should therefore focus on telomere length and in particular on telomere length change over time.

REFERENCES

1. Schachter, F., et al., Genetic associations with human longevity at the APOE and ACE loci. Nat Genet, 1994. 6(1): p. 29-32.
2. Deelen, J., et al., Genome-wide association study identifies a single major locus contributing to survival into old age; the APOE locus revisited. Aging Cell, 2011. 10(4): p. 686-98.
3. Malovini, A., et al., Association study on long-living individuals from Southern Italy identifies rs10491334 in the CAMKIV gene that regulates survival proteins. Rejuvenation Res, 2011. 14(3): p. 283-91.
4. Nebel, A., et al., A genome-wide association study confirms APOE as the major gene influencing survival in long-lived individuals. Mech Ageing Dev, 2011. 132(6-7): p. 324-30.
5. Newman, A.B., et al., A meta-analysis of four genome-wide association studies of survival to age 90 years or older: the Cohorts for Heart and Aging Research in Genomic Epidemiology Consortium. J Gerontol A Biol Sci Med Sci, 2010. 65(5): p. 478-87.
6. Sebastiani, P., et al., Genetic signatures of exceptional longevity in humans. PLoS One, 2012. 7(1): p. e29848.
7. Walter, S., et al., A genome-wide association study of aging. Neurobiol Aging, 2011. 32(11): p. 2109 e15-28.
8. Gibson, G., Hints of hidden heritability in GWAS. Nat Genet, 2010. 42(7): p. 558-60.
9. Gibson, G., Rare and common variants: twenty arguments. Nat Rev Genet, 2011. 13(2): p. 135-45.
10. Atzmon, G., et al., Evolution in health and medicine Sackler colloquium: Genetic variation in human telomerase is associated with telomere length in Ashkenazi centenarians. Proc Natl Acad Sci U S A, 2010. 107 Suppl 1: p. 1710-7.
11. Murabito, J.M., R. Yuan, and K.L. Lunetta, The search for longevity and healthy aging genes: insights from epidemiological studies and samples of long-lived individuals. J Gerontol A Biol Sci Med Sci, 2012. 67(5): p. 470-9.
12. Evert, J., et al., Morbidity profiles of centenarians: survivors, delayers, and escapers. J Gerontol A Biol Sci Med Sci, 2003. 58(3): p. 232-7.
13. Cho, Y.S., et al., Meta-analysis of genome-wide association studies identifies eight new loci for type 2 diabetes in east Asians. Nat Genet, 2012. 44(1): p. 67-72.
14. Voight, B.F., et al., Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. Nat Genet, 2010. 42(7): p. 579-89.
15. Newman, A.B., et al., A physiologic index of comorbidity: relationship to mortality and disability. J Gerontol A Biol Sci Med Sci, 2008. 63(6): p. 603-9.
16. Kenyon, C.J., The genetics of ageing. Nature, 2010. 464(7288): p. 504-12.
17. Levy, D., et al., Genome-wide association study of blood pressure and hypertension. Nat Genet, 2009. 41(6): p. 677-87.
18. Westendorp, R.G., et al., Nonagenarian siblings and their offspring display lower risk of mortality and morbidity than sporadic nonagenarians: The Leiden Longevity Study. J Am Geriatr Soc, 2009. 57(9): p. 1634-7.
19. Cirulli, E.T. and D.B. Goldstein, Uncovering the roles of rare variants in common disease through whole-genome sequencing. Nat Rev Genet, 2010. 11(6): p. 415-25.
20. Morrow, G., et al., Overexpression of the small mitochondrial Hsp22 extends Drosophila life span and increases resistance to oxidative stress. FASEB J, 2004. 18(3): p. 598-9.
21. Walker, G.A., et al., Heat shock protein accumulation is upregulated in a long-lived mutant of Caenorhabditis elegans. J Gerontol A Biol Sci Med Sci, 2001. 56(7): p. B281-7.
22. Magrane, J., et al., Heat shock protein 70 participates in the neuroprotective response to intracellularly expressed beta-amyloid in neurons. J Neurosci, 2004. 24(7): p. 1700-6.
23. Muchowski, P.J. and J.L. Wacker, Modulation of neurodegeneration by molecular chaperones. Nat Rev Neurosci, 2005. 6(1): p. 11-22.

24. Sherman, M.Y. and A.L. Goldberg, Cellular defenses against unfolded proteins: a cell biologist thinks about neurodegenerative diseases. Neuron, 2001. 29(1): p. 15-32.
25. Kowalczyk, M., et al., Heat shock protein 70 gene polymorphisms are associated with paranoid schizophrenia in the Polish population. Cell Stress Chaperones, 2013.
26. Singh, R., et al., Heat-shock protein 70 genes and human longevity: a view from Denmark. Ann N Y Acad Sci, 2006. 1067: p. 301-8.
27. Zhang, Y., et al., Genetic polymorphisms of HSP70 in age-related cataract. Cell Stress Chaperones, 2013.
28. Hollingworth, P., et al., Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. Nat Genet, 2011. 43(5): p. 429-35.
29. Naj, A.C., et al., Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. Nat Genet, 2011. 43(5): p. 436-41.
30. Simon-Sanchez, J., et al., Genome-wide association study reveals genetic risk underlying Parkinson's disease. Nat Genet, 2009. 41(12): p. 1308-12.
31. Houtkooper, R.H., et al., Mitonuclear protein imbalance as a conserved longevity mechanism. Nature, 2013. 497(7450): p. 451-7.
32. Kampinga, H.H., et al., Guidelines for the nomenclature of the human heat shock proteins. Cell Stress Chaperones, 2009. 14(1): p. 105-11.
33. Wadhwa, R., K. Taira, and S.C. Kaul, An Hsp70 family chaperone, mortalin/mthsp70/PBP74/Grp75: what, when, and where? Cell Stress Chaperones, 2002. 7(3): p. 309-16.
34. Mayhew, M., et al., Protein folding in the central cavity of the GroEL-GroES chaperonin complex. Nature, 1996. 379(6564): p. 420-6.
35. Yokoyama, K., et al., Extended longevity of Caenorhabditis elegans by knocking in extra copies of hsp70F, a homolog of mot-2 (mortalin)/mthsp70/Grp75. FEBS Lett, 2002. 516(1-3): p. 53-7.
36. Peters, M.J., et al., Genome-wide association study meta-analysis of chronic widespread pain: evidence for involvement of the 5p15.2 region. Ann Rheum Dis, 2013. 72(3): p. 427-36.
37. Broer, L., et al., Meta-analysis of telomere length in 19 713 subjects reveals high heritability, stronger maternal inheritance and a paternal age effect. Eur J Hum Genet, 2013.
38. Codd, V., et al., Identification of seven loci affecting mean telomere length and their association with disease. Nat Genet, 2013. 45(4): p. 422-7, 427e1-2.
39. Diaz, V.A., et al., Telomere length and adiposity in a racially diverse sample. Int J Obes (Lond), 2010. 34(2): p. 261-5.
40. Fitzpatrick, A.L., et al., Leukocyte telomere length and cardiovascular disease in the cardiovascular health study. Am J Epidemiol, 2007. 165(1): p. 14-21.
41. Lee, M., et al., Inverse association between adiposity and telomere length: The Fels Longitudinal Study. Am J Hum Biol, 2011. 23(1): p. 100-6.
42. Njajou, O.T., et al., Shorter telomeres are associated with obesity and weight gain in the elderly. Int J Obes (Lond), 2012. 36(9): p. 1176-9.
43. Nordfjall, K., et al., Telomere length is associated with obesity parameters but with a gender difference. Obesity (Silver Spring), 2008. 16(12): p. 2682-9.
44. Valdes, A.M., et al., Obesity, cigarette smoking, and telomere length in women. Lancet, 2005. 366(9486): p. 662-4.
45. Zhu, H., et al., Leukocyte telomere length in healthy Caucasian and African-American adolescents: relationships with race, sex, adiposity, adipokines, and physical activity. J Pediatr, 2011. 158(2): p. 215-20.
46. Horn, T., B.C. Robertson, and N.J. Gemmell, The use of telomere length in ecology and evolutionary biology. Heredity (Edinb), 2010. 105(6): p. 497-506.
47. Achi, M.V., N. Ravindranath, and M. Dym, Telomere length in male germ cells is inversely correlated with telomerase activity. Biol Reprod, 2000. 63(2): p. 591-8.
48. Kong, A., et al., Rate of de novo mutations and the importance of father's age to disease risk. Nature, 2012. 488(7412): p. 471-5.

49. Halaschek-Wiener, J., et al., Reduced telomere length variation in healthy oldest old. Mech Ageing Dev, 2008. 129(11): p. 638-41.
50. Bakaysa, S.L., et al., Telomere length predicts survival independent of genetic influences. Aging Cell, 2007. 6(6): p. 769-74.
51. Bischoff, C., et al., No association between telomere length and survival among the elderly and oldest old. Epidemiology, 2006. 17(2): p. 190-4.
52. Cawthon, R.M., et al., Association between telomere length in blood and mortality in people aged 60 years or older. Lancet, 2003. 361(9355): p. 393-5.
53. Ehrlenbach, S., et al., Influences on the reduction of relative telomere length over 10 years in the population-based Bruneck Study: introduction of a well-controlled high-throughput assay. Int J Epidemiol, 2009. 38(6): p. 1725-34.
54. Epel, E.S., et al., The rate of leukocyte telomere shortening predicts mortality from cardiovascular disease in elderly men. Aging (Albany NY), 2009. 1(1): p. 81-8.
55. Harris, S.E., et al., The association between telomere length, physical health, cognitive ageing, and mortality in non-demented older people. Neurosci Lett, 2006. 406(3): p. 260-4.
56. Kimura, M., et al., Telomere length and mortality: a study of leukocytes in elderly Danish twins. Am J Epidemiol, 2008. 167(7): p. 799-806.
57. Martin-Ruiz, C.M., et al., Telomere length in white blood cells is not associated with morbidity or mortality in the oldest old: a population-based study. Aging Cell, 2005. 4(6): p. 287-90.
58. Njajou, O.T., et al., Association between telomere length, specific causes of death, and years of healthy life in health, aging, and body composition, a population-based cohort study. J Gerontol A Biol Sci Med Sci, 2009. 64(8): p. 860-4.
59. Zekry, D., et al., Telomere length, comorbidity, functional, nutritional and cognitive status as predictors of 5 years post hospital discharge survival in the oldest old. J Nutr Health Aging, 2012. 16(3): p. 225-30.
60. Al-Attas, O.S., et al., Adiposity and insulin resistance correlate with telomere length in middle-aged Arabs: the influence of circulating adiponectin. Eur J Endocrinol, 2010. 163(4): p. 601-7.
61. Aviv, A., et al., Menopause modifies the association of leukocyte telomere length with insulin resistance and inflammation. J Clin Endocrinol Metab, 2006. 91(2): p. 635-40.
62. Njajou, O.T., et al., Shorter telomeres are associated with obesity and weight gain in the elderly. Int J Obes (Lond), 2011.
63. Menni, C., et al., Metabolomic markers reveal novel pathways of ageing and early development in human populations. Int J Epidemiol, 2013.
64. Pritchard, J.K. and N.J. Cox, The allelic architecture of human disease genes: common disease-common variant...or not? Hum Mol Genet, 2002. 11(20): p. 2417-23.
65. Maher, B., Personal genomes: The case of the missing heritability. Nature, 2008. 456(7218): p. 18-21.
66. Al-Attas, O.S., et al., Telomere length in relation to insulin resistance, inflammation and obesity among Arab youth. Acta Paediatr, 2010. 99(6): p. 896-9.
67. Decary, S., et al., Shorter telomeres in dystrophic muscle consistent with extensive regeneration in young children. Neuromuscul Disord, 2000. 10(2): p. 113-20.
68. Blasco, M.A., et al., Telomere shortening and tumor formation by mouse cells lacking telomerase RNA. Cell, 1997. 91(1): p. 25-34.
69. Hoeijmakers, J.H., Genome maintenance mechanisms for preventing cancer. Nature, 2001. 411(6835): p. 366-74.
70. Gieger, C., et al., Genetics meets metabolomics: a genome-wide association study of metabolite profiles in human serum. PLoS Genet, 2008. 4(11): p. e1000282.
71. Li, J. and L. Ji, Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. Heredity (Edinb), 2005. 95(3): p. 221-7.
72. Deelen, J., et al., Gene set analysis of GWAS data for human longevity highlights the relevance of the insulin/IGF-1 signaling and telomere maintenance pathways. Age (Dordr), 2013. 35(1): p. 235-49.

Chapter 6

Summary & Samenvatting

SUMMARY

Longevity is usually defined as age at death or survival to an exceptional age, such as 90 years or older, or even 100 years or older in humans. The genetic contribution to longevity has been estimated to be up to 40%, yet no consistent genetic associations, except for candidate genes *APOE* and *FOXO3*, have been identified. There are three well-known biological pathways involved in longevity in animal models, namely Insulin-like growth factor signaling, stress resistance and senescence as mediated by telomere length.

The primary focus of this thesis was to find genetic markers associated with longevity and a longevity-associated trait telomere length using both genome-wide association studies (GWAS) and candidate gene association studies. Using a candidate gene approach we also made an effort to identify genetic variants associated with longevity-related diseases including Alzheimer's Disease (AD) and Parkinson's Disease (PD).

In **Chapter 2.1** we investigated the genetic architecture of longevity, defined as individuals reaching at least 90 years of age and controls defined as individuals that died between 55 and 80 years of age, in the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium (cases = 6,036, controls = 3,757). No new genome-wide significant associations were discovered. Moreover none of the previously suggested longevity loci from literature, except for *APOE* and *FOXO3*, were replicated illustrating the complexity of the longevity phenotype. This large scale GWAS was focused only on detecting variants that were common in populations and exerted small effects on longevity. A potential reason for not finding significant associations with longevity could be the involvement of rare genetic variants with relatively large effects on longevity, that were not captured by these GWAS. The only ways to identify these were to sequence or use imputed genetic variants from the 1000 Genomes Project for analysis. We studied the rare variants using both sequencing and 1000 Genomes imputed dataset in the Rotterdam Study to discover large effect rare variants that might influence longevity (**Chapter 2.2**). However, we could not find significant evidence for the involvement of rare variants with large effects on longevity. On the contrary, what we discovered was that the unidentified common variants present on genome-wide genotyping arrays explained over 80% of the total genetic variance of longevity. This implies that the effect of each individual common variation is very small and that these can be discovered by increasing the sample size in the GWAS.

Chapter 3 focusses on candidate genes involved in the stress response, namely Heat Shock Proteins (HSPs). We investigated the association of these genes with longevity and age-related diseases like AD and PD, which are often identified as protein folding disorders. The HSP families investigated included HSP10, HSP40, HSP60, HSP70, HSP90, BAG, prefoldin, small HSP (sHSP) and Heat Shock Factors (HSFs). We investigated the association of HSPs with AD, including 462 patients and 5,238 unaffected participants derived from the Rotterdam Study, and sought replication in the Framingham Heart Study and Cardiovascular Health Study (**Chapter 3.1**). Adjusted for multiple testing we found a small but consistent effect

of rs12118313 located 32kb from *PFDN2* (OR = 1.19; p-value = 0.003). In a formal pathway analysis we found nominally significant evidence for the association of three classes of co-chaperones (BAG, DNAJA and prefoldin) with AD. No effect of the HSP70 family on AD was identified. Three genes showed evidence of association with PD, namely Chaperonin (*HSP60* + *HSP10*), *HSPA13* and *HSF5* (**Chapter 3.2**). None of these genes were replicated in an online-available GWAS on PD with p-values lower than 10^{-4}. While three SNPs (rs12574703, rs7797781 and rs1416733) in genes *HSPA8*, *HSPB1* and *HSF2* showed evidence for association (p-values = 0.001, 0.002 and 0.003) with all-cause mortality replication in CHARGE was successful for *HSF2* (rs1416733, p-value = 0.003) only (**Chapter 3.3**). The identified SNP is a known cis-eQTL for *HSF2*, which is one of the main transcription factors influencing HSP transcription.

Next we focused on telomere length, which is considered as a biomarker of aging. As telomere length is a relatively new phenotype **Chapter 4** focuses on various exploratory analyses of telomere length including heritability analysis and factors that might influence telomere length. In **Chapter 4.1** we investigated the heritability, mode of inheritance and the influence of parental age on telomere length in six independent family-based cohorts including 19,713 participants. The heritability of telomere length was estimated to be 70% (p-value = $2.31*10^{-111}$). We additionally observed a stronger mother-offspring (r = 0.42; p-value = $3.60*10^{-61}$) than father-offspring correlation (r = 0.33; p-value = $7.01*10^{-5}$). This did not differ between male and female offspring, suggesting a maternal, non-X-chromosomal, inheritance component. We found a positive association between paternal age and telomere length (beta = 0.005; p-value = $7.01*10^{-5}$).

The high heritability of telomere length and the family-based structure of the ERF population enabled us to extrapolate the telomere length of the 90+ to younger ages and also study the so called survivor effect, i.e., the reduced variation in telomere length in the elderly (**Chapter 4.2**). Survivor effect has been cited as the reason for the inconsistent association between telomere length and mortality. We observed that the mean telomere length of the grandchildren of the oldest-old was significantly longer (p-value = $2.60*10^{-4}$) compared to others in the same age-range. This suggests that the oldest-old started life with long telomeres. Moreover, we observed a decreasing variability of telomere length from young adulthood to old age (p-value = $4.23*10^{-6}$), pointing to the presence of a survivor effect.

Average telomere length is significantly higher (p-value = $2.54*10^{-7}$) and telomere attrition rate lower (69.9%) in women compared to men. Differences in mean telomere length have been attributed to differences in height between genders as larger bodies require more cell replications which results in the reduction of telomeres. Building on this hypothesis, in **Chapter 4.3**, we investigated the association of height with telomere length in humans. We observed a significant correlation between telomere length and height in men (r = -0.08; p-value = $5.8*10^{-4}$) but not in women (r = -0.02; p-value = 0.30). This was in line with the

earlier findings of the studies performed in animals. Mendelian Randomization analysis suggested that height causally influences telomere length.

Other factors influencing telomere length are oxidative stress and inflammation, which also have an influence on the most prominent adipocyte-derived protein hormones, namely leptin and adiponectin. In **Chapter 4.4** we investigated the association of telomere length with these hormones. In seven independent cohorts from the ENGAGE consortium including 11,460 participants the partial correlation between telomere length and adiponectin/leptin was investigated adjusting for age and sex. In extended models we additionally adjusted for body mass index (BMI) and C-reactive protein (CRP). No association between adiponectin and telomere length was identified. The association between telomere length and leptin was highly significant ($r = -0.05$; p-value = $1.81*10^{-7}$). Telomere length showed no association with BMI or CRP. Adjusting for BMI or CRP did not change the results. Sex-stratified analysis revealed no differences between men and women.

Finally, in an effort towards a better understanding of the biology behind the association of telomere length with age-related diseases and mortality, we investigated the association between telomere length and serum metabolites in 8,192 individuals from the ENGAGE consortium (**Chapter 4.5**). Partial correlation analysis identified six metabolites to be significantly associated with telomere length consistently across populations. These metabolites included three phosphatidylcholines, two amino acids (methionine and tyrosine) and one acyl-carnitine. Together these metabolites implicate the involvement of lipid, homocysteine and thyroid metabolism in telomere length attrition.

This thesis shows that the longevity phenotype is highly complex and resistant to finding new genetic associations. For the biomarker telomere length we have only scratched the surface of understanding its association with longevity. In **Chapter 5** (general discussion) the difficulty in distinguishing true from false positive findings in genomics and other -omics studies is discussed, as well as further steps to investigate longevity and telomere length.

SAMENVATTING

Langlevendheid wordt meestal gedefinieerd als leeftijd van overlijden of overleving tot extreme leeftijd, bijvoorbeeld 90 jaar en ouder of zelfs 100 jaar en ouder in mensen. De genetische bijdrage aan langlevendheid wordt geschat op 40%, maar er zijn nog geen consistente genetische associaties ontdekt, behalve de kandidaatgenen *APOE* en *FOXO3*. Er zijn drie bekende biologische processen betrokken bij langlevendheid in dierstudies, namelijk insuline-achtige groeifactor signalering, stress resistentie en het beëindigen van de celdeling wat gereguleerd wordt door telomeerlengte.

De primaire focus van dit proefschrift was om genetische markers te vinden die geassocieerd zijn met langlevendheid en telomeerlengte, gebruik makend van zowel genoom-wijde associatie studies (GWAS) als kandidaatgen associatie studies. Gebruikmakend van de kandidaatgen methode heb ik ook een poging gedaan om genetische varianten geassocieerd met langlevendheid gerelateerde eigenschappen, Alzheimer(AD) en de ziekte van Parkinson (PD) te ontdekken.

In **Hoofdstuk 2.1** hebben we de genetische architectuur van langlevendheid, gedefinieerd als mensen die minstens 90 jaar oud zijn geworden en controles gedefinieerd als individuen die tussen de 55 en 80 jaar oud zijn overleden, in het CHARGE consortium onderzocht (langlevende deelnemers = 6.036, controles = 3.757). Geen nieuwe genoom-wijd significante associaties werden ontdekt. Bovendien werden voorheen in de literatuur genoemde associaties niet gerepliceerd, behalve *APOE* en *FOXO3*. Dit benadrukt de complexiteit van langlevendheid als fenotype. Deze grootschalige GWAS was gericht op het vinden van varianten die veelvoorkomend zijn in de populatie en een klein effect hebben op langlevendheid. Een mogelijke reden voor het niet vinden van significante associaties met langlevendheid kan zijn dat zeldzame genetische varianten met grote effecten betrokken zijn bij langlevendheid, welke niet door GWAS bestudeerd kunnen worden. De enige manier om deze varianten te identificeren is door sequentie data te gebruiken of door de varianten te imputeren met behulp van het 1000 Genomen Project. We bestudeerden de zeldzame varianten gebruikmakend van zowel sequentie data als met behulp van 1000 Genomen geïmputeerde data in de Rotterdam Study om zeldzame varianten met grote effecten te ontdekken, die invloed op langlevendheid hebben (**Hoofdstuk 2.2**). We hebben echter geen significant bewijs gevonden voor de betrokkenheid van zeldzame varianten met grote effecten bij langlevendheid. Daarentegen, hebben we ontdekt dat de ongeïdentificeerde veelvoorkomende varianten aanwezig op genoom-wijde genotypering arrays meer dan 80% van de totale genetische variatie van langlevendheid bepalen. Dit suggereert dat het effect op langlevendheid van iedere individuele veelvoorkomende variant erg klein is en dat deze ontdekt kunnen worden door het aantal individuen in GWAS te vergroten.

Hoofdstuk 3 richt zich op kandidaatgenen die betrokken zijn bij de stress respons, namelijk Heat Shock Proteins (HSPs). We onderzochten de associatie van deze genen met langlevendheid en leeftijd-gerelateerde ziektes zoals AD en PD welke beide als

eiwitvouwingsziektes bekent staan. De HSP families die onderzocht zijn bevatten HSP10, HSP40, HSP60, HSP70, HSP90, BAG, prefoldin, small HSP (sHSP) en Heat Shock Factors (HSFs). We onderzochten de associatie van HSPs met AD. We includeerden 462 patiënten en 5.238 controles van de Rotterdam Study en zochten replicatie in de Framingham Heart Study en Cardiovascular Health Study (**Hoofdstuk 3.1**). Gecorrigeerd voor meervoudig testen vonden we een klein maar consistent effect van rs12118313, wat zich op 32kb van *PDFN2* bevind (OR = 1,19; p-waarde = 0,003). In een formele netwerk analyse vonden we nominaal significant bewijs voor de associatie van drie groepen van co-chaperones (BAG, DNAJA en prefoldin) met AD. Geen effect van de HSP70 familie werd geïdentificeerd. Drie genen vertooneden bewijs van associatie met PD, namelijk Chaperonin (*HSP60* en *HSP10*), *HSPA13* en *HSF5* (**Hoofdstuk 3.2**). Geen van deze genen werden gerepliceerd met p-waarden lager dan 10^{-4} in een online-beschikbare GWAS naar PD. Terwijl drie SNPs (rs12574703, rs7797781 en rs1416733) in de genen *HSPA8*, *HSPB1* en *HSF2* bewijs voor associatie met sterfte (p-waardes = 0,001, 0,002 en 0,003) vertoonden was replicatie in CHARGE alleen succesvol voor *HSF2* (rs1416733, p-waarde = 0.003, **Hoofdstuk 3.3**). De geïdentificeerde variant is een bekende cis-eQTL van *HSF2*, wat een van de belangrijkste transcriptiefactoren is die transcriptie van HSPs reguleert.

Hierna concentreerden we ons op telomeerlengte, wat gezien wordt als een biomarker voor veroudering. Aangezien telomeerlengte een relatief nieuw fenotype is richt **Hoofdstuk 4** zich op verscheidende analyses naar telomeerlengte, onder meer erfelijkheidsanalyse en factoren die mogelijk invloed hebben op telomeerlengte. In **Hoofdstuk 4.1** onderzochten we de erfelijkheid, het overervingspatroon en de invloed van de leeftijd van de ouders op telomeerlengte in zes onafhankelijke cohorten met in totaal 19.713 deelnemers. De erfelijkheid van telomeerlengte werd geschat op 70% (p-waarde = $2.31*10^{-111}$). We vonden ook een sterkere moeder-kind (r = 0,42; p-waarde = $3,60*10^{-61}$) dan vader-kind (r = 0,33; p-waarde = $7,01*10^{-5}$) correlatie. Dit verschilde niet tussen mannelijke en vrouwelijke kinderen hetgeen suggereert dat er een maternale, niet X-chromosomale, overervingscomponent is. We vonden een positieve associatie met paternale leeftijd en telomeerlengte (beta = 0,005; p-waarde = $7,01*10^{-5}$).

De hoge erfelijkheid van telomeerlengte en de familiestructuur van de ERF populatie stelden ons instaat om de telomeerlengte van de 90+ers te extrapoleren naar jongere leeftijden en om het zogenoemde overlevingseffect, de gereduceerde variabiliteit in telomeerlengte in ouderen, te bestuderen (**Hoofdstuk 4.2**). Overlevingseffect word genoemd as de reden voor de inconsistente associaties van telomeerlengte met sterfte. We observeerden dat de gemiddelde telomeerlengte van de kleinkinderen van de oudste-ouden significant langer waren (p-waarde = $2.60*10^{-4}$) dan de telomeerlengte van andere mensen van hun leeftijd. Dit duidt er op dat de oudste-ouden hun leven begonnen met lange telomeren. Bovendien observeerden we een afnemende variabiliteit van telomeerlengte van jongvolwassen

leeftijd tot oudere leeftijd (p-waarde = $4.23*10^{-6}$), wat duidt op de aanwezigheid van een overlevingseffect.

Gemiddelde telomeerlengte was significant hoger (p-waarde = $2.54*10^{-7}$) en de afnamesnelheid van telomeerlengte was lager (69.9%) in vrouwen vergelijkend met mannen. Verschillen in gemiddeld telomeerlengte worden toegeschreven aan verschillen in lichaamslengte tussen de geslachten aangezien grotere lichamen meer cel replicaties nodig hebben wat resulteert in kortere telomeren. Bouwend op deze hypothese, in **Hoofdstuk 4.3**, onderzochten we de associatie van lichaamslengte met telomeerlengte in mensen. We observeerden een significante associatie tussen telomeerlengte en lichaamslengte in mannen (r = -0.08; p-waarde = $5.8*10^{-4}$), maar niet in vrouwen (r = -0.02; p-waarde = 0.30). Dit komt overeen met eerdere bevindingen in dierstudies. Mendeliaanse Randomisatie suggereerde dat lichaamslengte telomeerlengte causaal beïnvloed.

Andere factoren die invloed hebben op telomeerlengte zijn oxidatieve stress en inflammatie, wat ook invloed heeft op de meest prominente vetcel-gemaakte eiwit hormonen, namelijk leptine en adiponectine. In **Hoofdstuk 4.4** onderzochten we de associatie van telomeerlengte met deze hormonen. In zeven onafhankelijke cohorten van het ENGAGE consortium met 11.460 deelnemers onderzochten we de partiële correlatie tussen telomeerlengte en adiponectine/leptine, gecontroleerd voor leeftijd en geslacht. In additionele modellen controleerden we ook voor BMI en CRP. Er werd geen associatie tussen adiponectine en telomeerlengte geobserveerd. De associatie tussen telomeerlengte en leptine was zeer significant (r = -0,05; p-value = $1,81*10^{-7}$). Telomeerlengte vertoonde geen associatie met BMI of CRP. Corrigeren voor BMI en CRP veranderde de resultaten niet. Geslacht-gestratificeerde analyse liet geen verschil tussen mannen en vrouwen zien.

In een poging om de biologie achter de associatie tussen telomeerlengte en leeftijdsgerelateerde ziektes en sterfte beter te begrijpen, onderzochten we de associatie tussen telomeerlengte en metabolieten gemeten in het bloed in 8.192 individuen van het ENGAGE consortium (**Hoofdstuk 4.5**). Partiële correlatie analyse identificeerde zes metabolieten die significant geassocieerd zijn met telomeerlengte. Deze metabolieten bestonden uit drie phosphatidylcholines, twee aminozuren (metionine en tyrosine) en een acyl-carnitine. Samen impliceren deze metabolieten de betrokkenheid van lipide, homocysteine en thyroïd metabolisme in telomeerlengte bepaling.

Dit proefschrift laat zien dat langlevendheid een erg complex fenotype is en dat er geen nieuwe consistente genetische associaties gevonden zijn. Voor de biomarker telomeerlengte hebben we pas het topje van ijsberg ontdekt en ons begrip van de associatie met langlevendheid is nog verre van volledig. In **Hoofdstuk 5** (algemene discussie) hebben we de problematiek rond het onderscheiden van ware bevindingen en fout-positieve bevindingen in genomics en andere '-omics' studies beschreven, alsook verdere stappen in de studie naar langlevendheid en telomeerlengte beschreven.

C h a p t e r 7

Appendices (supplementary material)

Appendix of published chapters (3.1, 3.2, 3.3, 4.1 and 5.1) can be found in published articles online.

APPENDIX CHAPTER 2.1

Methods

Study population details

From 1991 to 1995 all inhabitants of Ommoord, a district of Rotterdam, The Netherlands, who were 55 years or older, were invited to participate in the **Rotterdam study (RS)** [1, 2]. Genotyping information was available for 5,974 participants. All of the participants were followed for incident diseases through linkage to the general practitioner data base and record review by trained medical investigators. General practitioners', hospital records as well as death certificates were used for identification of deaths and health events through 01.01.2009.

The **Study of Osteoporotic Fractures (SOF)** is a prospective multicenter study of risk factors for vertebral and non-vertebral fractures [3]. The cohort at the baseline visit is comprised of 9704 community dwelling women 65 years old or older recruited from population-based listings in four U.S. areas: Baltimore, Maryland; Minneapolis, Minnesota; Portland, Oregon; and the Monongahela Valley, Pennsylvania. Women enrolled in the study were 99% Caucasian with African American women initially excluded from the study due to their low incidence of hip fractures. The SOF inclusion criteria were: 1) 65 years or older, (2) ability to walk without the assistance of another, (3) absence of bilateral hip replacements, (4) ability to provide self-reported data, (5) residence near a clinical site for the duration of the study, (6) absence of a medical condition that (in the judgment of the investigator) would result in imminent death, and (7) ability to understand and sign an informed consent. The SOF participants were followed up every four months by postcard or telephone to ascertain the occurrence of falls, fractures and changes in address. To date, follow-up rates have exceeded 95% for vital status and fractures. DNA was extracted from buffy coat or whole blood samples collected at visit 2 (1989-1990) or visit 6 (1997-1998) and whole-genome genotyping was performed at the Broad Institute using DNA from 3924 participants. After QC, whole genome genotype data was available for 3625 SOF participants.

The **Cardiovascular Health Study (CHS)** is a prospective population-based cohort study of CVD, mortality, and other outcomes in 65+ year old Medicare-eligible adults living in four US communities [4]. Recruitment of the initial predominantly white cohort was completed in 1990 and 3,267 participants fulfilled the inclusion criteria of this study and had genotyping information available. Only European or European Americans were included in the analysis, so the CHS African Americans were excluded. Major incident health events and deaths were identified through several methods, including: questionnaires completed by participants at each semi-annual contact during follow-up; reports by family members;

and periodic searches of the Medicare Utilization database, the National Death Index, and local newspaper obituaries through 30.06.2005. All cardiovascular events were reviewed and adjudicated by an events committee.

The **Osteoporotic Fractures in Men Study (MrOS)** is a multi-center prospective, longitudinal, observational study of risk factors for vertebral and all non-vertebral fractures in older men, and of the sequelae of fractures in men [5, 6]. The MrOS study population at baseline consists of 5,994 community dwelling, ambulatory men aged 65 years or older from six communities in the United States (Birmingham, AL; Minneapolis, MN; Palo Alto, CA; Monongahela Valley near Pittsburgh, PA; Portland, OR; and San Diego, CA). Inclusion criteria were designed to provide a study cohort that is representative of the broad population of older men. The inclusion criteria were: (1) ability to walk without the assistance of another, (2) absence of bilateral hip replacements, (3) ability to provide self-reported data, (4) residence near a clinical site for the duration of the study, (5) absence of a medical condition that (in the judgment of the investigator) would result in imminent death, and (6) ability to understand and sign an informed consent. The baseline examination was completed over a 25-month period from March 2000 to April 2002. During the study period, participants complete a tri-annual questionnaire every four months that obtains information concerning the occurrence of incident falls and fractures. Follow-up is over 99% complete. All deaths are centrally adjudicated using death certificates and recent hospitalization records. DNA was extracted from whole blood samples collected at baseline and consent for use of DNA for genetic analysis was obtained through written consent. Whole-genome genotyping was performed at the Broad Institute using DNA from 5506 participants. After QC, whole genome genotype data was available for 5151 MrOS participants.

Framingham Heart Study (FHS) was initiated to study determinants of CVD and other major illnesses. The original cohort was recruited in 1948 and the offspring of the Original cohort participants and offspring spouses were enrolled in 1971 [7-10]. DNA was obtained for genetic studies in the 1990s from surviving Original cohort and Offspring participants. The exam at which DNA was obtained is considered the baseline exam for these analyses. All participants remain under continuous surveillance and deaths that occurred through 01.01.2009 were included.

The **Health and Retirement Study (HRS)** is a longitudinal survey of a representative sample of Americans over the age of 50. The current sample includes over 26,000 persons in 17,000 households. Respondents are interviewed every two years about income and wealth, health and use of health services, work and retirement, and family connections. DNA was extracted from saliva collected during a face-to-face interview in the respondents' homes. These data represent respondents who provided DNA samples and signed consent forms in 2006 and 2008. Deaths were validated using the Social Security Administration's Death Master File that included reports through the end of 2010. Cause of death was determined by linking to the National Death Index (NDI) Cause of Death file.

The **Age, Gene/Environment Susceptibility -Reykjavik Study (AGES)** was initiated to examine potential genetic susceptibility and gene/environment interaction. Between 2002 and 2006, baseline exams were conducted in survivors from the Reykjavik Study. 3,219 AGES-Reykjavik participants were eligible for participation and with available genotype information. Follow-up information was complete till 27.04.2009 via linkage to electronic medical records and vital status registry.

The Religious Orders Study (ROS), started in 1994, is a longitudinal, clinical-pathologic cohort study of common chronic conditions of aging. The study enrolls Catholic priests, nuns, and brothers from about 40 groups in 12 states of the United States. Since January 1994, over 1,100 participants completed their baseline evaluation. The study was approved by the institutional review board of Rush University Medical Center. The follow-up rate of survivors exceeds 90%. Participants were free of known dementia at enrollment, agreed to annual clinical evaluations, and signed both an informed consent and an Anatomic Gift Act form donating their brains at time of death. DNA was extracted from whole blood, lymphocytes, or frozen postmortem brain tissue. Genotyping was performed at the Broad Institute's Center for Genotyping and the Translational Genomics Research Institute.

Rush Memory and Aging Project (MAP), started in 1997, is a longitudinal, clinical-pathologic cohort study of common chronic conditions of aging. The study enrolled older men and women from assisted living facilities in the Chicago area with no evidence on dementia at baseline. Since October 1997, over 1,500 participants completed their baseline evaluation. The study was approved by the institutional review board of Rush University Medical Center. The follow-up rate of survivors exceeds 90%. Similar to the ROS, participants agreed to annual clinical evaluations and signed both an informed consent and an Anatomic Gift Act form donating their brains, spinal cords, and selected nerves and muscles to Rush investigators at the time of death. DNA was extracted from whole blood, lymphocytes, or frozen postmortem brain tissue. Genotyping was performed at the Broad Institute's Center for Genotyping and the Translational Genomics Research Institute.

The **Invecchiare nel Chianti (InCHIANTI)** study is a population-based epidemiological study aimed at evaluating the factors that influence mobility in the older population living in the Chianti region in Tuscany, Italy that was initiated in 1998. Presence of chronic diseases was ascertained by a combination of assessments by trained geriatrician, medication use, blood tests, and self-reported physician diagnosis. Death was determined using a death registry. 686 subjects with genotype and mortality and disease data assessed through to 06.03.2006 were used for this analysis.

The **Baltimore Longitudinal Study of Ageing (BLSA)** is a population-based study aimed to evaluate contributors of healthy aging in the older population residing predominantly in the Baltimore-Washington DC area. The study began in 1958, and follow up data through to 20.12.2004 for 599 subjects was used for this analysis. Participants returned for follow-up visits every 1-2 years where presence of chronic disease was assessed by a physician during

the physical exam, self-report or medication use. Death was ascertained through death certificates or report by family members.

The **NECS case-control** set of 801 centenarians and 914 genetically matched controls is described in [11]. The study included centenarians from the NECS [12] (median survival 104 years, age range 94-119) and a combination of controls from the NECS (median age 73 years, age range 50-93) and from the Illumina control database (iControlDB). Ages were unknown for the majority of Illumina controls. The algorithm for genetic matching is described in [13]. The data were genotyped using Illumina arrays 370, 610 and 1M, and the final list of 281 SNPs passed stringent quality control measures, including a sample call rate > 96% and a SNP call rate > 98% in all array types. The NECS protocol is approved by the Boston University Medical Campus Institutional Review Board.

The Long Life Family Study (**LLFS**) enrolled families enriched for longevity via 4 field centers (Boston, New York, and Pittsburgh in the USA, and Denmark) between 2006 and 2009. The recruitment protocol, described in [14], used the Family Longevity Selection Score (FLoSS) [15] to identify family enriched of exceptional longevity, and enrolled 583 families with a FLoSS $\geq$ 7 consisting of 1493 probands, their siblings and 192 spouses in the older generation, and 2437 offspring and 809 of their spouses. Genotype data for 4,567 subjects were generated from the Illumina Omni chip 2.5 (2.5M SNPs). The genotyping protocol, quality control and quality assessment included use of the GRR program to validate familial relations [16], the program Loki to validate Mendelian consistency [17], SNP call rate > 98% and sample call rate > 97.5%. These 4,567 subjects included 1562 from the parental generation (probands, their siblings and alive spouses, median age 93 years, range 52-110) and 3095 from the offspring generation (offspring and their spouses, median age 63, range 25-90). Participants underwent informed consent and the study was approved by the Institutional Review Boards of Boston University Medical Campus, University of Pittsburgh, Columbia University and Washington University St Louis, and the Ethics Board of the University of Denmark.

The **EU longevity** consortium GWAS was performed in 5409 longevity cases (≥ 90 years of age) and 16121 controls below 65 years of age, from 13 cohorts. All individuals were of European descent.

Analysis details

NECS genotype data were analyzed using allelic association tests and p-values are from 1dof c2 tests. LLFS genotype data were analyzed using a mixed effect Cox proportional hazard regression model implemented in the coxme package in R with a random effect per subject to account for within family correlation [18]. The random effects were modeled with a normal distribution with 0 mean and variance covariance matrix proportional to the kinship matrix [19]. Kinship coefficients were estimated using the kinship2 package in R.

The EU longevity consortium performed a GWAS on longevity using the same longevity definition and a slightly different comparison group definition (< 65 years of age). A standard logistic regression adjusting for sex was performed.

Significant genes were followed up using an online tool for pathway analysis: gene-networks (www.genenetworks.nl). We used this tool to infer biological function of newly identified longevity genes.

Additionally we performed look-ups in two previous longevity GWAS studies, namely Newman *et al* and Sebastiani *et al* [11, 20]. Newman *et al* was the original CHARGE longevity GWAS publication, partly overlapping with the current study [20]. The same phenotype description was used. Sebastiani *et al* concerns a recent GWAS on centenarians showing a profile of genetic associations together determining longevity. In this study 4 different genetic models were used [11]. We focused on those associations reaching suggestive p-values for the additive model.

A recent genome-wide linkage analysis in longevity (90+) sibpairs identified 4 regions significantly linked to longevity [21]. We investigated the association of all SNPs located under the linkage peaks. To adjust for multiple testing we corrected for the number of independent tests (VeffLi) [22] using an online available tool (http://gump.qimr.edu.au/general/daleN/matSpDlite/).

Finally we performed a literature search for genetic associations of *FOXO3* with longevity. For the SNP most commonly reported we added the literature-based associations to the current study in a meta-analysis.

REFERENCES

1. Hofman, A., et al., Determinants of disease and disability in the elderly: the Rotterdam Elderly Study. Eur J Epidemiol, 1991. 7(4): p. 403-22.
2. Hofman, A., et al., The Rotterdam Study: 2012 objectives and design update. Eur J Epidemiol, 2011. 26(8): p. 657-86.
3. Cummings, S.R., et al., Risk factors for hip fracture in white women. Study of Osteoporotic Fractures Research Group. N Engl J Med, 1995. 332(12): p. 767-73.
4. Fried, L.P., et al., The Cardiovascular Health Study: design and rationale. Ann Epidemiol, 1991. 1(3): p. 263-76.
5. Blank, J.B., et al., Overview of recruitment for the osteoporotic fractures in men study (MrOS). Contemp Clin Trials, 2005. 26(5): p. 557-68.
6. Orwoll, E., et al., Design and baseline characteristics of the osteoporotic fractures in men (MrOS) study--a large observational study of the determinants of fracture in older men. Contemp Clin Trials, 2005. 26(5): p. 569-85.
7. Dawber, T.R., W.B. Kannel, and L.P. Lyell, An approach to longitudinal studies in a community: the Framingham Study. Ann N Y Acad Sci, 1963. 107: p. 539-56.
8. Dawber, T.R., G.F. Meadors, and F.E. Moore, Jr., Epidemiological approaches to heart disease: the Framingham Study. Am J Public Health Nations Health, 1951. 41(3): p. 279-81.
9. Feinleib, M., et al., The Framingham Offspring Study. Design and preliminary data. Prev Med, 1975. 4(4): p. 518-25.
10. Kannel, W.B., et al., An investigation of coronary heart disease in families. The Framingham offspring study. Am J Epidemiol, 1979. 110(3): p. 281-90.
11. Sebastiani, P., et al., Genetic signatures of exceptional longevity in humans. PLoS ONE, 2012. 7(1): p. e29848.
12. Andersen, S.L., et al., Health span approximates life span among many supercentenarians: compression of morbidity at the approximate limit of life span. J Gerontol A Biol Sci Med Sci, 2012. 67(4): p. 395-405.
13. Solovieff, N., et al., Clustering by genetic ancestry using genome-wide SNP data. BMC Genet, 2010. 11: p. 108.
14. Newman, A.B., et al., Health and function of participants in the Long Life Family Study: A comparison with other cohorts. Aging (Albany NY), 2011. 3(1): p. 63-76.
15. Sebastiani, P., et al., A family longevity selection score: ranking sibships by their longevity, size, and availability for study. Am J Epidemiol, 2009. 170(12): p. 1555-62.
16. Abecasis, G.R., et al., GRR: graphical representation of relationship errors. Bioinformatics, 2001. 17(8): p. 742-3.
17. Heath, S.C., Markov chain Monte Carlo segregation and linkage analysis for oligogenic models. Am J Hum Genet, 1997. 61(3): p. 748-60.
18. Govindarajulu, U.S., et al., Frailty models: Applications to biomedical and genetic studies. Stat Med, 2011. 30(22): p. 2754-64.
19. Borecki, I.B. and M.A. Province, Genetic and genomic discovery using family studies. Circulation, 2008. 118(10): p. 1057-63.
20. Newman, A.B., et al., A meta-analysis of four genome-wide association studies of survival to age 90 years or older: the Cohorts for Heart and Aging Research in Genomic Epidemiology Consortium. J Gerontol A Biol Sci Med Sci, 2010. 65(5): p. 478-87.
21. Beekman, M., et al., Genome-wide linkage analysis for human longevity: Genetics of Healthy Aging Study. Aging Cell, 2013. 12(2): p. 184-93.
22. Li, J. and L. Ji, Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. Heredity (Edinb), 2005. 95(3): p. 221-7.

Supplement Table 1A | Genotyping/Imputation information per discovery cohort.

Study	RS1/RS2	SOF	CHS	MrOS	FHS	HRS	AGES	RADC	InCHIANTI	BLSA
Array type	Version 3 Illumina Infinium II HumanHap550 SNP chip array	Illumina Human Omni1-Quad	Illumina 370CNV	Illumina Human Omni1-Quad	Affymetrix 500K and MIPS 50K combined	Illumina Omni2.5	Illumina 370CNV	Affymetrix 6.0	Illumina 550K	Illumina 550K
Genotyping center	Erasmus Medical Center	Broad Institute	Cedars-Sinai Medical Center	Broad Institute	Affymetrix Core Laboratory	Center for Inherited Disease Research (CIDR)	NIA Laboratory of Neurogenetics	Broad Institute's Center for Genotyping; Translational Genomics Research Institute	Laboratory of Neurogenetics (NIA)	Laboratory of Neurogenetics (NIA)
Genotype calling	Illumina Beadstudio	Illumina Beadstudio	Illumina BeadStudio	Illumina Beadstudio	BRLMM	Illumina BeadStudio	Illumina BeadStudio	Birdsuite, Broad Institute	BeadStudio	BeadStudio
Exclusion on SNPs used for imputation	Call rate < 90%, no MAF/HWE filter	Call rate < 97%, HWE $p < 10^{-4}$, MAF < 0.01, > 1 discordant call among duplicates	Call rate ≤ 97%, HWE $p < 10^{-5}$, > 2 duplicate errors or Mendelian inconsistencies (for reference CEPH trios) heterozygote frequency = 0, not in HapMap	Call rate < 97%, HWE $p < 10^{-4}$, MAF < 0.01, > 1 discordant call among duplicates	Call rate < 97%, HWE $p < 1^{-6}$, Mishap $p < 1^{-9}$, Mendelian errors >100	Call rate < 98%, HWE $p<10^{-4}$, MAF<1%	Call rate < 97%, HWE $p < 1^{-6}$, MAF < 1%, Mishap $p < 1^{-9}$, A/T and G/C SNPs, Mismatches between Illumina, dbSNP and/or HapMap position	HWE $p < 1x10^{-6}$; MAF < 0.01, genotype call rate < 0.95; mishap test $< 1x10^{-9}$	call rate < 99%, HWE $p < 10^{-4}$, MAF < 1%,	call rate < 99%, HWE $p < 10^{-4}$, MAF < 1%,
Exclusion on a per sample basis	Call rate < 97.5%, sex mismatch, excess autosomal heterozygosity > 0.336, outliers identified by the IBS clustering analysis	Call rate < 97%, sex mismatch, IBD estimate of genetic relatedness > half-sibling, > 5 chromo-somal abnormalities, non-European descent	Call rate <95%, sex mismatch, sample failure	Call rate <97%, sex mismatch, IBD estimate of genetic relatedness > half-sibling, >5 chromo-somal abnormalities, non-European descent	Call rate < 97%, subject heterozygosity > 5 SD away from the mean, large Mendelian error rate	Call rate < 98%, sex mismatch, sample failure, duplicates and/or relatives with KD > 1/32	Sex mismatch, Sample failure, Genotype mismatch with reference panel	genotype success rate < 95%, genotype-derived gender discordant with reported gender, inbreeding coefficient F > 0.04	Call rate < 97%, heterozygosity < 0.3, sex mismatch	Non-European decent, call rate < 98.5%, sex mismatch, duplication

Study	RS1/RS2	SOF	CHS	MrOS	FHS	HRS	AGES	RADC	InCHIANTI	BLSA
Imputation	MACH 1.0	MACH (v 1.0.17) and minimac (v 2011-08-12 beta)	BIMBAM10 v0.99	MACH (v 1.0.17) and minimac (v 2011-08-12 beta)	MACH (version 1.00.15)	MaCH (v1.0.16)	MACH (version 1.00.16	MACH (version 1.0.16a)	MACH (version 1.0.16)	MACH (version 1.0.16)
Imputation backbone (NBCI build)	HapMap release 22 CEU (build 36)	HapMap release 22 CEU (build 36)	HapMap release 22 CEU (build 36)	HapMap release 22 CEU (build 36)	HapMap release 22 CEU (build 36)	HapMap release 22 CEU (build 36)	HapMap release 22 CEU (build 36)	HapMap release 22 CEU (build 36)	HapMap release 22 CEU (build 36)	HapMap release 22 CEU (build 36)
Data handling and statistical tests	PLINK, ProbABEL, Mach2QTL	R	R	R	R packages kinship, gee, coxph	PLINK and R	PLINK and R	ProbABEL, R	ProbABEL, R, Perl	ProbABEL, R, Perl

7

Supplement Table 1B | Genotyping information per replication cohort.

Study	EU_longevity	NECS	LLFS
Array type	Varies	Illumina 370;610;1M	Illumina 2.5
Genotyping center	Varies	BMC	CIDR
Genotype calling	Varies	Illumina beadchip	Illumina beadchip
Exclusion on SNPs	Varies	Call rate < 98%	Call rate < 98%
Exclusion on a per sample basis	Varies	Call rate < 98%	Call rate < 98%
Data handling and statistical tests	CC-assoc	R	R

Supplment Table 2 | Biological processes in which both GRIK2 and CADM2 are involved.

Biological process	GRIK2	CADM2
neuron cell-cell adhesion	$6.89*10^{-7}$	$1.48*10^{-13}$
regulation of synaptic transmission. glutamatergic	$2.25*10^{-6}$	$7.95*10^{-9}$
homophilic cell adhesion	$3.11*10^{-6}$	$3.18*10^{-4}$
glutamate signaling pathway	$9.26*10^{-6}$	$7.22*10^{-15}$
synapse assembly	$2.83*10^{-5}$	$9.60*10^{-3}$
synaptic transmission. glutamatergic	$1.01*10^{-4}$	$6.21*10^{-7}$
response to fungicide	$1.12*10^{-4}$	$1.61*10^{-4}$
prepulse inhibition	$2.41*10^{-4}$	$2.17*10^{-5}$
detection of mechanical stimulus	$3.36*10^{-4}$	$2.35*10^{-3}$
gamma-aminobutyric acid signaling pathway	$4.93*10^{-4}$	$1.07*10^{-9}$
detection of mechanical stimulus involved in sensory perception	$1.11*10^{-3}$	$3.88*10^{-4}$
retina morphogenesis in camera-type eye	$2.32*10^{-3}$	$2.84*10^{-4}$
learning or memory	$2.43*10^{-3}$	$1.39*10^{-6}$
startle response	$2.80*10^{-3}$	$1.36*10^{-11}$
neuron-neuron synaptic transmission	$2.95*10^{-3}$	$5.29*10^{-9}$
cognition	$2.99*10^{-3}$	$7.08*10^{-7}$
neural retina development	$3.61*10^{-3}$	$1.46*10^{-3}$
regulation of action potential	$3.74*10^{-3}$	$6.61*10^{-6}$
positive regulation of amine transport	$4.19*10^{-3}$	$9.81*10^{-3}$
memory	$4.35*10^{-3}$	$1.55*10^{-3}$
regulation of synaptic transmission	$7.66*10^{-3}$	$1.70*10^{-7}$
cell-cell adhesion	$7.73*10^{-3}$	$2.76*10^{-3}$
synapse organization	$8.07*10^{-3}$	$4.73*10^{-3}$
neurotransmitter uptake	$8.53*10^{-3}$	$1.56*10^{-5}$
long-term memory	$9.12*10^{-3}$	$1.62*10^{-4}$

P-values for involvement in different biological processes. Only those pathways shown with a p-value < 0.01 for both *CADM2* and *GRIK2*.

Supplement Table 3 | Replication of the findings of the original CHARGE longevity GWAS (Newman *et al*, 2010).

				original study								Current study meta-analysis				
				1st meta-analysis					2nd meta-analysis							
SNP	Chr	Gene	Distance(kb)	EA	EAF	OR	p-value	direction	direction	OR	p-value	EA	EAF	OR	p-value	direction
rs4443878	1	RGS7	34	T	0.04	0.41	$1.30*10^{-6}$	––––	++	0.83	0.068	T	0.03	0.56	$1.62*10^{-6}$	--+------+-
rs9825185	3	C3orf21	Intron	A	0.87	0.69	$2.50*10^{-6}$	––––	++	0.91	0.045	A	0.87	0.83	$2.90*10^{-4}$	--+-------+
rs954551	6	GRIK2	262	A	0.75	1.30	$5.30*10^{-6}$	++++	NA	NA	NA	A	0.74	1.23	$7.09*10^{-8}$	+-+++++++-+
rs7624691	3	IL20RB	133	T	0.57	1.25	$8.80*10^{-6}$	++++	––	1.05	0.092	T	0.56	1.16	$1.60*10^{-5}$	+-++++++-+-
rs10888267	1	OR2W3	Missense	T	0.55	0.80	$9.70*10^{-6}$	––––	NA	NA	NA	T	0.55	0.91	0.007	-++-------+
rs9972933	17	ACCN1	Intron	T	0.23	0.77	$1.10*10^{-5}$	––––	+–	0.89	0.003	T	0.22	0.94	0.138	-++-+-+-++-
rs2739532	4			C	0.27	1.48	$1.10*10^{-5}$	?+?+	NA	NA	NA	C	0.27	1.11	0.056	++-++--++++
rs8029244	15	LASS3	Intron	A	0.49	0.79	$1.20*10^{-5}$	––––	+–	0.90	0.002	A	0.49	0.96	0.205	-++-+-+--++
rs16850255	1	PAPPA2	6	T	0.79	1.33	$1.20*10^{-5}$	++++	––	1.09	0.041	T	0.79	1.15	0.002	++++++++---
rs1543505	14	REM2	5	A	0.72	0.79	$1.30*10^{-5}$	––––	+–	0.89	0.001	A	0.72	0.91	0.009	--+---+-++-
rs7321904	13	SPRY2	868	T	0.07	0.64	$1.30*10^{-5}$	––––	++	0.92	0.179	T	0.08	0.83	0.003	------+-++-
rs17401847	1	OTUD3	1	A	0.85	0.72	$1.40*10^{-5}$	––––	+–	0.89	0.015	A	0.86	0.84	0.001	-++-+-----+
rs3124736	10	CASP7	6	A	0.03	2.30	$1.40*10^{-5}$	+++?	NA	NA	NA	A	0.02	1.61	$1.79*10^{-4}$	+++?+++++++
rs690232	9	DIRAS2	Intron	A	0.30	1.27	$1.60*10^{-5}$	+++–	NA	NA	NA	A	0.30	1.17	$2.02*10^{-5}$	+--+++-++++
rs9664222	10	MINPP1	25	A	0.21	0.77	$1.60*10^{-5}$	––––	––	0.82	$6.8*10^{-7}$	A	0.22	0.85	$7.14*10^{-5}$	-+-----+---
rs11157721	14	LOC196913	33	T	0.39	0.79	$1.70*10^{-5}$	––––	+–	0.90	0.002	T	0.39	0.90	0.005	-++-----+++
rs4690810	4	SC4MOL	16	T	0.65	1.27	$1.90*10^{-5}$	++++	––	1.08	0.044	T	0.65	1.11	0.004	+-++++++++-
rs11605096	11	TMPRSS5	16	A	0.12	0.71	$1.90*10^{-5}$	––––	NA	NA	NA	A	0.12	0.82	$1.43*10^{-4}$	----+----+-
rs16972414	18	PIK3C3	2079	A	0.70	1.27	$2.00*10^{-5}$	++++	NA	NA	NA	A	0.69	1.05	0.220	+--+-+-++--
rs12935091	16	ZNF19	1	A	0.93	1.61	$2.00*10^{-5}$	++++	+–	1.25	0.002	A	0.93	1.21	0.020	+--?-+++---

7

				original study								Current study meta-analysis				
				1st meta-analysis					2nd meta-analysis							
SNP	Chr	Gene	Distance(kb)	EA	EAF	OR	p-value	direction	direction	OR	p-value	EA	EAF	OR	p-value	direction
rs210332	14	BMP4	223	T	0.81	0.75	$2.30*10^{-5}$	−−−−	NA	NA	NA	T	0.81	0.88	0.009	-+--+++---+
rs17369174	8	CRISPLD1	19	T	0.90	1.45	$2.30*10^{-5}$	++++	−+	1.16	0.014	T	0.91	1.19	0.004	++-+-+++-+-
rs6721003	2	SCN7A	38	A	0.45	1.23	$2.40*10^{-5}$	++++	−+	1.09	0.006	A	0.45	1.11	0.002	++-++++-++-+
rs4734457	8	ANKRD46	28	A	0.10	1.75	$2.50*10^{-5}$	++++	+−	1.10	0.098	A	0.11	1.21	0.015	+-+?-?++?-+

EA: effective allele; EAF: effective allele frequency; '+': effective allele overrepresented in 90+ cases; '-': effective allele underrepresented in 90+ cases; '?': not tested in this study.

Supplement Table 4 | Replication of the findings of Sebastiani *et al*, 2012.

				Sebastiani		Current Study			
SNP	**Chr**	**Gene**	**EA**	**OR**	**p-additive**	**EAF**	**OR**	**p-value**	**Direction**
rs2075650	19	TOMM40/APOE	A	2.07	$2.4*10^{-10}$	0.87	1.20	$4.8*10^{-4}$	+++++?+++++
rs12629971	3	EIF4E3	T	0.62	$1.9*10^{-6}$	0.14	0.93	0.118	-+-+-+--+--
rs4977756	9	NA	A	0.73	$8.0*10^{-6}$	0.60	0.93	0.031	-++--+---+-
rs6801173	3	EIF4E3	T	0.66	$8.2*10^{-6}$	0.17	0.95	0.249	-+-+-+--+--
rs1456669	3	NA	A	1.58	$8.6*10^{-6}$	0.13	0.96	0.353	-++--+-++++
rs4802234	19	CEACAM16	T	0.73	$9.2*10^{-6}$	0.44	0.94	0.095	----+------

EA: effective allele; EAF: effective allele frequency; '+': effective allele overrepresented in 90+ cases; '-': effective allele underrepresented in 90+ cases; '?': not tested in this study
Direction: RS1, RS2, SOF, CHS, MrOS, FHS, HRS, AGES, RADC, InCHIANTI, BLSA

Supplement Table 5 | Replication of linkage peaks (Beekman *et al*, 2013).

linkage regions			Current study								
Chr	**start pos**	**end pos**	**n_SNPs**	**VeffLi**	**Significance**	**SNPID**	**EA**	**EAF**	**OR**	**p-value**	**direction**
14	19494891	22794987	3918	796	$6.28*10^{-5}$	rs11622628	A	0.80	1.21	$1.23*10^{-4}$	+++++?++---
17	34001045	52871813	13428	2745	$1.82*10^{-5}$	rs2277624	T	0.23	0.87	$4.09*10^{-4}$	----------+
19	6644683	17313527	6435	1378	$3.08*10^{-5}$	rs10423231	A	0.56	0.86	$3.29*10^{-5}$	-----------
19	40003810	51262580	6467	1400	$3.57*10^{-5}$	rs4420638	A	0.83	1.31	$\mathbf{2.13*10^{-5}}$	+++++?+++++

Most significant SNP in the region presented in the table.
EA: effective allele; EAF: effective allele frequency; '+': effective allele overrepresented in 90+ cases;
'-': effective allele underrepresented in 90+ cases; '?': not tested in this study
VeffLi: Number of independent variables; Significance: p-value for significant finding (0.05/VeffLi)
Direction: RS1, RS2, SOF, CHS, MrOS, FHS, HRS, AGES, RADC, InCHIANTI, BLSA

Supplement Table 6 | meta-analysis of published and current results for FOXO3 (rs2802292; minor allele).

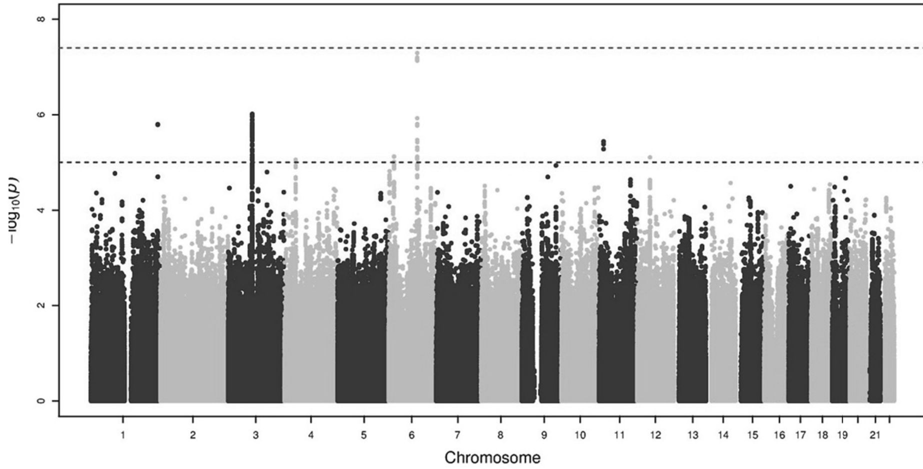
Study	Design	n_cases	n_controls	OR	95% CI	p-value
2008_Wilcox	95+	213	402	1.72	1.34-2.22	$2.44*10^{-5}$
2009_Anselmi	100+	281	195	1.57	1.17-2.10	0.002
2009_Li	100+	761	1056	1.36	1.16-1.60	$2.09*10^{-4}$
2010_Soerensen	95+	313	371	1.34	1.08-1.67	0.009
RS1	90+	899	1192	1.13	0.98-1.30	0.087
RS2	90+	69	161	1.39	0.86-2.24	0.181
SOF	90+	1720	124	0.95	0.72-1.25	0.710
CHS	90+	791	560	1.13	0.96-1.33	0.135
MrOS	90+	670	502	1.07	0.90-1.27	0.426
FHS	90+	320	484	0.99	0.80-1.24	0.961
HRS	90+	384	401	1.22	0.98-1.50	0.071
AGES	90+	541	145	1.02	0.80-1.30	0.882
RADC	90+	468	78	0.87	0.61-1.25	0.453
InCHIANTI	90+	101	75	1.21	0.80-1.81	0.366
BLSA	90+	73	75	1.09	0.64-1.85	0.754
meta	NA	7604	5821	1.18	1.11-1.25	**$1.44*10^{-8}$**

Design: definition of longevity cases

7

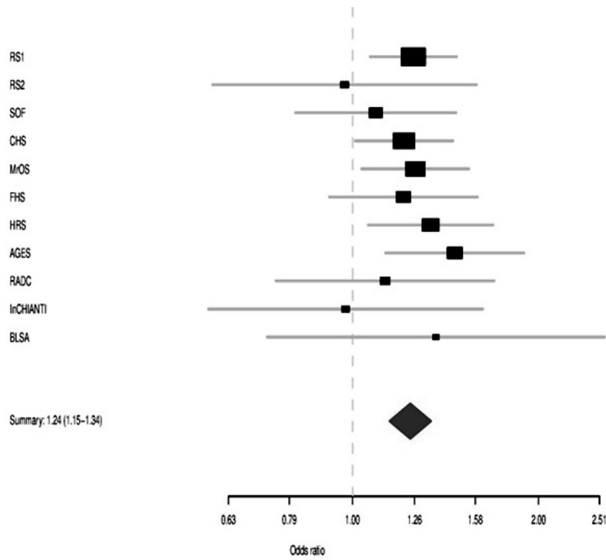
FIGURES

Supplement Figure 1 | Manhattan plot

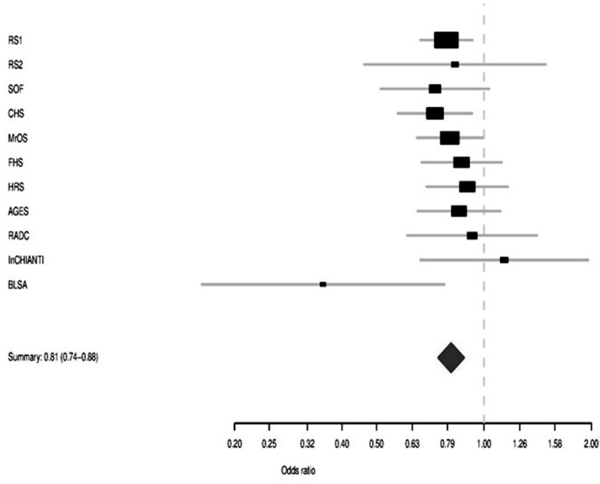


Supplement Figure 2 | Forest plots

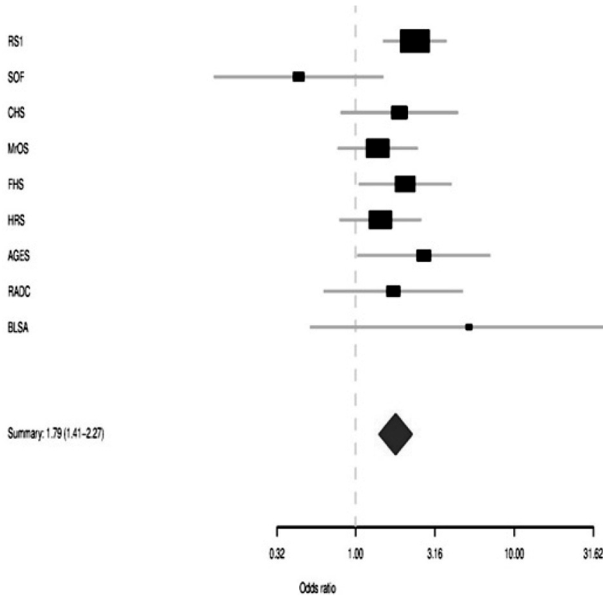
A. GRIK2



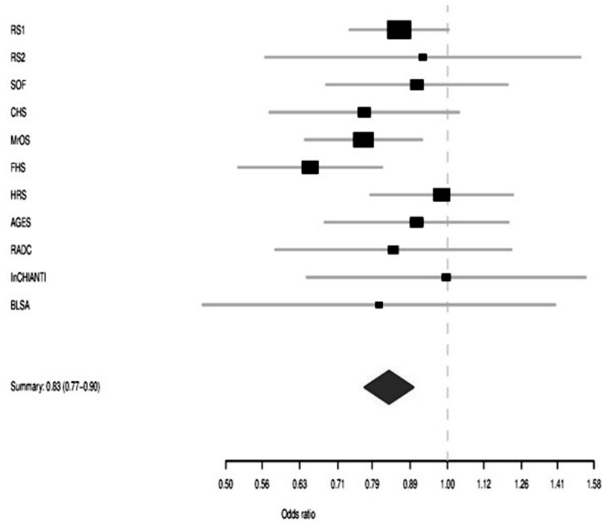
B. CADM2



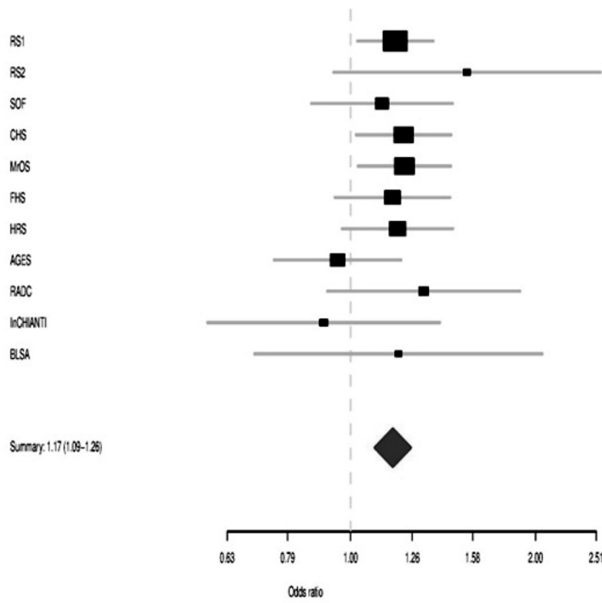
C. RGS7



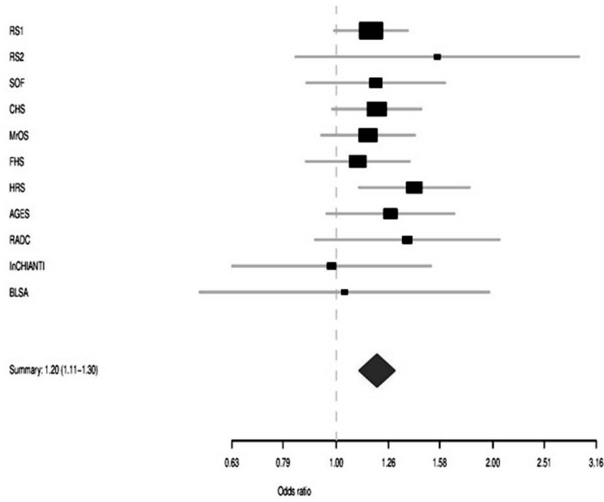
D. SOX6



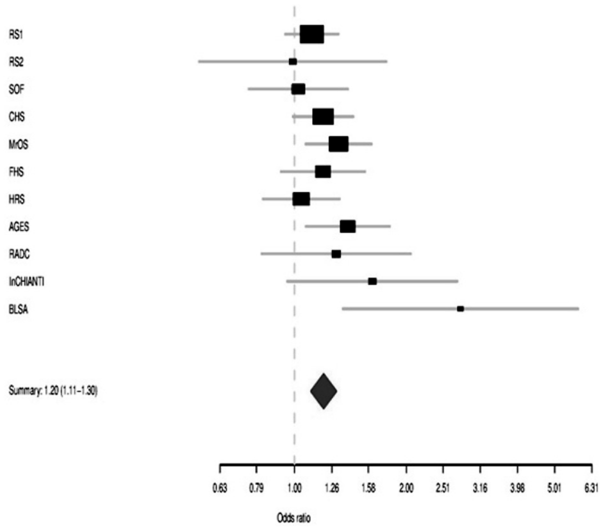
E. MBOAT1



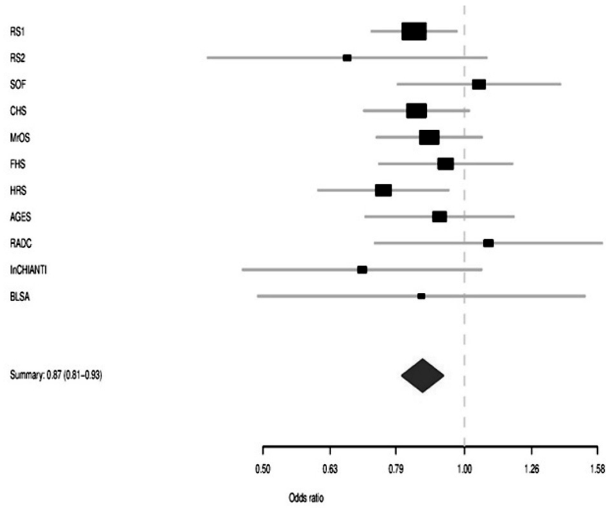
F. PFKM



G. LIMCH1



H. FOXO3

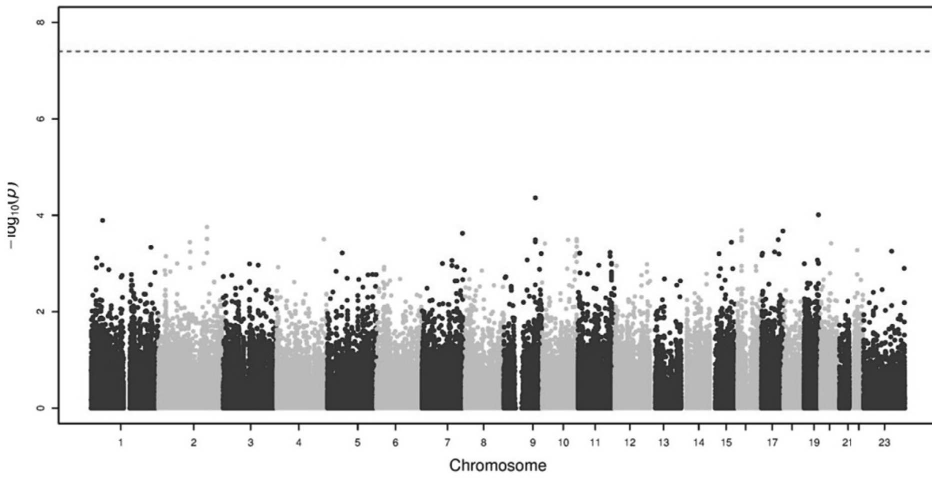


APPENDIX CHAPTER 2.2

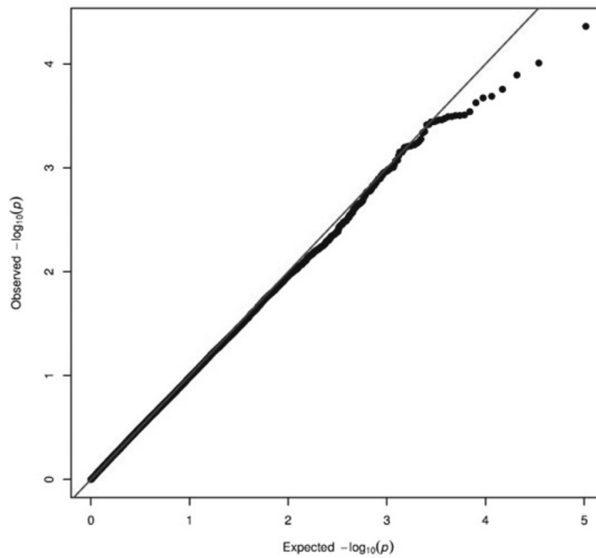
Supplementary figures

Supplement Figure 1 | Summary results of single variant analysis

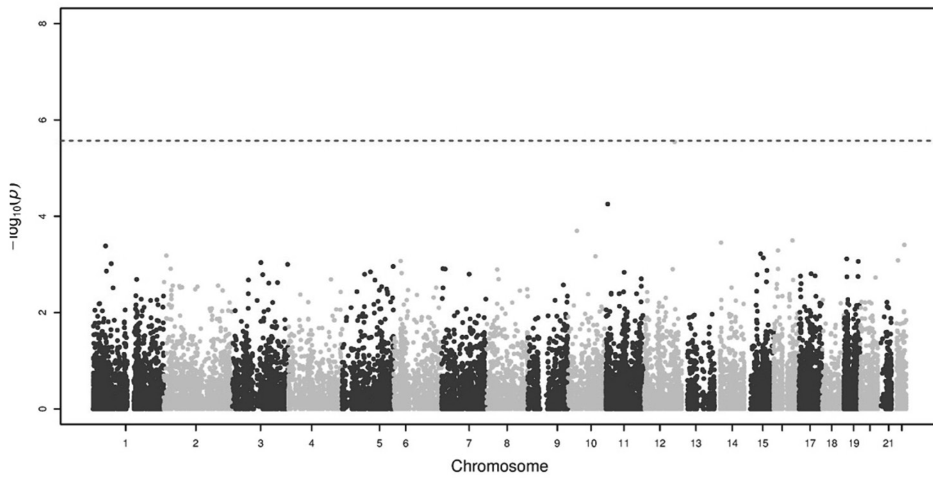
A. Manhattan plot



B. QQ plot (MAF > 0.05)



Supplement Figure 2 | Results of gene-based sequence kernel association test (SKAT) with longevity.



Supplementary Tables

Supplement Table 1 | Single Variant analysis results of candidate gene *APOE*.

SNP	CHR	BP	Gene	RA	EA	EAF	OR	P
19:45411057	19	45411057	APOE	G	A	0.005	3.70E-10	0.9993
19:45411110	19	45411110	APOE	T	C	0.004	9.62E+08	0.9993
19:45411145	19	45411145	APOE	G	C	0.001	1.69	0.7432
19:45411147	19	45411147	APOE	G	T	4.26E-04	3.80E-10	0.9993
19:45411149	19	45411149	APOE	A	C	4.17E-04	3.71E-10	0.9993
19:45411153	19	45411153	APOE	A	T	4.24E-04	3.70E-10	0.9993
19:45411941	19	45411941	APOE	T	C	0.195	0.8995	0.8308
19:45412408	19	45412408	APOE	C	G	0.002	4.13E-10	0.9993
19:45412498	19	45412498	APOE	C	A	0.005	7.46E+08	0.9993
19:45412500	19	45412500	APOE	A	T	0.005	7.54E+08	0.9993

Supplement Table 2 | PANTHER pathway analysis on SVA identified genes (GO: biological processes).

Biological Process	REFLIST (20000)	Observed	Expected	Over/Under	P-value
protein transport	1724	10	2.76	+	2.46E-04
intracellular protein transport	1724	10	2.76	+	2.46E-04
transport	2984	13	4.77	+	3.80E-04
cytokinesis	228	4	0.36	+	4.71E-04
vesicle-mediated transport	1240	8	1.98	+	5.97E-04
mitosis	621	5	0.99	+	0.003
homeostatic process	185	3	0.30	+	0.003
RNA localization	81	2	0.13	+	0.008
cellular component morphogenesis	1118	6	1.79	+	0.008
anatomical structure morphogenesis	1118	6	1.79	+	0.008
cellular glucose homeostasis	84	2	0.13	+	0.008
female gamete generation	301	3	0.48	+	0.012
nuclear transport	126	2	0.20	+	0.017
localization	160	2	0.26	+	0.027
cellular component organization	1496	6	2.39	+	0.029
protein targeting	193	2	0.31	+	0.038
JAK-STAT cascade	223	2	0.36	+	0.049

Supplement Table 3 | Overlap significant findings SVA and SKAT and functional information.

		SVA		SKAT	Function	
Genes	**Full name**	**SNP**	**p-value**	**p-value**	**GO: bio process**	**KEGG**
WSCD2	WSC domain containing 2	NA	NA	2.87E-06	long-term memory	calcium signaling pathway
OR52L1	olfactory receptor, family 52, subfamily L, member 1	rs61752597	6.09E-04	5.58E-05		
KIF5B	kinesin family member 5B	NA	NA	2.01E-04		
CES2	carboxylesterase 2	NA	NA	3.17E-04	lipid glycosylation	drug metabolism fatty acid metabolism
PSMB5	proteasome subunit beta type 5	NA	NA	3.51E-04		poteasome PD / AD / Huntington
LDOC1L	leucine zipper, down-regulated in cancer 1-like	NA	NA	3.92E-04		non-homologous end-joining
ELOVL1	elongation of very long chain fatty acid elongase 1	NA	NA	4.13E-04	lipid metabolism	lipid metabolism
MYH11	myosin, heavy chain 11, smooth muscle	rs1050163	2.05E-04	5.12E-04	muscle	muscle
TEX9	testis expressed 9	NA	NA	5.98E-04		
YWHAQ	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, theta polypeptide	NA	NA	6.54E-04		DNA replication
BLNK	B-cell linker	rs74151283	3.24E-04	6.77E-04	B cell	B cell receptor signaling pathway
MEGF11	multiple EGF-like domains 11	NA	NA	7.31E-04	glial cell fate commitment	
ZNF562	Zinc-finger 562	NA	NA	7.64E-04		
VPREB1	pre-B lymphocyte gene 1	NA	NA	8.22E-04		
HIST1H4B	histone cluster 1, H4b	NA	NA	8.46E-04	nucleosome / chromatin assembly	sytemic lupus erythematosus
ZNF534	Zinc-finger 534	rs1366258	9.79E-05	8.65E-04		
PCNP	PEST proteolytic signal containing nuclear protein	NA	NA	9.12E-04	nuclear transport	ubiquitin spliceosome

		SVA		SKAT	Function	
Genes	**Full name**	**SNP**	**p-value**	**p-value**	**GO: bio process**	**KEGG**
EFCAB7	EF-hand calcium binding domain 7	NA	NA	9.63E-04	centrosome cycle DNA integration	
SLC51A	solute carrier family 51, alpha subunit	NA	NA	9.96E-04		
DFNB31	deafness, autosomal recessive 31	rs4979386	4.35E-05	NA	mineralocorticoid metabolic process	
FOXJ3	forkhead box J3	rs343376	1.28E-04	NA		
NPLOC4	nuclear protein localization 4 homolog	rs3934711	2.12E-04	NA	ubiquitin protein catabolic processes	proteasome ubtiquitin
ABCF2	ATP-binding cassette sub-family F, member 2	rs7782699	2.36E-04	NA	RNA processing	RNA processing
ITGA4	integrin alpha 4	rs1143676	3.10E-04	NA	positive regulation of phosphatase activity	immune system calcium signaling
MKI67	antigen identified by monoclonal antibody Ki-67	rs2853344	3.13E-04	NA	M phase of mitotic cell cycle	
ASB5	ankyrin repeat and SOCS box containing 5	rs6827525	3.14E-04	NA	muscle	cardiac muscle
DDX18	DEAD box polypeptide 18	rs1052637	3.61E-04	NA	RNA processing	RNA Spliceosome
FAH	fumarylacetoacetate hydrolase	rs1370274	3.62E-04	NA	catabolic processes	(AA-)metabolism
PTPRT	protein tyrosine phosphatase receptor type, T	rs6072668	3.82E-04	NA	neuron-neuron synaptic transmission	neuroactive ligand-receptor interaction
USP6NL	USP6 N-terminal like	rs12570211	3.86E-04	NA		B cell receptor signaling pathway

7

Genes	Full name	SVA		SKAT	Function	
		SNP	p-value	p-value	GO: bio process	KEGG
HHIPL2	HHIP-like 2 (hedgehog interacting protein)	rs3748665	4.61E-04	NA		
AP1B1	adaptor-related protein complex 1, beta 1 subunit	rs2072051	5.29E-04	NA	endocytosis	endocytosis
MORC4	MORC family CW-type zinc finger 4	rs17253753	5.56E-04	NA		alanine and glutamate metabolism
MARCO	macrophage receptor with collagenous structure	rs6761637	5.73E-04	NA	complement activation	complement cascade prion disease
MYCBPAP	MYCBP associated protein	rs28498091	5.75E-04	NA	regulation of neural precursor cell proliferation	
HYOU1	hypoxia up-regulated 1	rs568922	5.87E-04	NA		protein export
MIR449C	microRNA 449C	rs35770269	6.01E-04	NA		
CTNS	cystinosin, lysosomal cystine transporter	rs2873624	6.12E-04	NA	ion transport	lysosome sphingolipid metabolism
UBAC1	UBA domain containing 1	rs1044193	6.22E-04	NA	cellular respiration	PD / AD / Huntington citrate cycle
SLC12A6	solute carrier family 12 (potassium/ chloride transporters), member 6	rs74010036	6.25E-04	NA		non-small cell lung cancer
CACNG1	calcium channel, voltage-dependent, gamma subunit 1	rs2363844	6.33E-04	NA	muscle	cardiac muscle
PTPRE	protein tyrosine phosphatase receptor type, E	rs41282874	6.36E-04	NA	hemopoietic progenitor cell differentiation	natural killer cell mediated cytotoxicity
WDR81	WD repeat domain 81	rs3809870	6.78E-04	NA	lysosome autophagy transferrin transport	lysosome glycosphingolipid biosynthesis

Genes	Full name	SVA		SKAT	Function	
		SNP	p-value	p-value	GO: bio process	KEGG
WDR11	WD repeat domain 11	rs10886789	7.06E-04	NA		Inositol phosphate metabolism
FAM179A	family with sequence similarity 179, member A	rs13009279	7.07E-04	NA	ciliary or flagellar motility	
KIF17	kinesin family member 17	rs479323	7.66E-04	NA	regulation of catenin import into nucleus	
HNRNPK	heterogeneous nuclear ribonucleoprotein K	rs3737135	8.45E-04	NA	RNA processing	spliceosome
NUCB1	nucleobindin 1	rs12460533	8.46E-04	NA	post-translational protein modification protein folding	N-glycan biosynthesis protein export
DOCK4	dedicator of cytokinesis 4	rs12154775	8.63E-04	NA		Glycolysis Axon guidance
OR8G2	olfactory receptor, family 8, subfamily G, member 2	rs11219508	9.86E-04	NA		
LRP2	low density lipoprotein receptor-related protein 2	rs62171262	9.89E-04	NA	tissue homeostasis	steroid hormone biosynthesis

NA = p-value not below 10^{-3} in this analysis

7

Supplement Table 4 | PANTHER pathway analysis on SKAT identified genes (GO: biological processes).

Biological Process	REFLIST (20000)	Observed	Expected	Over/Under	P-value
vitamin metabolic process	80	1	0.06	+	0.062
response to toxin	99	1	0.08	+	0.076
neuromuscular synaptic transmission	121	1	0.10	+	0.093
meiosis	166	1	0.13	+	0.125
cell cycle	1602	3	1.28	+	0.132
hemopoiesis	183	1	0.15	+	0.137
cytokinesis	193	1	0.15	+	0.144
protein amino acid glycosylation	203	1	0.16	+	0.151
chromosome segregation	210	1	0.17	+	0.155
fatty acid metabolic process	227	1	0.18	+	0.167
calcium-mediated signaling	256	1	0.20	+	0.186
polysaccharide metabolic process	261	1	0.21	+	0.190

Supplement Table 5 | Burden of damaging variants.

	score_cases	score_controls	beta	Se	p-value
damaging (all)	0.995 (0.0003)	0.995 (0.0003)	-370.12	308.72	0.231
damaging (MAF < 0.05)	0.999 (0.0003)	0.999 (0.0003)	-286.78	308.29	0.352
damaging (MAF < 0.01)	0.999 (0.0003)	0.999 (0.0003)	-276.58	319.89	0.387

APPENDIX CHAPTER 4.3

Supplementary Table 1 | Association of known height SNPs with telomere length (all).

				height_GIANT		telo_ENGAGE			
SNP	Chr	Gene	EA	beta	p-value	beta	se	p-value	N
rs17081935	4	POLR2B	T	0.030	4.77E-08	-0.027	0.009	0.002	37633
rs724016	3	ZBTB38	G	0.070	4.47E-52	-0.020	0.007	0.006	37323
rs2079795	17	TBX2	T	0.040	1.22E-16	0.018	0.008	0.022	37205
rs2780226	6	HMGA1	C	0.076	1.02E-18	-0.038	0.017	0.024	32536
rs634552	11	SERPINH1	T	0.039	1.35E-09	-0.026	0.012	0.026	29826
rs751543	9	PAPPA	T	0.026	4.51E-08	-0.019	0.009	0.034	32272
rs6569648	6	L3MBTL3	C	0.040	8.93E-12	0.019	0.009	0.039	34469
rs6684205	1	TGFB2	G	0.028	1.97E-11	-0.016	0.008	0.043	37638
rs6439167	3	C3orf47	C	0.034	7.20E-10	0.019	0.009	0.044	34618
rs6449353	4	LCORL	T	0.075	1.35E-27	0.021	0.011	0.052	37616
rs6879260	5	GFPT2	C	0.022	5.61E-10	0.015	0.009	0.075	29430
rs16942341	15	ACAN	C	0.130	1.28E-17	0.057	0.033	0.083	23329
rs42235	7	CDK6	T	0.057	7.33E-28	0.015	0.009	0.088	32470
rs6473015	8	PEX2	C	0.029	1.67E-10	0.014	0.008	0.090	37658
rs2237886	11	KCNQ1	T	0.046	3.12E-08	-0.021	0.013	0.093	33710
rs9360921	6	SENP6	G	0.042	4.56E-11	0.019	0.011	0.099	37676
rs473902	9	PTCH1/FANCC	T	0.065	1.70E-14	-0.030	0.018	0.108	27066
rs6457620	6	HLA_locus	G	0.029	3.65E-08	-0.012	0.008	0.138	32735
rs227724	17	NOG	T	0.030	1.18E-08	0.012	0.008	0.145	32509
rs12680655	8	ZFAT	C	0.028	4.83E-11	-0.011	0.007	0.146	37645
rs2341459	2	C2orf34	T	0.025	3.58E-08	0.012	0.008	0.150	37587
rs9844666	3	PCCB	G	0.024	3.05E-08	0.012	0.008	0.154	37669
rs7460090	8	SDR16C5	T	0.058	9.55E-16	0.015	0.011	0.158	37404
rs2154319	1	SCMH1	C	0.030	4.30E-10	0.013	0.009	0.164	32311
rs788867	4	PRKG2/BMP3	G	0.043	1.78E-15	-0.011	0.008	0.172	30251
rs2256183	6	MICA	A	0.040	2.67E-14	-0.010	0.008	0.186	32639
rs11259936	15	ADAMTSL3	C	0.044	2.18E-21	-0.009	0.007	0.198	37645
rs3129109	6	OR2J3	C	0.032	3.28E-08	-0.011	0.008	0.199	32662
rs720390	3	IGF2BP2	A	0.029	1.62E-10	0.010	0.008	0.213	32280
rs2145998	10	PPIF	T	0.026	2.68E-08	-0.009	0.007	0.214	37409
rs4665736	2	DNAJC27	T	0.029	1.44E-13	0.009	0.007	0.214	37339
rs7319045	13	GPC5	A	0.025	4.51E-10	-0.010	0.008	0.216	32666
rs2110001	7	TMEM176A	G	0.031	9.83E-10	0.010	0.009	0.254	32299
rs422421	5	FGFR4/NSD1	C	0.031	1.43E-09	-0.012	0.011	0.260	29617

				height_GIANT		telo_ENGAGE			
SNP	**Chr**	**Gene**	**EA**	**beta**	**p-value**	**beta**	**se**	**p-value**	**N**
rs17511102	2	CDC42EP3	T	0.060	1.26E-12	-0.016	0.014	0.268	34081
rs1173727	5	NPR3	T	0.034	3.97E-15	0.008	0.007	0.274	37423
rs10748128	12	FRS2	T	0.038	3.76E-11	-0.010	0.009	0.288	31074
rs9967417	18	DYM	G	0.038	2.57E-16	0.008	0.008	0.291	37336
rs2665838	17	CSH1/GH1	G	0.042	2.02E-13	-0.009	0.008	0.302	37151
rs310405	6	FAM46A	A	0.026	3.59E-11	0.008	0.008	0.312	34561
rs7466269	9	FUBP3	A	0.032	1.17E-14	-0.007	0.007	0.324	37404
rs2066807	12	STAT2	G	0.054	9.61E-09	-0.014	0.015	0.337	37399
rs143384	20	GDF5	G	0.063	4.94E-39	-0.008	0.008	0.349	30908
rs17780086	17	LRRC37B	A	0.028	4.39E-08	-0.010	0.011	0.350	37504
rs7849585	9	QSOX2	T	0.029	3.43E-11	-0.007	0.008	0.364	35917
rs2093210	14	SIX6	C	0.032	2.30E-12	0.007	0.008	0.367	32418
rs274546	5	SLC22A5	G	0.029	8.53E-10	-0.007	0.007	0.367	37537
rs3791675	2	EFEMP1	C	0.053	2.40E-20	0.012	0.013	0.367	37455
rs3764419	17	ATAD5/RNF135	C	0.035	8.87E-16	0.007	0.008	0.370	37152
rs889014	5	BOD1	C	0.030	4.51E-10	-0.007	0.007	0.379	37436
rs1351394	12	HMGA2	T	0.060	7.78E-34	-0.006	0.007	0.386	37210
rs2871865	15	IGF1R	C	0.057	1.07E-12	-0.013	0.015	0.388	29623
rs26868	16	CASKIN1	A	0.034	3.47E-08	-0.008	0.010	0.406	22771
rs11830103	12	SBNO1	G	0.035	3.79E-10	-0.008	0.010	0.410	32656
rs11684404	2	EIF2AK3	C	0.028	6.42E-09	-0.006	0.007	0.412	37478
rs10037512	5	MEF2C	T	0.032	3.82E-09	-0.006	0.008	0.434	34355
rs7274811	20	ZNF341	G	0.041	6.83E-14	-0.007	0.008	0.440	37574
rs806794	6	Histone_cluster	A	0.052	5.54E-26	-0.006	0.008	0.440	35315
rs4986172	17	ACBD4	C	0.032	7.12E-09	-0.006	0.008	0.446	34444
rs3782089	11	SSSCA1	C	0.058	5.89E-09	0.013	0.017	0.461	34332
rs11144688	9	PCSK5	G	0.049	1.48E-09	-0.013	0.018	0.466	16752
rs2336725	3	RTF1	C	0.027	3.52E-08	0.006	0.008	0.466	32128
rs7155279	14	TRIP11	G	0.024	8.91E-10	-0.008	0.011	0.470	37115
rs9969804	9	IPPK	A	0.030	5.61E-10	-0.005	0.007	0.471	37635
rs543650	6	ESR1	G	0.034	1.36E-09	-0.006	0.009	0.476	27510
rs10152591	15	TLE3	A	0.041	3.54E-08	0.009	0.013	0.477	37495
rs526896	5	PITX1	T	0.030	1.93E-09	0.006	0.009	0.488	31009
rs8052560	16	CTU2/GALNS	A	0.029	1.45E-08	-0.009	0.013	0.497	24409
rs237743	20	ZNFX1	A	0.041	7.20E-10	0.006	0.009	0.505	37654
rs1013209	8	ADAM28	C	0.025	4.51E-08	-0.005	0.008	0.515	37411
rs2145272	20	BMP2	G	0.039	5.87E-16	0.005	0.008	0.530	34480
rs3118905	13	DLEU7	G	0.056	2.99E-25	-0.007	0.013	0.578	37592

				height_GIANT		telo_ENGAGE			
SNP	**Chr**	**Gene**	**EA**	**beta**	**p-value**	**beta**	**se**	**p-value**	**N**
rs11118346	1	LYPLAL1	C	0.025	2.19E-09	0.004	0.007	0.581	37593
rs10838801	11	PTPRJ/SLC39A13	G	0.027	1.76E-10	0.005	0.008	0.584	32748
rs2834442	21	KCNE2	A	0.026	7.30E-09	0.006	0.011	0.602	37420
rs7759938	6	LIN28B	C	0.045	8.69E-18	-0.004	0.008	0.605	37527
rs9428104	1	SPAG17	G	0.041	8.88E-13	-0.004	0.008	0.616	37571
rs572169	3	GHSR	T	0.033	9.90E-14	0.004	0.008	0.619	37300
rs3812163	6	BMP6	T	0.036	6.66E-16	-0.003	0.008	0.644	37363
rs2629046	2	SERPINE2	T	0.024	2.17E-08	0.003	0.007	0.661	37333
rs7763064	6	GPR126	G	0.048	6.41E-19	-0.003	0.008	0.690	37316
rs9472414	6	SUPT3H/RUNX2	T	0.026	2.43E-08	0.003	0.009	0.707	37640
rs6457821	6	PPARD/FANCE	C	0.104	1.82E-11	0.014	0.037	0.715	22813
rs12982744	19	DOT1L	G	0.030	2.79E-12	-0.003	0.007	0.727	37444
rs1741344	20	SMOX	C	0.023	3.52E-08	-0.003	0.008	0.755	33811
rs10799445	1	JMJD4	A	0.032	1.18E-08	-0.003	0.009	0.756	37473
rs6470764	8	GSDMC	C	0.050	5.88E-17	-0.003	0.009	0.758	37498
rs4282339	5	SLIT3	G	0.036	3.38E-10	0.003	0.009	0.768	37537
rs1046934	1	TSEN15	C	0.044	6.43E-22	-0.002	0.007	0.776	37587
rs11205277	1	SF3B4	G	0.046	1.20E-18	0.002	0.009	0.787	25438
rs2856321	12	ETV6	G	0.029	1.48E-10	-0.002	0.008	0.798	37517
rs12470505	2	CCDC108/IHH	T	0.041	1.29E-10	0.003	0.012	0.799	37635
rs798489	7	GNA12	C	0.048	8.47E-25	-0.002	0.008	0.818	37076
rs11107116	12	SOCS2	T	0.052	1.75E-23	0.002	0.009	0.823	37626
rs7689420	4	HHIP	C	0.073	1.38E-29	0.002	0.010	0.823	34731
rs4800452	18	CABLES1	T	0.051	2.36E-17	0.002	0.009	0.824	32684
rs4965598	15	ADAMTS17	C	0.028	1.36E-13	-0.002	0.008	0.830	34140
rs2284746	1	MFAP2	G	0.040	5.64E-15	-0.002	0.008	0.847	32347
rs11958779	5	SLC38A9	G	0.027	8.04E-09	0.002	0.012	0.855	32735
rs1325598	1	PAPPA2	G	0.022	1.61E-08	-0.001	0.007	0.856	37571
rs1582931	5	CEP120	G	0.023	2.08E-08	0.001	0.008	0.869	32425
rs5742915	15	PML	C	0.031	2.99E-10	-0.001	0.008	0.878	32227
rs1490384	6	C6orf173	T	0.034	3.22E-16	0.001	0.007	0.879	37406
rs4470914	7	TWISTNB	T	0.029	3.78E-08	0.001	0.010	0.899	37266
rs17318596	19	ATP5SL	A	0.032	3.01E-09	-0.001	0.008	0.903	34797
rs12153391	5	FBXW11	C	0.030	8.75E-10	-0.001	0.009	0.910	32573
rs7027110	9	ZNF462	A	0.031	1.34E-10	0.001	0.009	0.918	37657
rs10770705	12	SLCO1C1	A	0.033	4.62E-11	-0.001	0.008	0.920	37391
rs1950500	14	NFATC4	T	0.034	3.94E-11	0.000	0.008	0.954	37608
rs17346452	1	DNM3	C	0.040	3.28E-14	0.000	0.008	0.962	37439

				height_GIANT		telo_ENGAGE			
SNP	**Chr**	**Gene**	**EA**	**beta**	**p-value**	**beta**	**se**	**p-value**	**N**
rs9863706	3	RYBP	C	0.031	1.47E-08	0.000	0.009	0.966	37213
rs2580816	2	NPPC	C	0.045	1.78E-12	0.000	0.010	0.968	35820
rs2638953	12	CCDC91	C	0.032	8.45E-14	0.000	0.008	0.970	36180
rs4821083	22	SYN3	T	0.031	4.79E-08	0.000	0.011	0.990	34613
rs1708299	7	JAZF1	A	0.040	1.48E-17	0.000	0.008	0.994	37519
OVERALL						0.000	0.001	0.925	

EA = Effective allele
Telo_ENGAGE = look up of known height SNPs in telomere length GWAS
Height_GIANT = genome-wide significant results from published height GWAS

APPENDIX CHAPTER 4.4

Supplementary methods

Telomere length

ERF; KORA; LLS; TwinsUK

Mean leucocyte telomere length was measured established and validated quantitative PCR-based technique. This method expresses telomere length as a ratio (T/S) of telomere repeat length (T) to copy number of a single copy gene, 36B4(S), within each sample. DNA samples were run in duplicate in 25µL reactions containing 1x SensiMix NoRef SYBR Green master mix (Quantace, UK), 300nM of Tel1b (5′ CGGTTTGTTTGGGTTTGGGTTTGGGTTTGGGTTTGGGTT 3′), 300nM Tel2b (5′ GGCTTGCCTTACCCTTACCCTTACCCTTACCCTTACCCT 3′)primers and 30ng of template DNA. For the 36b4 PCR the above was performed, substituting the Tel primers for 300nM 36b4U (5′ CAGCAAGTGGGAAGGTGTAATCC 3′) and 500nM of 36B4D (5′ CCCATTCTATCATCAACGGGTACAA 3′).

All reactions were set up using a Qiagility liquid handling system (Qiagen, UK) and run on a Rotorgene-Q Real Time Thermal Cycler (Qiagen, UK). Alongside the samples each run also contained a Calibrator sample (genomic DNA from the K562 cell line) in duplicate and a no template control. Cycling conditions were 95^{0}C incubation for 10mins followed by either 20 cycles (telomere) or 30 cycles (36B4) of 95C for 15 sec and 58C for 1min. Dilution series (100-1.56ng in two-fold dilutions) were run for both telomere and 36B4 assays to establish the linear range. Good linearity was observed across this range ($r^2 > 0.99$) and the input amount of 30ng was subsequently set. Any samples found to run outside this range were diluted and run again.

Analysis of the PCR output was performed using Comparative Quantitation (Corbett Life Science Rotorgene analysis software version 1.7, Corbett Life Science, UK) as previously described.15 A takeoff value was produced for each sample based on the second derivative of the amplification plot, which represents the start of the exponential phase. The amplification efficiency for each sample and subsequently the mean amplification across the run was calculated. The relative concentration of the sample was then calculated relative to the calibrator sample using the takeoff value and amplification efficiency. This was performed for both the telomere (T) and 36B4 (S) assays and telomere length expressed as the ratio of these (T/S). For quality control all samples were checked for concordance between duplicate values. Samples showing a difference between the duplicate measurements of greater than 0.2 cycles in the takeoff value were excluded and re-run. Mean inter-run coefficients of variation for the cohorts measured with this technique ranged from 2.7-3.6%.

SAPHIR

Genomic DNA was extracted from blood samples with the Qiagen EZ1®DNA Blood 200 µl Kit using the Qiagen EZ1 advanced Biorobot. DNA concentrations were quantified with the Tecan NanoQuant infinite M200. Samples were normalized in 96-well microtiter plates.

Samples were ascertained in a singleplex, quadruplicate approach to measure the T/S-ratios. These T/S-ratios are proportional to individual relative telomere length (RTL).

RTL was measured with some modifications by using a quantitative real-time polymerase chain reaction (qPCR) assay, which was first described by Cawthon [1] and we modified as described below.

Each qPCR was carried out in 384-well format which was vertically segmented in two parts: telomere (T) and housekeeping gene 36B4 (S).Telomere (T) and housekeeping gene (S) PCRs were identical composed except for the primers. DNA samples were run in 15µl reactions containing 1x Quantifast™ SYBR® Green PCR mastermix (Qiagen), 10 ng of DNA, 1µM of telomere primer and 250nm of housekeeping gene 36B4 primer, respectively. The primer sequences (5'→3') were:

tel1b CGGTTTGTTTGGGTTTGGGTTTGGGTTTGGGTTTGGGTT;

tel2b GGCTTGCCTTACCCTTACCCTTACCCTTACCCTTACCCT;

36B4u CAGCAAGTGGGAAGGTGTAATCC;

36B4d CCCATTCTATCATCAACGGGTACAA [2].

Each 384-well plate contained the standard DNA, a positive control (commercially available DNA-Human Genomic DNA, Roche) and a non template control (NTC) in quadruplicate.

All sample transfers and dilution steps were performed with a Tecan robotic workstation.

Relative qPCR was carried out on an Applied Biosystems Taqman Fast Real-Time PCR 7900HT System. The thermal cycling began with the initial polymerase activation step (10 min at 95°C) and was followed by 40 cycles of 95°C for 15 s, 60°C for 1 min. A melting curve analysis to verify the specificity and identity of the products was performed.

The relative quantities were determined by the efficiency correction method [3], which does not need calibration curves and includes the individual real-time PCR efficiencies. This mathematical model calculates the ratio of a target gene (telomere) from the efficiencies and C_t-value of an experimental sample versus a standard in comparison to a reference gene (housekeeping gene). To calculate PCR efficiencies of both the reference gene and the target gene PCR raw data were imported into the program LinRegPCR (version 12.5.0) [4]. Efficiencies were computed for all replicates of each sample. To check PCR data for outliers, the coefficients of variation (intra-assay CVs) were assessed for the C_t-values and the efficiency-values of each sample in each gene. All single outlying values (CV > 5%) were removed from further analyses. In the case of the housekeeping gene 36B4, less than 1% and in case of telomere, about 3% were removed. For further mathematical analysis, the mean value of all efficiency-values of each gene on each plate and the mean C_t-value of the

four replicates for each gene and each sample were used. Relative T/S-ratios reflect relative telomere length differences of the samples.

$$Relative\, T/S - Ratio = \frac{eff\ (\text{tel,sample})^{Ct\ (\text{tel,sample})}}{eff\ (\text{ref,sample})^{Ct\ (\text{ref,sample})}} \div \frac{eff\ (\text{tel,standard})^{Ct\ (\text{tel,standard})}}{eff\ (\text{ref,standard})^{Ct\ (\text{ref,standard})}}$$

To test the reproducibility of RTL measurement, the inter-assay CV of T/S-ratios was calculated according to the following formula

$$\frac{\sqrt[2]{\frac{\sum_{i=1}^{n}(x_i - y_i)^2}{2n}}}{(\bar{X}+\bar{Y})\div 2}.$$

6.8% of all samples were analyzed in duplicate. Duplicate samples were never positioned on the same plate or at the same plate position. Inter-assay CV of T/S-ratios was 6%. As second RTL measurement quality control, we analyzed the relative telomere length of a commercially available DNA, which was positioned on each 384-well plate (a total of 39 plates). Inter-assay CV of T/S-ratios of this 39 times analyzed sample was 7.7%.

Finnish Twins

Leukocyte telomere length was measured from DNA extracted from peripheral blood. We used a qPCR-based method [1], as described previously [5]. We used β-hemoglobin (Hgb) as a single copy reference gene. Separate reactions for telomere and Hgb reaction were carried out in paired 384-well plates in which matched sample well positions were used. Ten nanograms of DNA were used for each reaction, performed in triplicate. Every plate included a 7-point standard curve, which was used to create a standard curve and to perform absolute quantification of each sample. Samples and standard dilutions were transferred into the plates using a DNA Hydra 96 robot and dried overnight at +37°C. Specific reaction mix for telomere reaction included 270 nM tel1b primer (5´-CGGTTT(GTTTGG)5GTT-3´) and 900 nM tel2b primer (5´-GGCTTG(CCTTAC)5CCT-3´), 150 nM ROX (Invitrogen), 0.2X SYBR Green I (Invitrogen), 5 mM DTT (Sigma-Aldrich), 1% DMSO (Sigma-Aldrich), 0.2 mM of each dNTP (Fermentas), and 1.25 U AmpliTaq Gold DNA polymerase (Applied Biosystems) in a total volume of 15 µl AmpliTaq Gold Buffer I. Hgb reaction mix included 300 nM Hgb1 primer (5´-GCTTCTGACACAACTGTGTTCACTAGC-3´) and Hgb2 primer (5´-CACCAACTTCATCCACGTTCACC-3´) in a total volume of 15 µl of iQ SyBrGreen supermix (BioRad). The cycling conditions for telomere amplification were: 10 minutes at 95°C followed by 25 cycles at 95°C for 15 s and 54°C for 2 min with signal acquisition. The cycling conditions for Hgb amplification were: 95 °C for 10 min followed by 35 cycles at 95°C for 15 s, 58°C for 20 s, 72°C for 20 s with signal acquisition. Reactions were performed with CFX384 Real-Time PCR Detection System (Bio-Rad). Melt-curve analysis was carried out in the end of the run to ensure specific primer binding.

We used the Bio-Rad CFX Manager software to perform quality control, and samples with standard deviation of > 0.5 between triplicates were omitted from the analysis. Plate effect was taken into account by analyzing five genomic DNA control samples on every plate. We normalized the telomere and Hgb signal values separately to the mean of these control samples before taking the T/S ratio. The control samples were used for calculating the coefficient of variation (CV) values, which was 2.8% for the telomere reaction, 6.4% for the Hgb reaction, and 8.2% for their ratio (T/S).

Adiponectin/leptin/CRP

ERF

Blood from participants was obtained in a fasted state. Total plasma adiponectin measurements were analyzed with a human adiponectin RIA kit (Linco Research). Plasma leptin was determined using a custom made duplex assay (cat.#: N45ZA-1) of Meso Scale Discovery (MSD). The Leptin/Resistin duplex assay was analyzed using a sector 2400 imager. Total plasma CRP measurements were analyzed with the US C-reactive protein ELISA (cat.# DSL-10-42100) of Diagnostic Systems Laboratories, Inc. All measurements were performed conform the manufactures protocol.

KORA F3 and F4

In the KORA F3 and F4 studies, non-fasting and fasting venous blood samples stored at -80°C were used to measure blood markers. Concentrations of CRP were measured by nephelometry using a BNII and the test kit from Siemens (Erlangen, Germany). CRP concentrations were approximately lognormal distributed and therefore log-transformed if used as a continuous variable. Serum leptin concentrations were determined using ELISAs from Mercodia, Stockholm, Sweden. The intra- and inter- assay coefficients of variation for leptin were < 10.0%.

In KORA F3, adiponectin levels were measured by an ELISA assay from Mercodia as described recently [6]. The quality of the measurements was monitored by 4 controls samples on each microtitre plate. The intra-assay and inter-assay coefficient of variation for adiponectin was 3.78% and < 10%, respectively. In KORA-F4, serum adiponectin levels were measured using the Quantikine ELISA kit from R&D Systems (Wiesbaden, Germany). The intra- and inter-assay coefficients of variation were 3.8% and 8.0%, respectively.

LLS

In the LLS all standard serum measurements were performed using fully automated equipment. Adiponectin levels were determined with the DuoSet ELISA kit (R&D Systems Europe Ltd, Abingdon, United Kingdom) according to manufacturer's instructions. The lower and upper detection limits of the assay were respectively 62.5 and 4000 pg/ml. Serum leptin was measured using 'two-step' sandwich ELISA kit (Diagnostics Biochem Canada Inc.,

Dorchester, Ontario, Canada) according to manufacturer's instructions. Assay sensitivity was 0.5 ng/ml. High-sensitivity C-reactive protein (hsCRP) was measured using the Hitachi Modular P800 (Roche, Almere, the Netherlands) [7]. Since the adiponectin, leptin and hsCRP levels were not normally distributed the log transformed values were used for analysis. For all serum parameters, measurements with a deviation from the mean > 3 SD were removed.

TwinsUK

Fasting morning serum total adiponectin levels were measured with a two-site DELFIA assay using antibodies and standards from R&D Systems (Minneapolis, MN). The day-to-day coefficient of variation (CV) for adiponectin was 9.9% at a concentration of 3.2 ng/ml, 7.8% at 8.5 ng/ml, and 5.2% at 14.7 ng/ml. Serum leptin concentration was determined after an overnight fast using a RIA (Linco Research, St Louis, MO). CRP concentrations from serum were measured with the Human Cardiovascular Disease (CVD) Panel 2 (acute-phase proteins) LINCOplex Kit (HCVD2-67BK) from Linco (Millipore) and with the Extracellular Protein Buffer Reagent Kit (LHB0001) from Invitrogen. CRP concentrations were expected to be very high and a dilution step was required prior to analysis. The optimal dilution which was not specified in the assay procedure was set at 1:2000. Sample analyses were performed according to the manufacturers' protocol and assayed in duplicate. Data was collected using the Luminex-100 system (Qiagen LiquiChip).

SAPHIR

Venous blood (EDTA) was collected after an overnight fast and full blood as well as serum/plasma was stored at -80°C. Serum adiponectin levels were measured using the human adiponectin ELISA kit (BioCat GmbH Heidelberg, Germany). hsCRP was measured by applying the Tina-quant Cardiac C-reactive Protein (Latex) High Sensitivity kit as described by the manufacturer (Roche) and analyzed by the Hitachi 911.

Finnish Twins

Venous blood samples for plasma leptin and adiponectin were obtained between 07:30 AM and 08:30 AM after an overnight fast. Leptin and adiponectin were analysed by enzyme-linked immunosorbent assay (ELISA) by using DuoSet ELISA, R&D Systems Europe Ltd, Abindgon, UK. Serum hsCRP was measured by particle-enhanced immunoturbidimetric assay (Cobas CRP(Latex)HS, Roche Diagnostics) on Modular automatic analyzer (Hitachi Ltd, Tokyo, Japan).

Supplementary tables

Supplementary Table 1 | Association of telomere length with CRP.

	Total				Men				Women			
Study	**n**	**r**	**se**	**p-value**	**n**	**r**	**se**	**p-value**	**n**	**r**	**se**	**p-value**
ERF	1,980	-0.01	0.02	0.738	849	-0.04	0.03	0.262	1,131	0.04	0.03	0.213
KORA F3	259	-0.04	0.06	0.493	143	-0.01	0.08	0.929	116	-0.08	0.09	0.409
KORA F4	3,009	-0.004	0.02	0.842	1,444	-0.03	0.03	0.197	1,565	0.02	0.03	0.357
LLS	2,230	-0.01	0.02	0.797	1,014	0.004	0.03	0.896	1,216	-0.01	0.03	0.687
TwinsUK	1,250	-0.01	0.03	0.685	0	NA	NA	NA	1,250	-0.01	0.03	0.685
SAPHIR	1,520	-0.005	0.03	0.858	982	-0.04	0.03	0.233	538	0.08	0.04	0.077
FTC	190	-0.04	0.07	0.563	107	0.03	0.10	0.725	83	-0.09	0.11	0.421
Meta-analysis	10,438	-0.01	0.01	0.444	4,539	-0.02	0.01	0.095	5,899	0.01	0.01	0.329

Supplementary Table 2 | Adiponectin

A. Model 2 (age + gender + BMI)

	Total				Men				Women			
Study	**n**	**r**	**se**	**p-value**	**n**	**r**	**se**	**p-value**	**n**	**r**	**se**	**p-value**
ERF	2,092	0.09	0.02	$7.25*10^{-5}$	904	0.06	0.03	0.070	1,188	0.11	0.03	$9.08*10^{-5}$
KORA F3	3,094	0.02	0.02	0.193	1,501	0.001	0.03	0.994	1,593	0.05	0.03	0.068
KORA F4	1,110	-0.05	0.03	0.127	571	-0.08	0.04	0.052	539	-0.01	0.04	0.837
LLS	1,892	$-3.26*10^{-4}$	0.02	0.989	861	0.04	0.03	0.243	1,031	-0.03	0.03	0.343
TwinsUK	1,184	-0.01	0.03	0.628	NA	NA	NA	NA	1,184	-0.01	0.03	0.628
SAPHIR	1,523	0.001	0.03	0.969	984	-0.01	0.03	0.661	539	0.02	0.04	0.660
FTC	190	-0.09	0.07	0.231	107	0.01	0.10	0.919	83	-0.24	0.11	0.026
Meta-analysis	11,085	0.02	0.01	0.104	4,928	0.01	0.01	0.669	6,157	0.02	0.01	0.062

B. Model 3 (age + gender + CRP)

	Total				Men				Women			
Study	**n**	**r**	**se**	**p-value**	**n**	**r**	**se**	**p-value**	**n**	**r**	**se**	**p-value**
ERF	1,980	0.09	0.02	$4.72*10^{-5}$	849	0.07	0.03	0.042	1,131	0.13	0.03	$1.61*10^{-5}$
KORA F3	258	0.08	0.06	0.228	143	-0.003	0.08	0.972	115	0.19	0.09	0.044
KORA F4	1,113	-0.04	0.03	0.176	572	-0.07	0.04	0.089	541	-0.002	0.04	0.968
LLS	1,864	0.001	0.02	0.968	848	0.02	0.03	0.532	1,016	0.02	0.03	0.607
TwinsUK	998	-0.003	0.03	0.923	NA	NA	NA	NA	998	-0.003	0.03	0.923
SAPHIR	1,521	0.01	0.03	0.846	983	-0.002	0.03	0.950	538	0.01	0.04	0.799
FTC	190	-0.04	0.07	0.580	107	0.06	0.10	0.533	83	-0.20	0.11	0.065
Meta-analysis	7,924	0.02	0.01	0.083	3,502	0.01	0.02	0.490	4,422	0.04	0.02	0.011

Supplementary Table 3 | Leptin

A. Model 2 (age + gender + BMI)

	Total				Men				Women			
Study	**n**	**r**	**se**	**p-value**	**n**	**r**	**se**	**p-value**	**n**	**r**	**se**	**p-value**
ERF	1,592	-0.04	0.03	0.140	687	-0.03	0.04	0.417	905	-0.05	0.03	0.145
KORA F3	1,605	-0.02	0.02	0.426	793	0.01	0.04	0.700	812	-0.05	0.04	0.130
KORA F4	2,995	-0.04	0.02	0.038	1,439	-0.06	0.03	0.016	1,556	$2.66*10^{-4}$	0.03	0.992
LLS	1,912	-0.07	0.02	0.002	872	-0.07	0.03	0.032	1,040	-0.07	0.03	0.021
TwinsUK	1,426	-0.05	0.03	0.055	NA	NA	NA	NA	1,426	-0.05	0.03	0.055
SAPHIR	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
FTC	190	-0.05	0.07	0.520	107	-0.11	0.10	0.250	83	0.07	0.11	0.523
Meta-analysis	9,720	-0.04	0.01	$1.93*10^{-5}$	3,898	-0.05	0.02	0.005	5,822	-0.04	0.01	0.003

B. Model 3 (age + gender + CRP)

	Total				Men				Women			
Study	**n**	**r**	**se**	**p-value**	**n**	**r**	**se**	**p-value**	**n**	**r**	**se**	**p-value**
ERF	1,559	-0.05	0.03	0.068	683	-0.04	0.04	0.290	876	-0.08	0.03	0.014
KORA F3	180	-0.13	0.07	0.081	102	-0.19	0.10	0.050	78	-0.04	0.11	0.744
KORA F4	3,007	-0.08	0.02	$2.60*10^{-5}$	1,442	-0.11	0.03	$5.36*10^{-5}$	1,565	-0.05	0.03	0.059
LLS	1,883	-0.05	0.02	0.032	859	-0.05	0.03	0.139	1,024	-0.05	0.03	0.123
TwinsUK	1,250	-0.03	0.03	0.223	NA	NA	NA	NA	1,250	-0.03	0.03	0.223
SAPHIR	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
FTC	190	-0.01	0.07	0.870	107	-0.08	0.10	0.406	83	0.14	0.11	0.200
Meta-analysis	8,069	-0.06	0.01	$2.10*10^{-8}$	3,193	-0.09	0.02	$1.46*10^{-6}$	4,876	-0.05	0.01	$3.00*10^{-4}$

REFERENCES

1. Cawthon, R.M., Telomere measurement by quantitative PCR. Nucleic Acids Res, 2002. 30(10): p. e47.
2. Willeit, P., et al., Cellular aging reflected by leukocyte telomere length predicts advanced atherosclerosis and cardiovascular disease risk. Arterioscler Thromb Vasc Biol, 2010. 30(8): p. 1649-56.
3. Pfaffl, M.W., A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res, 2001. 29(9): p. e45.
4. Ruijter, J.M., et al., Amplification efficiency: linking baseline and bias in the analysis of quantitative PCR data. Nucleic Acids Res, 2009. 37(6): p. e45.
5. Kajantie, E., et al., No association between body size at birth and leucocyte telomere length in adult life--evidence from three cohort studies. Int J Epidemiol, 2012. 41(5): p. 1400-1408.
6. Heid, I.M., et al., Clear detection of ADIPOQ locus as the major gene for plasma adiponectin: results of genome-wide association analyses including 4659 European individuals. Atherosclerosis, 2010. 208(2): p. 412-20.
7. Rozing, M.P., et al., C-reactive protein and glucose regulation in familial longevity. Age (Dordr), 2011. 33(4): p. 623-30.

APPENDIX CHAPTER 4.5

Methods

Cohort descriptions

The **KORA** (Cooperative Health Research in the Region of Augsburg) study is a series of independent population-based epidemiological surveys and follow-up studies of participants living in the region of Augsburg, Southern Germany [1]. All survey participants are of German nationality, identified through the registration office. Informed consent has been given by all participants. The present study includes data of the KORA F4 (2006-2008) study which is a follow-up study of the KORA S4 survey (1999-2001).

The Netherlands Twin Register (**NTR**: http://www.tweelingenregister.org/) recruits twins and their family members to study the causes of individual differences in health, behavior and lifestyle. Participants are followed longitudinally; details about the cohort have been published previously [2]. A subsample of unselected twins and their family members has taken part in the NTR-Biobank [3] in which biological samples, including DNA and RNA, were collected in a standardized manner after overnight fasting. Study protocols were approved by the medical ethics board of the VUMC Amsterdam, the Netherlands.

The Estonian Genome Center, University of Tartu (**EGCUT**) is a population-based biobank of the Estonian Genome Project of University of Tartu (www.biobank.ee) [4]. The current cohort size is over 51,515, from 18 years of age and up, which reflects closely the age distribution in the adult Estonian population. The samples included in this study form a random subset of the cohort, with the exception of 500 female individuals aged 83+ which were specifically selected according to age and sex. Subjects are recruited by the general practitioners (GP) and physicians in the hospitals. Each participant filled out a Computer Assisted Personal interview, including personal data (place of birth, place(s) of living, nationality etc.), genealogical data (family history, three generations), educational and occupational history and lifestyle data (physical activity, dietary habits, smoking, alcohol consumption, women's health, quality of life). Anthropometric and physiological measurements were also taken.

The **TwinsUK** cohort (www.twinsuk.ac.uk) is an adult twin British registry shown to be representative of singleton populations and the United Kingdom population [5]. A total of 6,038 twins with RTL measurement were included in the analysis. The age range of the TwinsUK cohort was 16-99 years. Ethical approval was obtained from the Guy's and St. Thomas' Hospital Ethics Committee. Written informed consent was obtained from every participant in the study.

The Erasmus Rucphen Family (**ERF**) study is a cross-sectional cohort including 3,000 living descendants of 22 couples who had at least 6 children baptized in the community church around 1850-1900. The participants are not selected on any disease or other outcome. Details about the genealogy of the population have been described elsewhere [6, 7]. The study protocol was approved by the medical ethics board of the Erasmus MC Rotterdam, the Netherlands.

For the Leiden Longevity Study (**LLS**), long-lived siblings of Dutch descent were recruited together with their offspring and the partners of thereof. Families were included if at least two long-lived siblings were alive and fulfilled the age criterion of 89 years or older for males and 91 years or older for females, representing less than 0.5% of the Dutch population in 2001 [8]. In total, 944 long-lived proband siblings from 421 families with a mean age of 94 years (range, 89-104), 1,671 offspring (61 years, 39-81), and 744 partners (60 years, 36-79) were included in the study. DNA from the LLS was extracted from samples at baseline using conventional methods [9]. For the current analysis only the offspring and their partners were used.

The Queensland Institute of Medical Research (**QIMR**) adolescent study comprised twins and their non-twin siblings living in south-east Queensland, Australia [10]. Most (98% by self-report) are of mixed European ancestry, mainly from the British Isles. The participants are not selected on the basis of any disease or other outcome. Blood samples were collected at the end of testing sessions from participants and, if possible, from their parents. Pedigree relationships and zygosity were confirmed by genotype data. Further details are provided elsewhere [11].

Telomere length measurements

All samples from all studies were measured in the same laboratory under standard conditions. Mean leukocyte telomere length was measured by quantitative PCR-based technique as previously described [12, 13]. This method expresses telomere length as a ratio (T/S) of telomere repeat length (T) to copy number of a single copy gene, 36B4(S), within each sample. Samples were quantified relative to a calibrator sample used on each run (DNA from the K562 cell line) [13]. Mean inter-run coefficient of variations (CVs) were calculated for all study cohorts and these were less than 5% in all.

Metabolite measurements

Targeted metabolite profiling by electrospray ionization (ESI) tandem mass spectrometry (MS/MS) was performed on a fee-for-service basis on a quantitative metabolomics platform at Biocrates Life Sciences AG, Austria. The company had no access to genotype or phenotype information that would have permitted any data pre-filtering other than objective quality control for measurement errors based on internal controls and duplicates. The experimental metabolomics measurement technique is described in detail by patent US 2007/0004044 (accessible online at http://www.freepatentsonline.com/20070004044.html). A summary of the method can be found in [14, 15] and a comprehensive overview of the field and the related technologies is given in the review paper by Wenk [16]. Briefly, a targeted profiling scheme is used to quantitatively screen for known small molecule metabolites using multiple reaction monitoring, neutral loss and precursor ion scans. Quantification of

the metabolites of the biological sample is achieved by reference to appropriate internal standards.

Metabolite QC

Metabolite profile quality control was performed both for metabolites (variables) and samples (individuals).

Metabolite quality control

For each cohort, metabolite profile measurements for all included individuals were performed on multiple plates. For each metabolite *i* and plate *j*, the coefficient of variation ($CV_{i,j}$) was calculated as

$$CV_{i,j} = \frac{sd_{i,j}}{mean_{i,j}},$$

where the standard deviation (*sd*) and *mean* were calculated over all reference measurements per plate *j* (five per plate). A metabolite was excluded from further analyses if the mean CV_i over all plates exceeded 25%. Furthermore, metabolites for which the concentration data contained more than 5% missing values were excluded from further analysis.

Summary statistics for the metabolite concentration data for each cohort were compared with measurement detection limit specifications as reported by the manufacturer of the AbsoluteIDQ p150 kit (Biocrates AG, Innsbruck, Austria). For metabolites reported as absolute concentrations (μM), a metabolite was excluded from further analysis for that cohort if its median was below the lower limit of quantification (LLOQ) as reported by Biocrates AG. In the case of semiquantitatively determined metabolites (relative to a suitable internal standard), such metabolites were excluded for a given cohort if their median measurement values were below the limit of detection (LOD).

Assessment of sample quality

To assess whether individual study samples should be excluded from further analyses, we applied the following procedure. First, outlying data points per metabolite were defined as those deviating more than five SDs from the mean value for that metabolite calculated over all individuals in which this metabolite was detected. Outlying data points were set to missing. Outlying data points per metabolite were considered independent if they were located in a metabolite for which the correlations with all other metabolites containing outliers (for a given individual) were less than 70%. Samples containing more than three independent outlying data points were excluded from further analysis.

Imputation of missing values

Missing values in the remaining metabolite concentration data were imputed using a multiple imputation procedure as implemented in the 'mice' package in the statistical language and environment "R" [17].

REFERENCES

1. Wichmann, H.E., et al., KORA-gen--resource for population genetics, controls and a broad spectrum of disease phenotypes. Gesundheitswesen, 2005. 67 Suppl 1: p. S26-30.
2. Boomsma, D.I., et al., Netherlands Twin Register: from twins to twin families. Twin Res Hum Genet, 2006. 9(6): p. 849-57.
3. Willemsen, G., et al., The Netherlands Twin Register biobank: a resource for genetic epidemiological studies. Twin Res Hum Genet, 2010. 13(3): p. 231-45.
4. Nelis, M., et al., Genetic structure of Europeans: a view from the North-East. PLoS One, 2009. 4(5): p. e5472.
5. Moayyeri, A., et al., Cohort Profile: TwinsUK and Healthy Ageing Twin Study. Int J Epidemiol, 2012.
6. Aulchenko, Y.S., et al., Linkage disequilibrium in young genetically isolated Dutch population. Eur J Hum Genet, 2004. 12(7): p. 527-34.
7. Pardo, L.M., et al., The effect of genetic drift in a young genetically isolated population. Ann Hum Genet, 2005. 69(Pt 3): p. 288-95.
8. Schoenmaker, M., et al., Evidence of genetic enrichment for exceptional survival using a family approach: the Leiden Longevity Study. Eur J Hum Genet, 2006. 14(1): p. 79-84.
9. Beekman, M., et al., Chromosome 4q25, microsomal transfer protein gene, and human longevity: novel data and a meta-analysis of association studies. J Gerontol A Biol Sci Med Sci, 2006. 61(4): p. 355-62.
10. Wright, M.J. and N.G. Martin, Brisbane adolescent twin study: outline of study methods and research projects. Aust J Psychol 2004. 56(2): p. 65-78.
11. Medland, S.E., et al., Common variants in the trichohyalin gene are associated with straight hair in Europeans. Am J Hum Genet, 2009. 85(5): p. 750-5.
12. Cawthon, R.M., Telomere measurement by quantitative PCR. Nucleic Acids Res, 2002. 30(10): p. e47.
13. Codd, V., et al., Common variants near TERC are associated with mean telomere length. Nat Genet, 2010. 42(3): p. 197-9.
14. Unterwurzacher, I., et al., Rapid sample preparation and simultaneous quantitation of prostaglandins and lipoxygenase derived fatty acid metabolites by liquid chromatography-mass spectrometry from small sample volumes. Clin Chem Lab Med, 2008. 46(11): p. 1589-97.
15. Weinberger, K.M., [Metabolomics in diagnosing metabolic diseases]
16. Einsatz von Metabolomics zur Diagnose von Stoffwechselkrankheiten. Ther Umsch, 2008. 65(9): p. 487-91.
17. Wenk, M.R., The emerging field of lipidomics. Nat Rev Drug Discov, 2005. 4(7): p. 594-610.
18. R Development Core Team, R: A language and environment for statistical computing. 2010, R foundation for Statistical Computing: Vienna, Austria.

Tables

Supplement Table 1 | Exclusion reasons per cohort for each metabolite.

	KORA			NTR			EGCUT			TwinsUK			ERF			LLS			QIMR		
Mtb	**C1**	**C2**	**C3**	**C1**	**C2**	**C3**	**C1**	**C2**	**C3**	**C1**	**C2**	**C3**	**C1**	**C2**	**C3**	**C1**	**C2**	**C3**	**C1**	**C2**	**C3**
C0	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
C10	+	+	+	+	+	-	-	+	-	+	+	-	+	+	+	+	+	-	+	+	-
C10:1	+	+	+	+	+	+	-	+	-	+	+	+	+	+	+	+	+	+	+	+	-
C10:2	+	+	+	+	+	+	-	+	-	+	+	+	+	+	+	+	+	+	+	+	-
C12	+	+	-	+	+	-	-	+	-	+	+	-	+	+	-	+	+	-	+	+	-
C12-DC	+	+	-	+	+	-	-	+	-	+	+	-	+	+	-	+	+	-	+	+	-
C12:1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
C14	+	+	-	+	+	-	-	+	-	+	+	-	+	+	-	+	+	-	+	+	-
C14:1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
C14:1-OH	+	+	-	+	+	-	-	+	-	+	+	+	+	+	-	+	+	+	+	+	-
C14:2	+	+	+	+	+	+	-	+	-	+	+	+	+	+	+	+	+	+	+	+	-
C14:2-OH	+	+	-	+	+	-	-	+	-	+	+	-	+	+	-	+	+	-	+	+	-
C16	+	+	-	+	+	-	+	+	-	+	+	-	+	+	-	+	+	-	+	+	-
C16-OH	-	+	-	-	+	-	-	+	-	-	+	-	+	+	-	+	+	-	+	+	-
C16:1	+	+	-	+	+	-	-	+	-	+	+	-	+	+	-	+	+	-	+	+	-
C16:1-OH	+	+	-	-	+	-	-	+	-	+	+	-	+	+	-	+	+	-	+	+	-
C16:2	-	+	-	-	+	-	-	+	-	-	+	+	+	+	+	+	+	-	+	+	-
C16:2-OH	+	+	-	-	+	-	-	+	-	+	+	-	-	+	-	+	+	-	+	+	-
C18	+	+	-	+	+	-	-	+	-	+	+	-	+	+	-	+	+	-	+	+	-
C18:1	+	+	+	+	+	+	+	+	-	+	+	+	+	+	+	+	+	+	+	+	+
C18:1-OH	-	+	-	-	+	-	-	+	-	-	+	-	+	+	-	+	+	-	+	+	-
C18:2	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
C2	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
C3	+	+	-	+	+	-	+	+	-	+	+	-	+	+	+	+	+	+	+	+	-
C3-DC (C4-OH)	+	+	-	+	+	-	-	+	-	+	+	-	+	+	-	+	+	-	+	+	-
C3-OH	-	+	-	-	+	-	-	+	-	-	+	-	+	+	+	+	+	+	+	+	-
C3:1	-	+	-	-	+	-	-	+	-	-	+	-	+	+	-	+	+	-	+	+	-
C4	+	+	-	+	+	-	+	+	-	+	+	-	+	+	-	+	+	-	+	+	-
C4:1	-	+	-	+	+	-	-	+	-	-	+	-	+	+	-	+	+	-	+	+	-
C5	+	+	-	+	+	-	+	+	-	+	+	-	+	+	-	+	+	-	+	+	-
C5-DC (C6-OH)	+	+	-	+	+	-	-	+	-	+	+	-	+	+	-	+	+	-	+	+	-
C5-M-DC	+	+	-	+	+	-	-	+	-	+	+	-	+	+	-	+	+	-	+	+	-
C5-OH (C3-DC-M)	+	+	-	+	+	-	-	+	-	+	+	-	+	+	-	+	+	-	+	+	-
C5:1	+	+	-	+	+	-	-	+	-	+	+	-	+	+	-	+	+	-	+	+	-

7

	KORA			NTR			EGCUT			TwinsUK			ERF			LLS			QIMR		
Mtb	**C1**	**C2**	**C3**	**C1**	**C2**	**C3**	**C1**	**C2**	**C3**	**C1**	**C2**	**C3**	**C1**	**C2**	**C3**	**C1**	**C2**	**C3**	**C1**	**C2**	**C3**
C5:1-DC	+	+	-	+	+	+	-	+	-	+	+	-	+	+	+	+	+	-	+	+	-
C6 (C4:1-DC)	+	+	-	+	+	-	-	+	-	+	+	-	+	+	-	+	+	-	+	+	-
C6:1	+	+	-	+	+	-	-	+	-	+	+	-	+	+	-	+	+	-	+	+	-
C7-DC	+	+	+	+	+	+	-	+	-	+	+	+	+	+	+	+	+	+	+	+	-
C8	+	+	+	+	+	-	-	+	-	+	+	-	+	+	+	+	+	-	+	+	-
C8:1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
C9	+	+	+	+	+	-	-	+	-	+	+	+	+	+	+	+	+	+	+	+	-
Arg	+	+	+	+	+	+	+	+	-	+	+	+	+	+	+	+	+	+	+	+	-
Gln	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Gly	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
His	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Met	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Orn	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Phe	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Pro	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Ser	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Thr	+	+	+	+	+	+	+	+	-	+	+	+	-	+	+	+	+	+	+	+	+
Trp	+	+	+	+	+	+	+	+	-	+	+	+	-	+	+	+	+	+	+	+	+
Tyr	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	+	+	+	+	+
Val	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	+	+	+	+	+
xLeu	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	+	+	+	+	+
PC aa C24:0	+	+	+	-	+	+	-	+	-	+	+	+	+	+	-	+	+	-	+	+	+
PC aa C26:0	+	+	-	+	+	-	-	+	-	+	+	-	+	+	-	+	+	-	+	+	-
PC aa C28:1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PC aa C30:0	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PC aa C30:2	-	+	-	-	-	-	-	-	+	-	+	+	-	-	-	-	+	-	+	-	+
PC aa C32:0	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PC aa C32:1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PC aa C32:2	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PC aa C32:3	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PC aa C34:1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PC aa C34:2	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PC aa C34:3	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PC aa C34:4	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PC aa C36:0	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PC aa C36:1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PC aa C36:2	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PC aa C36:3	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

	KORA			NTR			EGCUT			TwinsUK			ERF			LLS			QIMR		
Mtb	**C1**	**C2**	**C3**	**C1**	**C2**	**C3**	**C1**	**C2**	**C3**	**C1**	**C2**	**C3**	**C1**	**C2**	**C3**	**C1**	**C2**	**C3**	**C1**	**C2**	**C3**
PC aa C36:4	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PC aa C36:5	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PC aa C36:6	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PC aa C38:0	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PC aa C38:1	-	+	-	-	+	+	-	+	+	-	+	+	-	+	+	+	+	+	+	+	+
PC aa C38:3	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PC aa C38:4	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PC aa C38:5	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PC aa C38:6	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PC aa C40:1	+	+	+	+	+	+	+	+	-	+	+	+	+	+	-	+	+	+	+	+	-
PC aa C40:2	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PC aa C40:3	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PC aa C40:4	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PC aa C40:5	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PC aa C40:6	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PC aa C42:0	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PC aa C42:1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PC aa C42:2	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PC aa C42:4	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PC aa C42:5	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PC aa C42:6	+	+	+	+	+	+	+	+	-	+	+	+	+	+	+	+	+	+	+	+	+
PC ae C30:0	+	+	+	+	+	+	+	+	-	+	+	+	+	+	+	+	+	+	+	+	+
PC ae C30:1	-	+	-	-	+	+	-	-	+	-	+	+	-	-	+	-	+	-	+	-	+
PC ae C30:2	+	+	-	+	+	-	-	+	-	+	+	-	+	+	-	+	+	-	+	+	-
PC ae C32:1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PC ae C32:2	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PC ae C34:0	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PC ae C34:1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PC ae C34:2	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PC ae C34:3	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PC ae C36:0	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PC ae C36:1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PC ae C36:2	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PC ae C36:3	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PC ae C36:4	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PC ae C36:5	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PC ae C38:0	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PC ae C38:1	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	+	+	+	+	+

7

	KORA			NTR			EGCUT			TwinsUK			ERF			LLS			QIMR		
Mtb	**C1**	**C2**	**C3**	**C1**	**C2**	**C3**	**C1**	**C2**	**C3**	**C1**	**C2**	**C3**	**C1**	**C2**	**C3**	**C1**	**C2**	**C3**	**C1**	**C2**	**C3**
PC ae C38:2	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PC ae C38:3	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PC ae C38:4	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PC ae C38:5	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PC ae C38:6	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PC ae C40:0	+	+	-	+	+	-	-	+	-	+	+	-	+	+	-	+	+	-	+	+	-
PC ae C40:1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PC ae C40:2	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PC ae C40:3	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PC ae C40:4	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PC ae C40:5	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PC ae C40:6	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PC ae C42:0	+	+	+	+	+	+	+	+	-	+	+	+	+	+	+	+	+	+	+	+	+
PC ae C42:1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PC ae C42:2	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PC ae C42:3	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PC ae C42:4	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PC ae C42:5	+	+	+	+	+	+	+	+	-	+	+	+	+	+	+	+	+	+	+	+	+
PC ae C44:3	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PC ae C44:4	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PC ae C44:5	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PC ae C44:6	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
lysoPC a C14:0	+	+	-	+	+	-	-	+	-	+	+	-	+	+	-	+	+	-	+	+	-
lysoPC a C16:0	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
lysoPC a C16:1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
lysoPC a C17:0	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
lysoPC a C18:0	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
lysoPC a C18:1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
lysoPC a C18:2	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
lysoPC a C20:3	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
lysoPC a C20:4	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
lysoPC a C24:0	+	+	-	+	+	-	-	+	-	+	+	-	+	+	-	+	+	-	+	+	-
lysoPC a C26:0	-	+	-	+	+	+	-	+	-	-	+	-	+	+	-	+	+	-	+	+	-
lysoPC a C26:1	+	+	-	+	+	-	-	+	-	+	+	-	+	+	-	+	+	-	+	+	-
lysoPC a C28:0	+	+	+	+	+	+	+	+	-	+	+	+	+	+	+	+	+	+	+	+	+
lysoPC a C28:1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
lysoPC a C6:0	-	+	-	-	+	-	-	-	-	-	+	-	-	-	-	-	+	-	+	-	-
SM (OH) C14:1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

	KORA			NTR			EGCUT			TwinsUK			ERF			LLS			QIMR		
Mtb	**C1**	**C2**	**C3**	**C1**	**C2**	**C3**	**C1**	**C2**	**C3**	**C1**	**C2**	**C3**	**C1**	**C2**	**C3**	**C1**	**C2**	**C3**	**C1**	**C2**	**C3**
SM (OH) C16:1	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	+	+	+	+	+
SM (OH) C22:1	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	+	+	+	+	+
SM (OH) C22:2	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	+	+	+	+	+
SM (OH) C24:1	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	+	+	+	+	+
SM C16:0	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
SM C16:1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
SM C18:0	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
SM C18:1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
SM C20:2	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
SM C22:3	-	+	-	-	-	-	-	-	-	-	+	+	-	-	-	-	+	-	+	-	-
SM C24:0	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
SM C24:1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
SM C26:0	+	+	+	+	+	+	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+
SM C26:1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
H1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

C1: Median(Concentration) > LOD, where LOD is limit of detection specified by Biocrates
C2: Excluded based on local QC of the metabolomics data.
C3: f(0) < 5%, where f(0)=N(0)/Samplesize. For the Biocrates data set a value of exactly 0 is considered a missing value.
'+': passed QC for this criterion; '-': failed QC for this criterion

Supplement Table 2 | Complete results of meta-analysis of partial correlation analysis of telomere length and metabolites.

	Model 1: age + sex				Model 2: age + sex + BMI				
metabolite	**n**	**direction**	**r**	**p-value**	**n**	**direction**	**r**	**p-value**	**metabolite full name**
C0	8192	-+---++	-0.01	0.502	7977	++---+	-0.01	0.648	Carnitine
C10	3809	+???+??	0.01	0.421	3794	+???+?	0.01	0.653	Decanoylcarnitine
C10:1	6915	-+?-++?	0.00	0.933	6893	-+?-++	0.00	0.987	Decenoylcarnitine
C10:2	6915	-+?---?	-0.02	0.066	6893	-+?---	-0.02	0.098	Decadienylcarnitine
C12:1	8192	-++-+--	-0.01	0.300	7977	-++-+-	-0.01	0.365	Dodecenoylcarnitine
C14:1	8192	++--+-+	0.00	0.775	7977	+0--+-	-0.01	0.461	Tetradecenoylcarnitine
C14:1-OH	1792	???-?+?	-0.04	0.121	1792	???-?-	-0.04	0.121	Hydroxytetradecenoylcarnitine
C14:2	6915	-+?-+-?	0.00	0.900	6893	-+?-+-	-0.01	0.657	Tetradecadienylcarnitine
C16:2	806	????-??	-0.01	0.836	806	????-?	-0.01	0.875	Hexadecadienylcarnitine
C18:1	7108	-+?-+-+	-0.01	0.518	6893	-+?-+-	-0.01	0.348	Octadecenoylcarnitine
C18:2	8192	-++--++	0.01	0.626	7977	-++--+	0.00	0.933	Octadecadienylcarnitine
C2	8192	--+---+	-0.03	0.014	7977	--+---	-0.03	0.015	Acetylcarnitine
C3	1449	????++?	0.01	0.780	1449	????++	0.01	0.758	Propionylcarnitine
C3-OH	1449	????--?	-0.10	$2.64*10^{-4}$	1449	????--	-0.10	$2.75*10^{-4}$	Hydroxypropionylcarnitine
C5:1-DC	2120	?+??-??	0.01	0.559	2113	?+??-?	0.01	0.516	Glutaconylcarnitine
C7-DC	6915	-+?---?	-0.02	0.055	6893	-+?---	-0.03	0.017	Pimelylcarnitine
C8	3809	+???+??	0.01	0.459	3794	+???+?	0.01	0.625	Octanoylcarnitine
C8:1	8192	-++--+-	-0.01	0.440	7977	+++--+	0.00	0.916	Octenoylcarnitine
C9	5601	+??-++?	0.04	0.003	5586	+??-++	0.04	0.011	Nonaylcarnitine
Arg	6915	+-?++-?	0.01	0.546	6893	+-?++-	0.01	0.602	Arginine
Gln	8192	+---+++	0.01	0.488	7977	+-+-++	0.00	0.847	Glutamine
Gly	8192	++-+-++	0.01	0.261	7977	+0-+-+	0.00	0.828	Glycine

metabolite	Model 1: age + sex				Model 2: age + sex + BMI				metabolite full name
	n	direction	r	p-value	n	direction	r	p-value	
His	8192	+--+-++	0.01	0.562	7977	+--+-+	0.00	0.834	Histidine
Met	8192	------+	-0.04	$9.20*10^{-5}$	7977	------	-0.05	$7.51*10^{-5}$	Methionine
Orn	8192	+-+--++	-0.01	0.611	7977	+-+--+	-0.01	0.527	Ornithine
Phe	8192	--+-+-+	-0.01	0.363	7977	--+-+-	-0.01	0.597	Phenylalanine
Pro	8192	------+	-0.02	0.183	7977	------	-0.01	0.230	Proline
Ser	8192	++++-++	0.02	0.081	7977	-+++-+	0.01	0.202	Serine
Thr	6302	-+?+?++	0.02	0.241	6087	-+?+?+	0.01	0.340	Threonine
Trp	6302	+-?-?++	-0.01	0.619	6087	+-?-?+	-0.01	0.500	Tryptophan
Tyr	7386	----?-+	-0.04	$2.14*10^{-4}$	7171	----?-	-0.04	0.001	Tyrosine
Val	7386	----?++	-0.01	0.541	7171	+---?+	0.00	0.977	Valine
xLeu	7386	----?++	-0.01	0.517	7171	+---?+	0.00	0.952	Leucine / Isoleucine
PC aa C24:0	4345	-??-??+	-0.02	0.290	4137	-??-??	-0.03	0.054	Phosphatidylcholine diacyl C24:0
PC aa C28:1	8192	----+++	-0.01	0.555	7977	----++	-0.01	0.282	Phosphatidylcholine diacyl C28:1
PC aa C30:0	8192	----+++	-0.02	0.070	7977	----++	-0.03	0.021	Phosphatidylcholine diacyl C30:0
PC aa C32:0	8192	----+++	-0.03	0.021	7977	----++	-0.03	0.006	Phosphatidylcholine diacyl C32:0
PC aa C32:1	8192	----+-+	-0.04	$2.44*10^{-4}$	7977	----++	-0.04	$3.38*10^{-4}$	Phosphatidylcholine diacyl C32:1
PC aa C32:2	8192	----+++	-0.01	0.352	7977	----++	-0.02	0.188	Phosphatidylcholine diacyl C32:2
PC aa C32:3	8192	+---+++	0.01	0.495	7977	+---++	0.01	0.623	Phosphatidylcholine diacyl C32:3
PC aa C34:1	8192	----+++	-0.02	0.031	7977	----++	-0.03	0.013	Phosphatidylcholine diacyl C34:1
PC aa C34:2	8192	-----++	-0.02	0.090	7977	-----+	-0.02	0.038	Phosphatidylcholine diacyl C34:2
PC aa C34:3	8192	----+++	-0.02	0.065	7977	----++	-0.02	0.033	Phosphatidylcholine diacyl C34:3
PC aa C34:4	8192	----+++	0.00	0.753	7977	----++	-0.01	0.625	Phosphatidylcholine diacyl C34:4
PC aa C36:0	8192	+---+++	0.01	0.424	7977	----++	0.00	0.977	Phosphatidylcholine diacyl C36:0

7

metabolite	Model 1: age + sex				Model 2: age + sex + BMI				metabolite full name
	n	direction	r	p-value	n	direction	r	p-value	
PC aa C36:1	8192	----+++	-0.03	0.009	7977	----++	-0.03	0.004	Phosphatidylcholine diacyl C36:1
PC aa C36:2	8192	-----++	-0.02	0.030	7977	-----+	-0.03	0.010	Phosphatidylcholine diacyl C36:2
PC aa C36:3	8192	----+++	-0.02	0.063	7977	----++	-0.02	0.053	Phosphatidylcholine diacyl C36:3
PC aa C36:4	8192	----+++	0.00	0.913	7977	----++	0.00	0.840	Phosphatidylcholine diacyl C36:4
PC aa C36:5	8192	----+++	-0.01	0.508	7977	----++	-0.01	0.229	Phosphatidylcholine diacyl C36:5
PC aa C36:6	8192	+---+++	0.00	0.715	7977	----++	0.00	0.823	Phosphatidylcholine diacyl C36:6
PC aa C38:0	8192	+-+-+++	0.01	0.355	7977	--+-++	0.00	0.735	Phosphatidylcholine diacyl C38:0
PC aa C38:1	836	?????++	0.08	0.026	643	?????+	0.05	0.173	Phosphatidylcholine diacyl C38:1
PC aa C38:3	8192	----+++	-0.03	0.007	7977	----++	-0.02	0.035	Phosphatidylcholine diacyl C38:3
PC aa C38:4	8192	----+++	-0.01	0.514	7977	----++	-0.01	0.646	Phosphatidylcholine diacyl C38:4
PC aa C38:5	8192	+---+++	-0.01	0.652	7977	----++	-0.01	0.342	Phosphatidylcholine diacyl C38:5
PC aa C38:6	8192	+---+-+	0.00	0.987	7977	+---+-	-0.01	0.611	Phosphatidylcholine diacyl C38:6
PC aa C40:1	6109	+0?+?-?	0.01	0.453	6087	++?+?-	0.00	0.740	Phosphatidylcholine diacyl C40:1
PC aa C40:2	8192	-++++++	0.01	0.321	7977	-+++++	0.00	0.758	Phosphatidylcholine diacyl C40:2
PC aa C40:3	8192	-++---+	0.00	0.852	7977	-++---	-0.01	0.435	Phosphatidylcholine diacyl C40:3
PC aa C40:4	8192	----+++	-0.01	0.225	7977	----++	-0.01	0.255	Phosphatidylcholine diacyl C40:4
PC aa C40:5	8192	----+++	-0.03	0.022	7977	----++	-0.03	0.017	Phosphatidylcholine diacyl C40:5
PC aa C40:6	8192	----+-+	-0.02	0.073	7977	----+-	-0.02	0.070	Phosphatidylcholine diacyl C40:6
PC aa C42:0	8192	+++++++	0.03	0.020	7977	+++-++	0.02	0.089	Phosphatidylcholine diacyl C42:0
PC aa C42:1	8192	+-+-+++	0.02	0.086	7977	+-+-++	0.01	0.389	Phosphatidylcholine diacyl C42:1
PC aa C42:2	8192	+++++++	0.01	0.236	7977	-0++++	0.00	0.907	Phosphatidylcholine diacyl C42:2
PC aa C42:4	8192	--+-+++	0.00	0.854	7977	--+-++	-0.01	0.599	Phosphatidylcholine diacyl C42:4
PC aa C42:5	8192	--+-+-+	-0.02	0.152	7977	--+-+-	-0.02	0.078	Phosphatidylcholine diacyl C42:5

metabolite	Model 1: age + sex				Model 2: age + sex + BMI				metabolite full name
	n	direction	r	p-value	n	direction	r	p-value	
PC aa C42:6	7108	--?-+-+	0.00	0.999	6893	--?-+-	-0.01	0.478	Phosphatidylcholine diacyl C42:6
PC ae C30:0	7108	+-?-+++	0.01	0.220	6893	--?-++	0.00	0.732	Phosphatidylcholine acyl-alkyl C30:0
PC ae C32:1	8192	----+++	0.00	0.756	7977	----++	-0.01	0.244	Phosphatidylcholine acyl-alkyl C32:1
PC ae C32:2	8192	--+-+++	0.00	0.771	7977	--+-++	-0.01	0.306	Phosphatidylcholine acyl-alkyl C32:2
PC ae C34:0	8192	+---+++	0.00	0.755	7977	----++	-0.01	0.643	Phosphatidylcholine acyl-alkyl C34:0
PC ae C34:1	8192	+---+++	0.00	0.835	7977	----++	-0.01	0.579	Phosphatidylcholine acyl-alkyl C34:1
PC ae C34:2	8192	+---+++	0.01	0.204	7977	----++	0.00	0.808	Phosphatidylcholine acyl-alkyl C34:2
PC ae C34:3	8192	--+++++	0.01	0.538	7977	--+-++	-0.01	0.590	Phosphatidylcholine acyl-alkyl C34:3
PC ae C36:0	8192	+---+++	0.00	0.982	7977	----++	-0.01	0.460	Phosphatidylcholine acyl-alkyl C36:0
PC ae C36:1	8192	+-+-+++	0.03	0.007	7977	+-+-++	0.02	0.058	Phosphatidylcholine acyl-alkyl C36:1
PC ae C36:2	8192	++--+++	0.03	0.014	7977	+---++	0.02	0.132	Phosphatidylcholine acyl-alkyl C36:2
PC ae C36:3	8192	+---+++	0.01	0.301	7977	----++	0.00	0.789	Phosphatidylcholine acyl-alkyl C36:3
PC ae C36:4	8192	+---+++	0.01	0.240	7977	+---++	0.01	0.401	Phosphatidylcholine acyl-alkyl C36:4
PC ae C36:5	8192	----+++	0.00	0.920	7977	----++	-0.01	0.635	Phosphatidylcholine acyl-alkyl C36:5
PC ae C38:0	8192	++--+++	0.01	0.567	7977	++--++	0.00	0.828	Phosphatidylcholine acyl-alkyl C38:0
PC ae C38:1	7193	+-++?-?	0.01	0.250	7171	+-++?-	0.01	0.482	Phosphatidylcholine acyl-alkyl C38:1
PC ae C38:2	8192	+++++++	0.02	0.099	7977	-++-++	0.01	0.402	Phosphatidylcholine acyl-alkyl C38:2
PC ae C38:3	8192	+++-+++	0.03	0.017	7977	+++-++	0.02	0.036	Phosphatidylcholine acyl-alkyl C38:3
PC ae C38:4	8192	+-+-+++	0.04	0.001	7977	+-+-++	0.03	0.005	Phosphatidylcholine acyl-alkyl C38:4
PC ae C38:5	8192	+-+-+++	0.01	0.251	7977	+-+-++	0.01	0.510	Phosphatidylcholine acyl-alkyl C38:5
PC ae C38:6	8192	+---+++	0.01	0.543	7977	+---++	0.00	0.967	Phosphatidylcholine acyl-alkyl C38:6
PC ae C40:1	8192	+-++-++	0.02	0.073	7977	+-++-+	0.01	0.402	Phosphatidylcholine acyl-alkyl C40:1
PC ae C40:2	8192	+-+++++	0.02	0.069	7977	+-++++	0.02	0.147	Phosphatidylcholine acyl-alkyl C40:2

7

metabolite	Model 1: age + sex				Model 2: age + sex + BMI				metabolite full name
	n	direction	r	p-value	n	direction	r	p-value	
PC ae C40:3	8192	+++++++	0.04	0.002	7977	++++++	0.03	0.009	Phosphatidylcholine acyl-alkyl C40:3
PC ae C40:4	8192	+0+++++	0.03	0.004	7977	+-++++	0.03	0.028	Phosphatidylcholine acyl-alkyl C40:4
PC ae C40:5	8192	+-+++++	0.04	0.002	7977	+-++++	0.03	0.019	Phosphatidylcholine acyl-alkyl C40:5
PC ae C40:6	8192	+-+-+++	0.03	0.009	7977	+-+-++	0.02	0.085	Phosphatidylcholine acyl-alkyl C40:6
PC ae C42:0	7108	+0?-+-+	0.00	0.841	6893	-+?-+-	0.00	0.805	Phosphatidylcholine acyl-alkyl C42:0
PC ae C42:1	8192	--+-+++	0.01	0.246	7977	-0+-++	0.01	0.610	Phosphatidylcholine acyl-alkyl C42:1
PC ae C42:2	8192	+-+--++	0.01	0.285	7977	--+--+	0.00	0.906	Phosphatidylcholine acyl-alkyl C42:2
PC ae C42:3	8192	+++++++	0.02	0.052	7977	++++++	0.01	0.414	Phosphatidylcholine acyl-alkyl C42:3
PC ae C42:4	8192	+-+++++	0.02	0.035	7977	+-++++	0.01	0.209	Phosphatidylcholine acyl-alkyl C42:4
PC ae C42:5	7108	+-?++++	0.02	0.084	6893	+-?+++	0.01	0.290	Phosphatidylcholine acyl-alkyl C42:5
PC ae C44:3	8192	--+++++	0.02	0.111	7977	-0++++	0.01	0.518	Phosphatidylcholine acyl-alkyl C44:3
PC ae C44:4	8192	+-+--++	0.01	0.418	7977	+-+---	0.00	0.994	Phosphatidylcholine acyl-alkyl C44:4
PC ae C44:5	8192	+-+--++	0.01	0.212	7977	+-+--+	0.01	0.527	Phosphatidylcholine acyl-alkyl C44:5
PC ae C44:6	8192	++++-++	0.02	0.099	7977	+++--+	0.01	0.353	Phosphatidylcholine acyl-alkyl C44:6
lysoPC a C16:0	8192	++-++++	0.02	0.120	7977	++-+++	0.01	0.417	lysoPhosphatidylcholine acyl C16:0
lysoPC a C16:1	8192	-+--+-+	-0.01	0.368	7977	-+--+-	-0.01	0.207	lysoPhosphatidylcholine acyl C16:1
lysoPC a C17:0	8192	++-++++	0.05	$7.10*10^{-6}$	7977	++-+++	0.04	$4.72*10^{-4}$	lysoPhosphatidylcholine acyl C17:0
lysoPC a C18:0	8192	++--+++	0.02	0.096	7977	+0--++	0.01	0.332	lysoPhosphatidylcholine acyl C18:0
lysoPC a C18:1	8192	++---++	0.01	0.264	7977	-0---+	0.00	0.771	lysoPhosphatidylcholine acyl C18:1
lysoPC a C18:2	8192	++++-++	0.02	0.154	7977	-+++-+	0.00	0.872	lysoPhosphatidylcholine acyl C18:2
lysoPC a C20:3	8192	-+---++	0.00	0.956	7977	-+---+	-0.01	0.620	lysoPhosphatidylcholine acyl C20:3
lysoPC a C20:4	8192	++++-++	0.02	0.065	7977	++++-+	0.01	0.314	lysoPhosphatidylcholine acyl C20:4
lysoPC a C26:0	1314	?+?????	0.02	0.565	1307	?+????	0.02	0.457	lysoPhosphatidylcholine acyl C26:0

metabolite	Model 1: age + sex				Model 2: age + sex + BMI				metabolite full name
	n	direction	r	p-value	n	direction	r	p-value	
lysoPC a C28:0	7108	-0?-+-+	-0.01	0.447	6893	-+?-+-	-0.02	0.175	lysoPhosphatidylcholine acyl C28:0
lysoPC a C28:1	8192	+++-+++	0.02	0.096	7977	-++-++	0.01	0.279	lysoPhosphatidylcholine acyl C28:1
SM (OH) C14:1	8192	+---+++	0.02	0.079	7977	+---++	0.01	0.331	Hydroxysphingomyeline C14:1
SM (OH) C16:1	7386	+-+-?++	0.03	0.011	7171	+-+-?+	0.02	0.053	Hydroxysphingomyeline C16:1
SM (OH) C22:1	7386	+0++?++	0.02	0.129	7171	+-++?+	0.01	0.310	Hydroxysphingomyeline C22:1
SM (OH) C22:2	7386	+-+-?++	0.02	0.051	7171	+-+-?+	0.02	0.213	Hydroxysphingomyeline C22:2
SM (OH) C24:1	7386	+0+-?++	0.01	0.209	7171	+-+-?+	0.01	0.457	Hydroxysphingomyeline C24:1
SM C16:0	8192	--+-+++	0.00	0.864	7977	--+-++	-0.01	0.330	Sphingomyeline C16:0
SM C16:1	8192	--+-+++	-0.02	0.066	7977	--+-++	-0.02	0.074	Sphingomyeline C16:1
SM C18:0	8192	--+-+++	0.00	0.933	7977	+-+-++	0.00	0.864	Sphingomyeline C18:0
SM C18:1	8192	--+-+++	0.00	0.822	7977	+-+-++	0.01	0.598	Sphingomyeline C18:1
SM C20:2	8192	+-+-+++	0.03	0.002	7977	+-+-++	0.03	0.003	Sphingomyeline C20:2
SM C24:0	8192	--+++++	0.00	0.906	7977	--+-++	-0.01	0.657	Sphingomyeline C24:0
SM C24:1	8192	--+--++	-0.01	0.402	7977	--+--+	-0.01	0.236	Sphingomyeline C24:1
SM C26:0	5794	+??-+++	0.03	0.043	5586	+??-++	0.02	0.229	Sphingomyeline C26:0
SM C26:1	6878	+?+++++	0.01	0.241	6670	-?+-++	0.01	0.604	Sphingomyeline C26:1
H1	8192	----+--	-0.02	0.073	7977	----++	-0.01	0.363	Hexose

*Direction: KORA, NTR, EGCUT, TwinsUK, ERF, LLS, QIMR

7

Supplement Table 3 | Study specific results for age and sex adjusted model.

	KORA			NTR			EGCUT			TwinsUK			ERF			LLS			QIMR		
metabolite	**n**	**r**	**p-value**	**n**	**r**	**p-value**	**n**	**r**	**p-value**	**n**	**r**	**p-value**	**n**	**r**	**p-value**	**n**	**r**	**p-value**	**n**	**r**	**p-value**
lysoPC a C17:0	3003	0.08	$8.19*10^{-6}$	1314	0.07	0.016	1084	-0.02	0.455	1149	0.01	0.849	806	0.02	0.597	643	0.08	0.032	193	0.09	0.234
Met	3003	-0.05	0.006	1314	-0.04	0.157	1084	-0.01	0.708	1149	-0.01	0.865	806	-0.13	$1.24*10^{-4}$	643	-0.04	0.300	193	0.04	0.537
Tyr	3003	-0.05	0.005	1314	-0.06	0.023	1084	-0.02	0.606	1149	-0.05	0.117	NA	NA	NA	643	-0.04	0.286	193	0.05	0.521
PC aa C32:1	3003	-0.06	0.001	1314	-0.05	0.087	1084	-0.04	0.163	1149	-0.06	0.030	806	0.00	0.925	643	0.00	0.989	193	0.05	0.494
C3-OH	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	806	-0.15	$3.11*10^{-5}$	643	-0.03	0.420	NA	NA	NA
PC ae C38:4	3003	0.05	0.003	1314	-0.01	0.723	1084	0.00	0.920	1149	-0.01	0.699	806	0.06	0.112	643	0.13	0.001	193	0.11	0.141
PC ae C40:3	3003	0.04	0.041	1314	0.00	0.918	1084	0.03	0.280	1149	0.01	0.851	806	0.07	0.048	643	0.08	0.049	193	0.09	0.193
PC ae C40:5	3003	0.05	0.009	1314	-0.01	0.798	1084	0.01	0.648	1149	0.02	0.578	806	0.05	0.170	643	0.09	0.025	193	0.09	0.227
SM C20:2	3003	0.02	0.190	1314	-0.04	0.128	1084	0.07	0.029	1149	-0.02	0.609	806	0.21	$1.04*10^{-9}$	643	0.00	0.937	193	0.09	0.224
C9	3003	0.06	0.003	NA	NA	NA	NA	NA	NA	1149	-0.03	0.290	806	0.03	0.346	643	0.08	0.052	NA	NA	NA
PC ae C40:4	3003	0.04	0.021	1314	0.00	1.000	1084	0.01	0.654	1149	0.01	0.803	806	0.03	0.452	643	0.11	0.006	193	0.09	0.237
PC aa C38:3	3003	-0.04	0.024	1314	-0.05	0.071	1084	-0.02	0.493	1149	-0.09	0.003	806	0.02	0.518	643	0.03	0.483	193	0.04	0.553
PC ae C36:1	3003	0.03	0.071	1314	-0.01	0.798	1084	0.00	0.894	1149	-0.01	0.667	806	0.06	0.077	643	0.12	0.003	193	0.15	0.035
PC aa C36:1	3003	-0.05	0.013	1314	-0.07	0.016	1084	-0.03	0.283	1149	-0.06	0.045	806	0.03	0.359	643	0.05	0.199	193	0.09	0.210
PC ae C40:6	3003	0.05	0.005	1314	-0.01	0.819	1084	0.01	0.832	1149	-0.01	0.740	806	0.05	0.176	643	0.04	0.281	193	0.10	0.147
SM (OH) C16:1	3003	0.05	0.010	1314	-0.02	0.507	1084	0.02	0.571	1149	-0.01	0.835	NA	NA	NA	643	0.09	0.021	193	0.12	0.102
C2	3003	-0.03	0.143	1314	-0.06	0.020	1084	0.03	0.378	1149	-0.08	0.005	806	-0.01	0.747	643	-0.02	0.593	193	0.04	0.603
PC ae C36:2	3003	0.03	0.062	1314	0.00	0.913	1084	-0.01	0.750	1149	-0.01	0.753	806	0.06	0.104	643	0.09	0.017	193	0.12	0.088
PC ae C38:3	3003	0.03	0.105	1314	0.00	0.892	1084	0.01	0.702	1149	-0.03	0.397	806	0.06	0.092	643	0.09	0.020	193	0.10	0.165
PC aa C42:0	3003	0.03	0.155	1314	0.01	0.692	1084	0.06	0.037	1149	0.00	0.962	806	0.02	0.662	643	0.04	0.364	193	0.05	0.515
PC aa C32:0	3003	-0.05	0.005	1314	-0.06	0.036	1084	-0.03	0.360	1149	-0.03	0.369	806	0.04	0.222	643	0.05	0.234	193	0.07	0.324
PC aa C40:5	3003	-0.03	0.074	1314	-0.06	0.039	1084	-0.03	0.396	1149	-0.05	0.076	806	0.02	0.505	643	0.01	0.836	193	0.08	0.262

	KORA			NTR			EGCUT			TwinsUK			ERF			LLS			QIMR		
metabolite	n	r	p-value	n	r	p-value	n	r	p-value	n	r	p-value	n	r	p-value	n	r	p-value	n	r	p-value
PC aa C38:1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	643	0.06	0.134	193	0.14	0.057
PC aa C36:2	3003	-0.04	0.051	1314	-0.07	0.014	1084	-0.02	0.563	1149	-0.03	0.260	806	-0.01	0.854	643	0.05	0.170	193	0.10	0.154
PC aa C34:1	3003	-0.03	0.066	1314	-0.05	0.056	1084	-0.05	0.131	1149	-0.04	0.131	806	0.00	0.954	643	0.07	0.091	193	0.11	0.140
PC ae C42:4	3003	0.04	0.049	1314	-0.01	0.654	1084	0.04	0.169	1149	0.01	0.787	806	0.00	0.895	643	0.04	0.292	193	0.07	0.366
SM C26:0	3003	0.03	0.075	NA	NA	NA	NA	NA	NA	1149	-0.03	0.341	806	0.05	0.170	643	0.01	0.744	193	0.14	0.048
SM (OH) C22:2	3003	0.03	0.144	1314	-0.01	0.786	1084	0.01	0.754	1149	0.00	0.936	NA	NA	NA	643	0.09	0.016	193	0.13	0.078
PC ae C42:3	3003	0.02	0.201	1314	0.01	0.699	1084	0.03	0.337	1149	0.01	0.678	806	0.01	0.837	643	0.03	0.505	193	0.12	0.085
C7-DC	3003	-0.02	0.368	1314	0.03	0.216	NA	NA	NA	1149	-0.09	0.001	806	-0.06	0.074	643	-0.04	0.331	NA	NA	NA
PC aa C36:3	3003	-0.04	0.050	1314	-0.04	0.152	1084	-0.03	0.300	1149	-0.05	0.069	806	0.00	0.890	643	0.08	0.048	193	0.09	0.193
lysoPC a C20:4	3003	0.02	0.180	1314	0.02	0.552	1084	0.01	0.834	1149	0.01	0.700	806	-0.03	0.469	643	0.07	0.066	193	0.14	0.055
PC aa C34:3	3003	-0.03	0.076	1314	-0.03	0.223	1084	-0.03	0.274	1149	-0.03	0.277	806	0.02	0.623	643	0.02	0.578	193	0.06	0.392
SM C16:1	3003	-0.04	0.015	1314	-0.07	0.013	1084	0.01	0.765	1149	-0.04	0.169	806	0.03	0.375	643	0.07	0.088	193	0.08	0.250
C10:2	3003	-0.03	0.115	1314	0.04	0.127	NA	NA	NA	1149	-0.09	0.001	806	-0.01	0.872	643	-0.06	0.144	NA	NA	NA
PC ae C40:2	3003	0.03	0.091	1314	-0.01	0.772	1084	0.01	0.699	1149	0.00	0.934	806	0.02	0.655	643	0.06	0.148	193	0.08	0.281
PC aa C30:0	3003	-0.04	0.018	1314	-0.06	0.034	1084	-0.04	0.177	1149	-0.01	0.664	806	0.06	0.075	643	0.06	0.158	193	0.08	0.288
H1	3003	-0.02	0.173	1314	-0.03	0.263	1084	-0.01	0.654	1149	-0.04	0.193	806	0.02	0.555	643	0.00	1.000	193	-0.08	0.287
PC aa C40:6	3003	-0.02	0.308	1314	-0.04	0.109	1084	-0.03	0.343	1149	-0.05	0.109	806	0.05	0.164	643	-0.05	0.248	193	0.08	0.294
PC ae C40:1	3003	0.02	0.279	1314	0.00	0.904	1084	0.02	0.601	1149	0.02	0.576	806	-0.01	0.850	643	0.07	0.067	193	0.17	0.020
SM (OH) C14:1	3003	0.02	0.180	1314	-0.03	0.278	1084	-0.01	0.783	1149	0.00	0.967	806	0.05	0.168	643	0.09	0.020	193	0.17	0.017
Ser	3003	0.01	0.711	1314	0.04	0.198	1084	0.00	0.918	1149	0.02	0.482	806	-0.02	0.563	643	0.10	0.013	193	0.11	0.129
PC ae C42:5	3003	0.04	0.047	1314	-0.02	0.398	NA	NA	NA	1149	0.01	0.692	806	0.02	0.616	643	0.05	0.186	193	0.03	0.629
PC aa C42:1	3003	0.02	0.245	1314	0.00	0.909	1084	0.03	0.278	1149	-0.01	0.746	806	0.02	0.520	643	0.04	0.288	193	0.10	0.162
PC aa C34:2	3003	-0.03	0.170	1314	-0.06	0.044	1084	-0.02	0.563	1149	-0.02	0.507	806	-0.05	0.123	643	0.09	0.027	193	0.11	0.132

	KORA			NTR			EGCUT			TwinsUK			ERF			LLS			QIMR		
metabolite	n	r	p-value	n	r	p-value	n	r	p-value	n	r	p-value	n	r	p-value	n	r	p-value	n	r	p-value
lysoPC a C18:0	3003	0.03	0.063	1314	0.01	0.856	1084	-0.01	0.821	1149	0.00	0.970	806	0.01	0.727	643	0.04	0.262	193	0.05	0.513
lysoPC a C28:1	3003	0.00	0.840	1314	0.00	0.918	1084	0.02	0.556	1149	-0.03	0.372	806	0.09	0.007	643	0.03	0.428	193	0.20	0.006
PC ae C44:6	3003	0.02	0.186	1314	0.01	0.779	1084	0.04	0.192	1149	0.00	0.867	806	-0.01	0.812	643	0.03	0.517	193	0.04	0.620
PC ae C38:2	3003	0.01	0.736	1314	0.01	0.819	1084	0.01	0.740	1149	0.00	0.989	806	0.04	0.222	643	0.08	0.045	193	0.12	0.105
PC ae C44:3	3003	-0.01	0.566	1314	0.00	0.924	1084	0.03	0.309	1149	0.01	0.763	806	0.08	0.030	643	0.06	0.132	193	0.19	0.010
lysoPC a C16:0	3003	0.02	0.296	1314	0.03	0.319	1084	-0.01	0.662	1149	0.00	0.869	806	0.02	0.666	643	0.05	0.223	193	0.06	0.389
C14:1-OH	NA	NA	NA	NA	NA	NA	NA	NA	NA	1149	-0.07	0.012	NA	NA	NA	643	0.00	0.964	NA	NA	NA
SM (OH) C22:1	3003	0.01	0.613	1314	NA	1.000	1084	0.01	0.629	1149	0.00	0.869	NA	NA	NA	643	0.10	0.015	193	0.10	0.185
PC aa C42:5	3003	-0.03	0.085	1314	-0.01	0.692	1084	0.01	0.787	1149	-0.03	0.351	806	0.03	0.373	643	-0.06	0.101	193	0.06	0.401
lysoPC a C18:2	3003	0.01	0.476	1314	0.02	0.507	1084	0.01	0.839	1149	0.02	0.458	806	-0.04	0.229	643	0.07	0.067	193	0.14	0.057
Pro	3003	-0.01	0.614	1314	-0.03	0.321	1084	-0.02	0.523	1149	-0.03	0.260	806	0.00	0.952	643	-0.03	0.428	193	0.08	0.269
PC ae C34:2	3003	0.01	0.551	1314	-0.02	0.477	1084	-0.01	0.690	1149	-0.01	0.679	806	0.04	0.291	643	0.11	0.004	193	0.14	0.061
SM (OH) C24:1	3003	0.01	0.620	1314	NA	1.000	1084	0.02	0.412	1149	0.00	0.959	NA	NA	NA	643	0.06	0.107	193	0.06	0.396
PC ae C44:5	3003	0.02	0.233	1314	-0.02	0.504	1084	0.05	0.115	1149	-0.02	0.490	806	-0.02	0.631	643	0.06	0.113	193	0.04	0.596
PC ae C30:0	3003	0.01	0.616	1314	-0.01	0.609	NA	NA	NA	1149	-0.02	0.604	806	0.05	0.137	643	0.06	0.119	193	0.12	0.101
PC aa C40:4	3003	-0.04	0.036	1314	-0.03	0.346	1084	0.00	0.901	1149	-0.05	0.091	806	0.06	0.087	643	0.04	0.352	193	0.08	0.296
PC aa C42:2	3003	0.00	0.838	1314	0.00	0.913	1084	0.02	0.581	1149	0.02	0.521	806	0.00	0.922	643	0.04	0.344	193	0.15	0.032
PC ae C36:4	3003	0.01	0.643	1314	-0.03	0.219	1084	-0.01	0.744	1149	-0.01	0.708	806	0.05	0.194	643	0.13	0.001	193	0.13	0.066
Thr	3003	0.00	0.935	1314	0.01	0.819	NA	NA	NA	1149	0.02	0.530	NA	NA	NA	643	0.09	0.018	193	0.06	0.391
SM C26:1	3003	0.00	0.858	NA	NA	NA	1084	0.02	0.588	1149	0.00	0.977	806	0.05	0.199	643	0.01	0.720	193	0.11	0.131
PC ae C42:1	3003	-0.01	0.614	1314	0.00	0.904	1084	0.04	0.194	1149	0.00	0.883	806	0.05	0.174	643	0.04	0.356	193	0.17	0.015
PC ae C38:1	3003	0.03	0.122	1314	-0.01	0.764	1084	0.02	0.485	1149	0.02	0.432	NA	NA	NA	643	-0.03	0.422	NA	NA	NA
PC ae C38:5	3003	0.02	0.313	1314	-0.05	0.098	1084	0.00	0.914	1149	-0.01	0.659	806	0.04	0.278	643	0.10	0.014	193	0.11	0.143

metabolite	KORA			NTR			EGCUT			TwinsUK			ERF			LLS			QIMR		
	n	r	p-value	n	r	p-value	n	r	p-value	n	r	p-value	n	r	p-value	n	r	p-value	n	r	p-value
Gly	3003	0.03	0.132	1314	0.01	0.654	1084	-0.02	0.607	1149	0.01	0.727	806	-0.01	0.807	643	0.02	0.642	193	0.02	0.742
lysoPC a C18:1	3003	0.02	0.400	1314	0.02	0.559	1084	-0.01	0.727	1149	-0.01	0.735	806	-0.02	0.516	643	0.07	0.060	193	0.11	0.112
PC ae C42:2	3003	0.00	0.793	1314	0.00	0.896	1084	0.03	0.316	1149	0.00	0.987	806	-0.01	0.825	643	0.05	0.205	193	0.14	0.055
PC aa C24:0	3003	-0.03	0.072	NA	NA	NA	NA	NA	NA	1149	-0.02	0.489	NA	NA	NA	NA	NA	NA	193	0.25	0.001
C12:1	3003	-0.01	0.616	1314	0.03	0.339	1084	0.00	0.985	1149	-0.09	0.002	806	0.00	0.900	643	-0.01	0.797	193	-0.13	0.082
PC ae C36:3	3003	0.01	0.553	1314	-0.02	0.400	1084	-0.01	0.786	1149	-0.03	0.365	806	0.05	0.183	643	0.10	0.012	193	0.10	0.185
PC aa C40:2	3003	0.00	0.994	1314	0.00	0.909	1084	0.04	0.239	1149	0.00	0.873	806	0.01	0.843	643	0.02	0.697	193	0.14	0.058
PC aa C32:2	3003	-0.02	0.208	1314	-0.02	0.502	1084	-0.04	0.243	1149	-0.03	0.258	806	0.06	0.091	643	0.01	0.710	193	0.10	0.171
PC aa C38:0	3003	0.01	0.742	1314	-0.01	0.707	1084	0.02	0.608	1149	-0.02	0.561	806	0.04	0.236	643	0.03	0.397	193	0.10	0.159
Phe	3003	-0.01	0.498	1314	-0.04	0.168	1084	0.02	0.593	1149	-0.04	0.165	806	0.02	0.630	643	0.00	0.959	193	0.05	0.477
lysoPC a C16:1	3003	-0.01	0.519	1314	0.01	0.749	1084	-0.02	0.603	1149	-0.05	0.096	806	0.01	0.853	643	-0.02	0.601	193	0.05	0.521
SM C24:1	3003	-0.03	0.146	1314	-0.02	0.425	1084	0.01	0.651	1149	-0.01	0.749	806	0.00	0.987	643	0.03	0.399	193	0.03	0.650
PC ae C44:4	3003	0.02	0.360	1314	-0.01	0.610	1084	0.05	0.075	1149	-0.01	0.630	806	-0.03	0.428	643	0.01	0.758	193	0.04	0.581
C10	3003	0.01	0.506	NA	NA	NA	NA	NA	NA	NA	NA	NA	806	0.02	0.639	NA	NA	NA	NA	NA	NA
PC aa C36:0	3003	0.00	0.843	1314	-0.01	0.717	1084	0.00	0.994	1149	-0.02	0.580	806	0.04	0.265	643	0.03	0.431	193	0.18	0.011
C8:1	3003	-0.01	0.757	1314	0.02	0.562	1084	0.01	0.848	1149	-0.06	0.054	806	-0.02	0.519	643	0.01	0.877	193	-0.10	0.183
lysoPC a C28:0	3003	-0.02	0.382	1314	NA	1.000	NA	NA	NA	1149	-0.04	0.213	806	0.04	0.244	643	-0.08	0.034	193	0.19	0.009
PC aa C40:1	3003	0.02	0.370	1314	NA	1.000	NA	NA	NA	1149	0.01	0.685	NA	NA	NA	643	0.00	0.945	NA	NA	NA
C8	3003	0.01	0.498	NA	NA	NA	NA	NA	NA	NA	NA	NA	806	0.01	0.762	NA	NA	NA	NA	NA	NA
Gln	3003	0.04	0.029	1314	-0.09	0.002	1084	0.00	0.918	1149	-0.02	0.401	806	0.02	0.597	643	0.08	0.034	193	0.05	0.521
PC aa C32:3	3003	0.00	0.934	1314	-0.04	0.150	1084	-0.01	0.802	1149	-0.03	0.385	806	0.10	0.007	643	0.07	0.092	193	0.09	0.209
C0	3003	0.00	0.938	1314	0.02	0.440	1084	-0.02	0.566	1149	-0.07	0.027	806	-0.03	0.368	643	0.01	0.876	193	0.05	0.464
PC aa C36:5	3003	0.00	0.844	1314	-0.02	0.517	1084	-0.06	0.063	1149	0.00	0.866	806	0.01	0.785	643	0.00	0.984	193	0.17	0.018

7

	KORA			NTR			EGCUT			TwinsUK			ERF			LLS			QIMR		
metabolite	n	r	p-value	n	r	p-value	n	r	p-value	n	r	p-value	n	r	p-value	n	r	p-value	n	r	p-value
PC aa C38:4	3003	-0.01	0.498	1314	-0.05	0.083	1084	-0.01	0.706	1149	-0.04	0.151	806	0.04	0.267	643	0.06	0.145	193	0.11	0.141
xLeu	3003	-0.01	0.626	1314	-0.03	0.269	1084	0.00	0.947	1149	-0.02	0.545	NA	NA	NA	643	0.03	0.399	193	0.04	0.581
C18:1	3003	-0.01	0.532	1314	0.00	0.892	NA	NA	NA	1149	-0.07	0.027	806	0.00	0.937	643	-0.01	0.798	193	0.17	0.016
PC ae C34:3	3003	-0.01	0.424	1314	-0.02	0.559	1084	0.00	0.887	1149	0.01	0.837	806	0.05	0.157	643	0.08	0.047	193	0.10	0.156
Val	3003	-0.01	0.515	1314	-0.03	0.250	1084	-0.01	0.823	1149	-0.01	0.648	NA	NA	NA	643	0.05	0.242	193	0.08	0.284
PC ae C38:6	3003	0.01	0.619	1314	-0.04	0.169	1084	-0.01	0.664	1149	-0.02	0.548	806	0.05	0.129	643	0.05	0.175	193	0.14	0.045
Arg	3003	0.02	0.301	1314	-0.04	0.180	NA	NA	NA	1149	0.01	0.792	806	0.05	0.187	643	0.00	0.913	NA	NA	NA
PC aa C28:1	3003	-0.01	0.651	1314	-0.08	0.004	1084	-0.02	0.461	1149	-0.01	0.668	806	0.04	0.208	643	0.07	0.088	193	0.17	0.021
C5:1-DC	NA	NA	NA	1314	0.02	0.373	NA	NA	NA	NA	NA	NA	806	-0.01	0.849	NA	NA	NA	NA	NA	NA
His	3003	0.03	0.159	1314	-0.04	0.106	1084	0.00	0.892	1149	0.05	0.090	806	-0.06	0.114	643	0.04	0.362	193	0.10	0.181
lysoPC a C26:0	NA	NA	NA	1314	0.02	0.565	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
PC ae C38:0	3003	0.01	0.546	1314	0.01	0.806	1084	-0.03	0.407	1149	-0.01	0.657	806	0.02	0.639	643	0.01	0.823	193	0.14	0.045
Orn	3003	0.01	0.676	1314	-0.06	0.037	1084	0.01	0.706	1149	-0.03	0.327	806	-0.03	0.336	643	0.04	0.316	193	0.11	0.141
Trp	3003	0.00	0.913	1314	-0.05	0.084	NA	NA	NA	1149	0.00	0.883	NA	NA	NA	643	0.02	0.604	193	0.05	0.534
C18:2	3003	0.00	0.987	1314	0.00	0.929	1084	0.04	0.176	1149	-0.03	0.275	806	-0.02	0.545	643	0.01	0.842	193	0.18	0.013
PC aa C38:5	3003	0.00	0.920	1314	-0.04	0.142	1084	-0.04	0.201	1149	-0.03	0.321	806	0.02	0.494	643	0.04	0.304	193	0.15	0.041
PC aa C36:6	3003	0.00	0.841	1314	-0.01	0.764	1084	-0.03	0.295	1149	-0.02	0.442	806	0.06	0.066	643	0.00	0.902	193	0.16	0.030
PC aa C34:4	3003	-0.02	0.385	1314	-0.02	0.522	1084	-0.04	0.185	1149	-0.02	0.438	806	0.06	0.095	643	0.05	0.182	193	0.13	0.082
PC ae C34:0	3003	0.01	0.691	1314	-0.03	0.244	1084	-0.04	0.240	1149	-0.03	0.366	806	0.05	0.171	643	0.08	0.041	193	0.09	0.203
PC ae C32:1	3003	-0.01	0.570	1314	-0.03	0.306	1084	0.00	0.883	1149	-0.02	0.457	806	0.01	0.824	643	0.06	0.151	193	0.11	0.137
PC ae C32:2	3003	-0.02	0.403	1314	-0.04	0.105	1084	0.01	0.854	1149	-0.03	0.272	806	0.05	0.185	643	0.05	0.213	193	0.15	0.031
C14:1	3003	0.01	0.671	1314	0.00	0.935	1084	-0.01	0.858	1149	-0.09	0.003	806	0.01	0.688	643	0.00	0.963	193	0.07	0.302
C3	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	806	0.00	0.945	643	0.01	0.732	NA	NA	NA

	KORA			NTR			EGCUT			TwinsUK			ERF			LLS			QIMR		
metabolite	n	r	p-value	n	r	p-value	n	r	p-value	n	r	p-value	n	r	p-value	n	r	p-value	n	r	p-value
SM C18:1	3003	0.00	0.921	1314	-0.04	0.185	1084	0.02	0.413	1149	-0.05	0.068	806	0.07	0.060	643	0.04	0.353	193	0.07	0.341
PC ae C34:1	3003	0.00	0.983	1314	-0.02	0.407	1084	-0.02	0.461	1149	-0.03	0.260	806	0.03	0.347	643	0.09	0.025	193	0.08	0.269
C16:2	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	806	-0.01	0.836	NA	NA	NA	NA	NA	NA
PC ae C42:0	3003	0.00	0.908	1314	NA	1.000	NA	NA	NA	1149	-0.02	0.569	806	0.04	0.287	643	-0.04	0.302	193	0.10	0.152
PC aa C40:3	3003	-0.02	0.380	1314	0.00	0.935	1084	0.01	0.679	1149	-0.01	0.810	806	0.00	0.969	643	0.00	0.933	193	0.13	0.082
PC aa C42:4	3003	-0.03	0.166	1314	-0.01	0.631	1084	0.03	0.293	1149	0.00	0.954	806	0.04	0.306	643	0.03	0.420	193	0.14	0.051
SM C16:0	3003	-0.03	0.122	1314	-0.03	0.331	1084	0.02	0.572	1149	-0.01	0.681	806	0.02	0.479	643	0.09	0.017	193	0.08	0.257
C14:2	3003	0.00	0.924	1314	0.03	0.210	NA	NA	NA	1149	-0.07	0.012	806	0.03	0.474	643	-0.02	0.662	NA	NA	NA
SM C24:0	3003	-0.03	0.133	1314	-0.01	0.616	1084	0.01	0.623	1149	0.00	0.983	806	0.04	0.245	643	0.05	0.181	193	0.03	0.637
PC aa C36:4	3003	-0.01	0.705	1314	-0.04	0.119	1084	-0.02	0.426	1149	-0.02	0.549	806	0.02	0.498	643	0.10	0.013	193	0.13	0.062
PC ae C36:5	3003	-0.01	0.508	1314	-0.03	0.213	1084	-0.03	0.341	1149	0.00	0.907	806	0.04	0.246	643	0.10	0.014	193	0.15	0.033
SM C18:0	3003	0.00	0.823	1314	-0.04	0.163	1084	0.01	0.695	1149	-0.03	0.297	806	0.06	0.105	643	0.04	0.365	193	0.07	0.329
C10:1	3003	0.00	0.923	1314	0.03	0.363	NA	NA	NA	1149	-0.07	0.019	806	0.02	0.538	643	0.03	0.482	NA	NA	NA
lysoPC a C20:3	3003	0.00	0.923	1314	0.02	0.388	1084	-0.01	0.627	1149	-0.04	0.203	806	-0.04	0.254	643	0.06	0.129	193	0.10	0.166
PC ae C36:0	3003	0.01	0.696	1314	-0.01	0.600	1084	-0.04	0.159	1149	-0.01	0.671	806	0.01	0.872	643	0.03	0.430	193	0.14	0.046
PC aa C38:6	3003	0.01	0.716	1314	-0.03	0.338	1084	-0.02	0.474	1149	-0.02	0.514	806	0.04	0.280	643	0.00	0.906	193	0.13	0.071
PC aa C42:6	3003	-0.01	0.695	1314	-0.02	0.467	NA	NA	NA	1149	-0.02	0.453	806	0.08	0.023	643	-0.03	0.458	193	0.11	0.139

7

Chapter

8

Acknowledgements / Publications

Dankwoord

Het is zover. Mijn proefschrift is af. Het heeft wat moeite, bloed, zweet en tranen gekost, maar dan heb je ook wat. En nu komt dan het moment om de mensen die me hebben geholpen zo ver te komen hartelijk te bedanken voor al hun inspanning. Te beginnen met mijn promotoren.

Beste Cock, sinds ik binnen kwam lopen voor de mini-stage van mijn bachelor biologie ben je een enorme steun geweest. Toen ik liet weten verder te willen studeren in de genetische epidemiologie heb jij dat onmiddellijk gesteund en ik hoop dat ik je hier geen spijt van heb gegeven.

Beste Ben, hartelijk bedankt voor alle steun die ik van je heb mogen ontvangen op zakelijk en persoonlijk gebied. Je bood altijd een luisterend oor. Ik hoop dat je geniet van je pensioen en ben stiekem een beetje trots dat ik een van je laatste studenten heb mogen zijn.

Dear Najaf, thank you for the help you have offered me over the past few years and in particular the last few months. I'm grateful for the time and energy you've invested to make this thesis happen. Succes met je Nederlands.

En voor mijn commissieleden. Dank jullie allen voor het lid zijn van mijn commissie. Eline, ik heb graag met je samengewerkt. Ik ben je zeer dankbaar voor al het advies dat je mij hebt gegeven over de laatste paar jaar. Henning, zeer hartelijk bedankt voor je steun en opbouwend commentaar voor mijn artikelen. André, de enkele keren dat wij persoonlijk contact hadden over een project heb ik je kennis en hulp zeer op prijs gesteld. Je aanwijzingen voor mijn artikelen waren ook altijd zeer welkom. Beste Gert-Jan en Bert, helaas heb ik hier weinig te zeggen omdat we niet met elkaar aan een project gewerkt hebben. Ik ben in ieder geval zeer dankbaar dat u beiden de tijd hebben genomen om in mijn commissie plaats te nemen. Ik hoop dat we elkaar in de toekomst nog eens mogen treffen.

My dear paranimfs, Ayse and Adriana. Ayse, we started out around the same time, with me taking a year longer as I had to finish my master before I could start my PhD. We have shared an office the entire time (until the move to the new building) and have often shared a car to ride to conferences. I really like your humor and company and hope that no matter where we end up, this friendship is one that will stay. Adriana, for the longest time you were here I hardly knew you. You were one of the 'fish-girls' and that was about it. Since joining me and Ayse in our office together (pre-move) I have gotten to know you and your humor and find you are a great person to be around. I deeply admire your drive and dedication to find the answers you are looking for. Thank you for everything and I wish you good luck with your own thesis! By the way, it was fun at the Archeon, I hope you both will have good memories about it too.

Elza, jij verdient deze plaats van eer meteen na mijn paranimfen. Wat heb ik de afgelopen jaren een steun aan jou gehad. Zoals ik je al vaker heb gezegd, jij bent de beste secretaresse die de afdeling heeft gehad in al de tijd dat ik er rondloop. De dingen die jij voor ons gedaan

hebt en de tijd die je in ons gestoken hebt is echt geweldig. En dat boek dat je wilt schrijven komt er vast. Een (beta)lezer heb je alvast. Ik wens je veel succes in al je toekomstige bezigheden. Zet hem op.
Mijn familie. Ma, je bent er altijd voor mij geweest. Hoe hectisch de dingen ook werden. Hoeveel drama er ook was, je hebt altijd tijd voor mij. Ik hoop dat we altijd zo close blijven. Pa, veel succes met je duikschool. Je maakt je droom waar en dat is top. Nathan, sinds Inge geboren is ben je veel meer een broer geworden voor me. Ons leeftijdsverschil heeft dat vroeger altijd onmogelijk gemaakt. Die keren dat we samen op stap gingen om naar het worstelen te kijken waren helemaal top. Hopelijk kunnen we dat nog eens over doen. Misschien als ze weer terug naar Ahoy komen. Vorige keer had ik er geen erg in eerst jou te bellen voor ik ging boeken. Lida, je steun en enthousiasme, zelfs als dingen thuis met de kinderen niet lekker gingen, heb ik zeer gewaardeerd. Lieve Inge, wat zou ik graag willen dat je dit een keer kon lezen meissie. Het leven zit je ook niet mee. Al die vervelende ziekenhuizen ook. Ik hoop dat dingen wat makkelijker voor je worden in de toekomst. Lieve Koen, ben je lief voor je grote zus? Als je wat groter bent en dit zelf kan lezen dan hoop ik dat je weet dat ik ontzettend veel van je hou en dat ik er altijd voor je zal zijn. Er zijn vast leuke foto's van hoe je door de pannenkoekenboot hebt gerent tijdens mijn feestje. Je moeder was van te voren in ieder geval overtuigd dat jij de show zou stelen.
Dear Aaron, I remember well that the first thesis defense I witnessed was yours. My eyes were big like saucers as I had never seen anything like it. Even though you hardly knew me, you were happy I showed up for your defense and have been supportive of me throughout the years since. I hope life will continue to smile on you and your family.
Claudia, I hope you're doing well and are happy. Good luck with your own thesis defense.
Carla, Sara zet hem op. Jullie kunnen het. Ik weet zeker dat jullie dit jaar (of volgend jaar) ook je boekje af zullen maken en wens jullie veel succes tijdens dat proces. Sven, Elisa, Henriët, jullie moeten nog even, maar dat komt vast goed. Heel veel succes en verlies je humor niet.
Lennart, Maarten ontzettend bedankt voor jullie technische hulp. Zonder jullie hulp hadden een aantal projecten toch een stuk ingewikkelder geweest. Lennart, bedankt voor je hulp met mijn Nederlandse samenvatting. Veel geluk ook met jullie beider gezinnetjes.
Andy, Jeannette, hartstikke bedankt voor jullie hulp tijdens mijn bachelor stage en ook later de enkele keer dat ik het lab in moest voor een project. En dat geldt natuurlijk ook voor Andrea en Bernadette. Nano, bedankt voor alle computer steun. Peter, bedankt voor je hulp met de metabolomics en je informatie over metabool syndroom. Ik kon ook altijd zeer met je lachen.
Hilde, iedere keer dat je naar Rotterdam komt ben je oprecht geïnteresseerd in wat ik doe en hoe het met me gaat op persoonlijk niveau. Ik stel dit zeer op prijs. En ondanks dat ik er geen gebruik van heb gemaakt, heb ik het ook zeer gewaardeerd dat je bereid was je huis open te stellen voor mij als ik weer eens naar Brussel kwam om naar het worstelen te kijken in het Vorst Nationaal. November kom ik weer.

Rogier Hintzen, zeer hartelijk bedankt voor je hulp toen ik net kwam. Ik heb onze discussies zeer gewaardeerd. Naghmeh, Immy, Julia, jullie ook bedankt voor alles en ik wens jullie verder veel succes.
Dorret Boomsma, ik heb graag met je samengewerkt en je hulp met de heritabiliteit paper zeer op prijs gesteld.
Joris, Harmen, Marian en René, dank jullie allen voor jullie hulp bij mijn telomeer lengte artikelen.
Beste Albert Hofman en Ko Willems van Dijk, dank jullie beiden voor de tijd die jullie gestoken hebben in mijn artikelen.
Dear Veryan, I'm so glad we actually got to meet during the ENGAGE meeting here in Rotterdam. It's always nice to know whom your talking to on the phone all the time. All your help with coordinating the telomere length projects and keeping everyone on track is much appreciated. Nilesh, thank you for making my thesis possible. Without your help and support all the telomere length projects would never have existed in the first place. Dale, thank you for reading my papers so carefully. You've been a great help.
Dear Joanne, thank you for your help over the years. I really appreciate your input on the CHARGE GWAS 90+ paper. It was also nice meeting you first in Washington and more recently in Rotterdam. Hopefully we'll work together on some project in the future.
Also a big thank you for my co-authors in no particular order: Arfan Ikram, Joshua Bis, Bruce Psaty, Anne Newman, Gil Atzmon, Doug Kiel, Massimo Mangino, Eva Albrecht, Tim Spector, Nick Martin, Aysel Mueezzinler, Daan Kromhout, Hermann Brenner, Julia Raschenberger, Kirsi Pietiläinen, Christian Gieger, Karl-Heinz Ladwig, Florian Kronenberg, Jaakko Kaprio, Mait Raag, Harish Dharuri, Christina Lill, Lars Bertram, John Ioannidis, Maaike Schuur, Antonia Trichopoulou, Paolo Boffetta.
And finally to all my other co-authors and colleagues not mentioned above: I'm sorry I've missed you, but

thank you

About author

Linda Broer was born on July 27, 1985 in Rotterdam, the Netherlands. After finished her pre-university education at the Rijks Scholen Gemeenschap in Oud-Beijerland, she started to study Biology at the University of Groningen. After 1.5 years she changed universities to Nijmegen and obtained het Bachelor of Science degree in 2007. She obtained her Master of Science degree in Genetic Epidemiology from NIHES in 2009. As part of her internship she worked at the genetic epidemiology unit of the department of epidemiology of the Erasmus MC Rotterdam. In September 2009 she continued the work presented in this thesis at the genetic epidemiology unit of the department of epidemiology of the Erasmus MC Rotterdam under the supervision of Prof.Dr. C.M. van Duijn, Prof.Dr. B.A. Oostra and Dr. N. Amin.

Portfolio

	Year	ECTs
Courses		
• Barcelona: ENGAGE-IT	2010	1.4
• Leiden: Next Generation Sequencing	2010	1.4
• Skill Course: Linux	2011	1.4
• GE13: Next Generation Sequencing	2012	1.4
Seminars and workshops		
• Weekly scientific Seminars	2009 – 2013	5
Presentations		
• Gen. Epi: "Heat shock proteins and Alzheimer's"	2009	1
• Gen. Epi: "Het shock proteins and Parkinson's"	2009	1
• Gen. Epi: "Heat shock proteins and mortality"	2010	1
• Gen. Epi: "SPECIAL TOPIC: NGS – how to analyze the data"	2010	1
• Gen. Epi: "Venice Criteria"	2011	1
• Gen. Epi: "Telomere length and MetS"	2011	1
• NCHA longevity meeting: "Telomere length and Survivor Effect"	2012	1
• NCHA outreach meeting:		
• "Telomere length and Survivor Effect"	2012	1
• Gen. Epi: "Family relationships in Telomere length"	2012	1
• Gen. Epi: "Survivor Effect in telomere length"	2012	1
• NCHA longevity meeting: "Update on telomere length projects ENGAGE/NCHA"	2012	1
• Gen. Epi: "Investigation of old longevity hypothesis"	2012	1
• CHANCES: "Height and telomere length: investigation of an old longevity hypothesis"	2013	1
• Gen. Epi.: "GWAS of longevity"	2013	1
• IUNS 2013: CHANCES parallel session: "Height and telomere length"	2013	1

	Year	ECTs
(Inter)national conferences		
• Longevity Consortium, Washington DC: Poster "HSPs and longevity"	2010	1
• Whole Genome Sequencing Meeting, Utrecht	2010	1
• ESHG, Gothenburg: Poster "Venice Criteria"	2010	1
• BBMRI - Biobanking for Science, Amsterdam	2010	1
• CMSB 7th annual, Leiden	2010	1
• BBMRI-NL, Amsterdam: Poster "Exome sequencing of the oldest old"	2010	1
• CHANCES, Heidelberg	2011	1
• ESHG, Amsterdam: Poster "Updated Venice Criteria"	2011	1
• ICAD, Paris: Poster "Updated Venice Criteria"	2011	1
• NCHA longevity meeting	2011	1
• CMSB 8th annual	2011	1
• BBMRI, Rotterdam	2011	1
• NCHA longevity meeting Presentation: "Survivor Effect in telomere length"	2012	1
• NCHA outreach meeting Presentation: "Survivor Effect in telomere length"	2012	1
• ESHG, Nuremberg Poster "Age-related effects in telomere length"	2012	1
• ENGAGE, Rotterdam	2012	1
• BBMRI, Utrecht	2012	1
• CHANCES, Rotterdam Presentation: "Height and telomere length: investigation of an old longevity hypothesis"	2013	1
• IUNS, Granada Presentation: "Height and telomere length"	2013	1
Teaching Assistant		
• ESP01 - 2010	2010	1
• ESP43 - 2010	2010	1
• GE03 - 2010	2011	1
• ESP43 - 2011	2011	1
• Rewriting exercises for ESP43	2011	1
• ESP01 - 2011	2011	1
• ESP01 - 2012	2012	1
• ESP43 - 2012	2012	1

	Year	ECTs
Other		
• Organizing ERF and GRIP data	2010 – 2011	2
• Organizing Luchmeetings	2010 – 2013	3
• Supervising Master student Andrea Gasten	2009 – 2011	7.5
• Review work	2009 – 2013	3

List of publications

Aulchenko, Y. S., Hoppenbrouwers, I. A., Ramagopalan, S. V., **Broer, L.**, Jafari, N., Hillert, J., Link, J., Lundstrom, W., Greiner, E., Dessa Sadovnick, A., Goossens, D., Van Broeckhoven, C., Del-Favero, J., Ebers, G. C., Oostra, B. A., Van Duijn, C. M. & Hintzen, R. Q. (2008) Genetic variation in the kif1b locus influences susceptibility to multiple sclerosis. *Nat Genet,* 40(12)**,** 1402-3.

Hoppenbrouwers, I. A., Aulchenko, Y. S., Janssens, A. C., Ramagopalan, S. V., **Broer, L.**, Kayser, M., Ebers, G. C., Oostra, B. A., Van Duijn, C. M. & Hintzen, R. Q. (2009) Replication of cd58 and clec16a as genome-wide significant risk genes for multiple sclerosis. *J Hum Genet,* 54(11)**,** 676-80.

Jafari, N., **Broer, L.**, Hoppenbrouwers, I. A., Van Duijn, C. M. & Hintzen, R. Q. (2010) Infectious mononucleosis-linked hla class i single nucleotide polymorphism is associated with multiple sclerosis. *Mult Scler,* 16(11)**,** 1303-7.

Broer, L., Ikram, M. A., Schuur, M., Destefano, A. L., Bis, J. C., Liu, F., Rivadeneira, F., Uitterlinden, A. G., Beiser, A. S., Longstreth, W. T., Hofman, A., Aulchenko, Y., Seshadri, S., Fitzpatrick, A. L., Oostra, B. A., Breteler, M. M. & Van Duijn, C. M. (2011) Association of hsp70 and its co-chaperones with alzheimer's disease. *J Alzheimers Dis,* 25(1)**,** 93-102.

Broer, L., Koudstaal, P. J., Amin, N., Rivadeneira, F., Uitterlinden, A. G., Hofman, A., Oostra, B. A., Breteler, M. M., Ikram, M. A. & Van Duijn, C. M. (2011) Association of heat shock proteins with parkinson's disease. *Eur J Epidemiol,* 26(12)**,** 933-5.

Jafari, N., **Broer, L.**, Van Duijn, C. M., Janssens, A. C. & Hintzen, R. Q. (2011) Perspectives on the use of multiple sclerosis risk genes for prediction. *PLoS One,* 6(12)**,** e26493.

Amin, N., Byrne, E., Johnson, J., Chenevix-Trench, G., Walter, S., Nolte, I. M., Vink, J. M., Rawal, R., Mangino, M., Teumer, A., Keers, J. C., Verwoert, G., Baumeister, S., Biffar, R., Petersmann, A., Dahmen, N., Doering, A., Isaacs, A., **Broer, L.**, Wray, N. R., Montgomery, G. W., Levy, D., Psaty, B. M., Gudnason, V., Chakravarti, A., Sulem, P., Gudbjartsson, D. F., Kiemeney, L. A., Thorsteinsdottir, U., Stefansson, K., Van Rooij, F. J., Aulchenko, Y. S., Hottenga, J. J., Rivadeneira, F. R., Hofman, A., Uitterlinden, A. G., Hammond, C. J., Shin, S. Y., Ikram, A., Witteman, J. C., Janssens, A. C., Snieder, H., Tiemeier, H., Wolfenbuttel, B. H., Oostra, B. A., Heath, A. C., Wichmann, E., Spector, T. D., Grabe, H. J., Boomsma, D. I., Martin, N. G. & Van Duijn, C. M. (2012) Genome-wide association analysis of coffee drinking suggests association with cyp1a1/cyp1a2 and nrcam. *Mol Psychiatry,* 17(11)**,** 1116-29.

Demirkan, A., Van Duijn, C. M., Ugocsai, P., Isaacs, A., Pramstaller, P. P., Liebisch, G., Wilson, J. F., Johansson, A., Rudan, I., Aulchenko, Y. S., Kirichenko, A. V., Janssens, A. C., Jansen, R. C., Gnewuch, C., Domingues, F. S., Pattaro, C., Wild, S. H., Jonasson, I., Polasek, O., Zorkoltseva, I. V., Hofman, A., Karssen, L. C., Struchalin, M., Floyd, J., Igl, W., Biloglav, Z., **Broer, L.**, Pfeufer, A., Pichler, I., Campbell, S., Zaboli, G., Kolcic, I., Rivadeneira, F., Huffman, J., Hastie, N. D., Uitterlinden, A., Franke, L., Franklin, C. S., Vitart, V., Consortium, D., Nelson, C. P., Preuss, M., Consortium, C. A., Bis, J. C., O'donnell, C. J., Franceschini, N., Consortium, C., Witteman, J. C., Axenovich, T., Oostra, B. A., Meitinger, T., Hicks, A. A., Hayward, C., Wright, A. F., Gyllensten, U., Campbell, H., Schmitz, G. & Consortium, E. (2012) Genome-wide association study identifies novel loci associated with circulating phospho- and sphingolipid concentrations. *PLoS Genet,* 8(2)**,** e1002490.

Franceschini, N., Van Rooij, F. J., Prins, B. P., Feitosa, M. F., Karakas, M., Eckfeldt, J. H., Folsom, A. R., Kopp, J., Vaez, A., Andrews, J. S., Baumert, J., Boraska, V., **Broer, L.**, Hayward, C., Ngwa, J. S., Okada, Y., Polasek, O., Westra, H. J., Wang, Y. A., Del Greco, M. F., Glazer, N. L., Kapur, K., Kema, I. P., Lopez, L. M., Schillert, A., Smith, A. V., Winkler, C. A., Zgaga, L., The Lifelines Cohort, S., Bandinelli, S., Bergmann, S., Boban, M., Bochud, M., Chen, Y. D., Davies, G., Dehghan, A., Ding, J., Doering, A., Durda, J. P., Ferrucci, L., Franco, O. H., Franke, L., Gunjaca, G., Hofman, A., Hsu, F. C., Kolcic, I., Kraja, A., Kubo, M., Lackner, K. J., Launer, L., Loehr, L. R., Li, G., Meisinger, C., Nakamura, Y., Schwienbacher, C., Starr, J. M., Takahashi, A., Torlak, V., Uitterlinden, A. G., Vitart, V., Waldenberger, M., Wild, P. S., Kirin, M., Zeller, T., Zemunik, T., Zhang, Q., Ziegler, A., Blankenberg, S., Boerwinkle, E., Borecki, I. B., Campbell, H., Deary, I. J., Frayling, T. M., Gieger, C., Harris, T. B., Hicks, A. A., Koenig, W., Cj, O. D., Fox, C. S., Pramstaller, P. P., Psaty, B. M., Reiner, A. P., Rotter, J. I., Rudan, I., Snieder, H., Tanaka, T., Van Duijn, C. M., Vollenweider, P., Waeber, G., Wilson, J. F., Witteman, J. C., Wolffenbuttel, B. H., Wright, A. F., Wu, Q., Liu, Y., Jenny, N. S., North, K. E., Felix, J. F., Alizadeh, B. Z., Cupples, L. A., et al. (2012) Discovery and fine mapping of serum protein loci through transethnic meta-analysis. *Am J Hum Genet,* 91(4)**,** 744-753.

Gasten, A. C., Ramdas, W. D., **Broer, L.**, Van Koolwijk, L. M., Ikram, M. K., De Jong, P. T., Aulchenko, Y. S., Wolfs, R. C., Hofman, A., Rivadeneira, F., Uitterlinden, A. G., Oostra, B. A., Lemij, H. G., Klaver, C. C., Jansonius, N. M., Vingerling, J. R. & Van Duijn, C. M. (2012) A genetic epidemiologic study of candidate genes involved in the optic nerve head morphology. *Invest Ophthalmol Vis Sci,* 53(3)**,** 1485-91.

Murabito, J. M., White, C. C., Kavousi, M., Sun, Y. V., Feitosa, M. F., Nambi, V., Lamina, C., Schillert, A., Coassin, S., Bis, J. C., **Broer, L.**, Crawford, D. C., Franceschini, N., Frikke-Schmidt, R., Haun, M., Holewijn, S., Huffman, J. E., Hwang, S. J., Kiechl, S., Kollerits, B., Montasser, M. E., Nolte, I. M., Rudock, M. E., Senft, A., Teumer, A., Van Der Harst, P., Vitart, V., Waite, L. L., Wood, A. R., Wassel, C. L., Absher, D. M., Allison, M. A., Amin, N., Arnold, A., Asselbergs, F. W., Aulchenko, Y., Bandinelli, S., Barbalic, M., Boban, M., Brown-Gentry, K., Couper, D. J., Criqui, M. H., Dehghan, A., Den Heijer, M., Dieplinger, B., Ding, J., Dorr, M., Espinola-Klein, C., Felix, S. B., Ferrucci, L., Folsom, A. R., Fraedrich, G., Gibson, Q., Goodloe, R., Gunjaca, G., Haltmayer, M., Heiss, G., Hofman, A., Kieback, A., Kiemeney, L. A., Kolcic, I., Kullo, I. J., Kritchevsky, S. B., Lackner, K. J., Li, X., Lieb, W., Lohman, K., Meisinger, C., Melzer, D., Mohler, E. R., 3rd, Mudnic, I., Mueller, T., Navis, G., Oberhollenzer, F., Olin, J. W., O'connell, J., O'donnell, C. J., Palmas, W., Penninx, B. W., Petersmann, A., Polasek, O., Psaty, B. M., Rantner, B., Rice, K., Rivadeneira, F., Rotter, J. I., Seldenrijk, A., Stadler, M., Summerer, M., Tanaka, T., Tybjaerg-Hansen, A., Uitterlinden, A. G., Van Gilst, W. H., Vermeulen, S. H., Wild, S. H., Wild, P. S., Willeit, J., Zeller, T., Zemunik, T., Zgaga, L., et al. (2012) Association between chromosome 9p21 variants and the ankle-brachial index identified by a meta-analysis of 21 genome-wide association studies. *Circ Cardiovasc Genet,* 5(1)**,** 100-12.

Stolk, L., Perry, J. R., Chasman, D. I., He, C., Mangino, M., Sulem, P., Barbalic, M., **Broer, L.**, Byrne, E. M., Ernst, F., Esko, T., Franceschini, N., Gudbjartsson, D. F., Hottenga, J. J., Kraft, P., Mcardle, P. F., Porcu, E., Shin, S. Y., Smith, A. V., Van Wingerden, S., Zhai, G., Zhuang, W. V., Albrecht, E., Alizadeh, B. Z., Aspelund, T., Bandinelli, S., Lauc, L. B., Beckmann, J. S., Boban, M., Boerwinkle, E., Broekmans, F. J., Burri, A., Campbell, H., Chanock, S. J., Chen, C., Cornelis, M. C., Corre, T., Coviello, A. D., D'adamo, P., Davies, G., De Faire, U., De Geus, E. J., Deary, I. J., Dedoussis, G. V., Deloukas, P., Ebrahim, S., Eiriksdottir, G., Emilsson, V., Eriksson, J. G., Fauser, B. C., Ferreli, L., Ferrucci, L., Fischer, K., Folsom, A. R., Garcia, M. E., Gasparini, P., Gieger, C., Glazer, N., Grobbee, D. E., Hall, P., Haller, T., Hankinson, S. E., Hass, M., Hayward, C., Heath, A. C., Hofman, A., Ingelsson, E., Janssens, A. C., Johnson, A. D., Karasik, D., Kardia, S. L., Keyzer, J., Kiel, D. P., Kolcic, I., Kutalik, Z., Lahti, J., Lai, S., Laisk, T., Laven, J. S., Lawlor, D. A., Liu, J., Lopez, L. M., Louwers, Y. V., Magnusson, P. K., Marongiu, M., Martin, N. G., Klaric, I. M., Masciullo, C., Mcknight, B., Medland, S. E., Melzer, D., Mooser, V., Navarro, P., Newman, A. B., Nyholt, D. R., Onland-Moret, N. C., Palotie, A., Pare, G., Parker, A. N., Pedersen, N. L., et al. (2012) Meta-analyses identify 13 loci associated with age at menopause and highlight DNA repair and immune pathways. *Nat Genet,* 44(3)**,** 260-8.

Broer, L., Codd, V., Nyholt, D. R., Deelen, J., Mangino, M., Willemsen, G., Albrecht, E., Amin, N., Beekman, M., De Geus, E. J., Henders, A., Nelson, C. P., Steves, C. J., Wright, M. J., De Craen, A. J., Isaacs, A., Matthews, M., Moayyeri, A., Montgomery, G. W., Oostra, B. A., Vink, J. M., Spector, T. D., Slagboom, P. E., Martin, N. G., Samani, N. J., Van Duijn, C. M. & Boomsma, D. I. (2013) Meta-analysis of telomere length in 19 713 subjects reveals high heritability, stronger maternal inheritance and a paternal age effect. *Eur J Hum Genet*.

Broer, L., Demerath, E. W., Garcia, M. E., Homuth, G., Kaplan, R. C., Lunetta, K. L., Tanaka, T., Tranah, G. J., Walter, S., Arnold, A. M., Atzmon, G., Harris, T. B., Hoffmann, W., Karasik, D., Kiel, D. P., Kocher, T., Launer, L. J., Lohman, K. K., Rotter, J. I., Tiemeier, H., Uitterlinden, A. G., Wallaschofski, H., Bandinelli, S., Dorr, M., Ferrucci, L., Franceschini, N., Gudnason, V., Hofman, A., Liu, Y., Murabito, J. M., Newman, A. B., Oostra, B. A., Psaty, B. M., Smith, A. V. & Van Duijn, C. M. (2013) Association of heat shock proteins with all-cause mortality. Age (Dordr), 35(4), 1367-76.

Broer, L., Lill, C. M., Schuur, M., Amin, N., Roehr, J. T., Bertram, L., Ioannidis, J. P. & Van Duijn, C. M. (2013) Distinguishing true from false positives in genomic studies: P values. *Eur J Epidemiol,* 28(2)**,** 131-8.

Codd, V., Nelson, C. P., Albrecht, E., Mangino, M., Deelen, J., Buxton, J. L., Hottenga, J. J., Fischer, K., Esko, T., Surakka, I., **Broer, L.**, Nyholt, D. R., Mateo Leach, I., Salo, P., Hagg, S., Matthews, M. K., Palmen, J., Norata, G. D., O'reilly, P. F., Saleheen, D., Amin, N., Balmforth, A. J., Beekman, M., De Boer, R. A., Bohringer, S., Braund, P. S., Burton, P. R., Craen, A. J., Denniff, M., Dong, Y., Douroudis, K., Dubinina, E., Eriksson, J. G., Garlaschelli, K., Guo, D., Hartikainen, A. L., Henders, A. K., Houwing-Duistermaat, J. J., Kananen, L., Karssen, L. C., Kettunen, J., Klopp, N., Lagou, V., Van Leeuwen, E. M., Madden, P. A., Magi, R., Magnusson, P. K., Mannisto, S., Mccarthy, M. I., Medland, S. E., Mihailov, E., Montgomery, G. W., Oostra, B. A., Palotie, A., Peters, A., Pollard, H., Pouta, A., Prokopenko, I., Ripatti, S., Salomaa, V., Suchiman, H. E., Valdes, A. M., Verweij, N., Vinuela, A., Wang, X., Wichmann, H. E., Widen, E., Willemsen, G., Wright, M. J., Xia, K., Xiao, X., Van Veldhuisen, D. J., Catapano, A. L., Tobin, M. D., Hall, A. S., Blakemore, A. I., Van Gilst, W. H., Zhu, H., Consortium, C., Erdmann, J., Reilly, M. P., Kathiresan, S., Schunkert, H., Talmud, P. J., Pedersen, N. L., Perola, M., Ouwehand, W., Kaprio, J., Martin, N. G., Van Duijn, C. M., Hovatta, I., Gieger, C., Metspalu, A., Boomsma, D. I., Jarvelin, M. R., Slagboom, P. E., Thompson, J. R., Spector, T. D., Van Der Harst, P. & Samani, N. J. (2013) Identification of seven loci affecting mean telomere length and their association with disease. *Nat Genet,* 45(4)**,** 422-7.

Fernandez-Rhodes, L., Demerath, E. W., Cousminer, D. L., Tao, R., Dreyfus, J. G., Esko, T., Smith, A. V., Gudnason, V., Harris, T. B., Launer, L., Mcardle, P. F., Yerges-Armstrong, L. M., Elks, C. E., Strachan, D. P., Kutalik, Z., Vollenweider, P., Feenstra, B., Boyd, H. A., Metspalu, A., Mihailov, E., **Broer, L.**, Zillikens, M. C., Oostra, B., Van Duijn, C. M., Lunetta, K. L., Perry, J. R., Murray, A., Koller, D. L., Lai, D., Corre, T., Toniolo, D., Albrecht, E., Stockl, D., Grallert, H., Gieger, C., Hayward, C., Polasek, O., Rudan, I., Wilson, J. F., He, C., Kraft, P., Hu, F. B., Hunter, D. J., Hottenga, J. J., Willemsen, G., Boomsma, D. I., Byrne, E. M., Martin, N. G., Montgomery, G. W., Warrington, N. M., Pennell, C. E., Stolk, L., Visser, J. A., Hofman, A., Uitterlinden, A. G., Rivadeneira, F., Lin, P., Fisher, S. L., Bierut, L. J., Crisponi, L., Porcu, E., Mangino, M., Zhai, G., Spector, T. D., Buring, J. E., Rose, L. M., Ridker, P. M., Poole, C., Hirschhorn, J. N., Murabito, J. M., Chasman, D. I., Widen, E., North, K. E., Ong, K. K. & Franceschini, N. (2013) Association of adiposity genetic variants with menarche timing in 92,105 women of european descent. *Am J Epidemiol*.

Peters, M. J., **Broer, L.**, Willemen, H. L., Eiriksdottir, G., Hocking, L. J., Holliday, K. L., Horan, M. A., Meulenbelt, I., Neogi, T., Popham, M., Schmidt, C. O., Soni, A., Valdes, A. M., Amin, N., Dennison, E. M., Eijkelkamp, N., Harris, T. B., Hart, D. J., Hofman, A., Huygen, F. J., Jameson, K. A., Jones, G. T., Launer, L. J., Kerkhof, H. J., De Kruijf, M., Mcbeth, J., Kloppenburg, M., Ollier, W. E., Oostra, B., Payton, A., Rivadeneira, F., Smith, B. H., Smith, A. V., Stolk, L., Teumer, A., Thomson, W., Uitterlinden, A. G., Wang, K., Van Wingerden, S. H., Arden, N. K., Cooper, C., Felson, D., Gudnason, V., Macfarlane, G. J., Pendleton, N., Slagboom, P. E., Spector, T. D., Volzke, H., Kavelaars, A., Van Duijn, C. M., Williams, F. M. & Van Meurs, J. B. (2013) Genome-wide association study meta-analysis of chronic widespread pain: Evidence for involvement of the 5p15.2 region. *Ann Rheum Dis,* 72(3)**,** 427-36.

International Multiple Sclerosis Genetics, C., Lill, C. M., Schjeide, B. M., Graetz, C., Ban, M., Alcina, A., Ortiz, M. A., Perez, J., Damotte, V., Booth, D., Lopez De Lapuente, A., **Broer, L.**, Schilling, M., Akkad, D. A., Aktas, O., Alloza, I., Antiguedad, A., Arroyo, R., Blaschke, P., Buttmann, M., Chan, A., Compston, A., Cournu-Rebeix, I., Dorner, T., Epplen, J. T., Fernandez, O., Gerdes, L. A., Guillot-Noel, L., Hartung, H. P., Hoffjan, S., Izquierdo, G., Kemppinen, A., Kroner, A., Kubisch, C., Kumpfel, T., Li, S. C., Lindenberger, U., Lohse, P., Lubetzki, C., Luessi, F., Malhotra, S., Mescheriakova, J., Montalban, X., Papeix, C., Paredes, L. F., Rieckmann, P., Steinhagen-Thiessen, E., Winkelmann, A., Zettl, U. K., Hintzen, R., Vandenbroeck, K., Stewart, G., Fontaine, B., Comabella, M., Urcelay, E., Matesanz, F., Sawcer, S., Bertram, L. & Zipp, F. (2013) Manba, cxcr5, sox8, rps6kb1 and zbtb46 are genetic risk loci for multiple sclerosis. *Brain,* 136(Pt 6)**,** 1778-82.